Cap 1, 2, 6, 7

Rethinking Social Inquiry

Diverse Tools, Shared Standards

Second Edition

Edited by
Henry E. Brady and David Collier

[illegible]

ROWMAN & LITTLEFIELD PUBLISHERS, INC.
Lanham • Boulder • New York • Toronto • Plymouth, UK

Contents

1

Refocusing the Discussion of Methodology

Henry E. Brady, David Collier, and Jason Seawright

MAINSTREAM QUANTITATIVE METHODS, QUALITATIVE METHODS, AND STATISTICAL THEORY

The quest for shared standards of methodology and research design is an abiding concern in the social sciences. A recurring tension in this quest is the relationship between quantitative and qualitative methods. This book aims to rethink the contribution of these alternative approaches and to consider how scholars can most effectively draw on their respective strengths.

One view of the relation between quantitative and qualitative methodology is provided by what we call "mainstream quantitative methods," an approach based on the use of regression analysis and related techniques for causal inference. Scholars who champion this approach often invoke norms identified with these tools to argue for the superiority of quantitative research, sometimes suggesting that qualitative research could be greatly improved by following such norms more closely. These scholars in effect propose a quantitative template for qualitative research. In doing so, they have made some valuable suggestions that qualitative researchers would do well to consider.

Qualitative methodologists,[1] for their part, have raised legitimate con-

1. We understand qualitative methods as encompassing partially overlapping approaches such as the case-study method, small-N analysis, the comparative method, concept analysis, the comparative-historical method, the ethnographic tra-

cerns about the limitations of the quantitative template. Some qualitative analysts are dubious that the quantitative approach provides the only appropriate model for qualitative analysis. Others consider the quantitative template entirely inappropriate. Still others argue that the qualitative approach has strengths often lacking in quantitative studies and that quantitative analysts have much to learn from the qualitative tradition.

Yet another perspective on quantitative and qualitative methods is provided by ideas drawn from what we call "statistical theory." In contrast to mainstream quantitative methods, these ideas reflect a long history of skepticism about applying the assumptions behind regression analysis and related tools to real-world data in the social sciences.[2] This methodological approach sometimes advocates alternative techniques that allow researchers to draw more limited inferences based on fewer untested assumptions. According to this perspective, it is by no means evident that conventional quantitative tools are more powerful than qualitative tools.

Indeed, it is possible to draw on statistical theory to provide what may be thought of as a "statistical rationale" for many standard practices of qualitative research. This does *not* involve an admonition that qualitative analysts, in designing research, are expected to prove theorems in order to demonstrate that they have adopted the right methods. Rather, this rationale provides other kinds of insight into the analytic contribution of qualitative methods. A basic theme of this volume is that many qualitative research practices can be justified both on their own terms, and on the basis of this statistical rationale.

Overall, a meaningful discussion of methodology must be grounded in the premise that strengths *and* weaknesses are to be found in both the qual-

dition of field research, interpretivism, and constructivism. For many purposes, the quantitative-qualitative distinction may be disaggregated. In chapter 9 and the glossary, we propose four component dimensions: level of measurement, number of cases, whether explicit statistical tests are employed, and what we call thick versus thin analysis. Yet the simple quantitative-qualitative dichotomy offers a heuristic distinction that productively structures much of the current discussion.

2. The tradition to which we refer grows out of debates among statisticians on causal inference in experiments and observational studies. It may be dated to Karl Pearson's 1896 critique of G. Udny Yule's causal assessment, based on a regression analysis of observational data, of the relation between welfare policy and poverty in Britain (Stigler 1986: 351–53, 358). For a recent statement about this debate, see Freedman (1999). In addition to work within the discipline of statistics, we consider this tradition to encompass studies in the fields of econometrics, psychometrics, and measurement theory that, like Pearson's critique, explore the foundations of inference. We would also include methodological contributions by some scholars in political science and sociology whose work stands outside of the basic regression framework.

itative and quantitative approaches. Regarding the weaknesses, as Brady (chap. 3, this volume) puts it, qualitative researchers are perhaps "handicapped by a lack of quantification and small numbers of observations," whereas quantitative researchers may sometimes suffer from "procrustean quantification and a jumble of dissimilar cases." The most productive way to reconcile these two approaches is not through the unilateral imposition of norms, but rather through mutual learning.

THE DEBATE ON *DESIGNING SOCIAL INQUIRY*

In the present volume, we explore the relationship between quantitative and qualitative methodology through an extended discussion of a book that exemplifies the approach of mainstream quantitative methods: *Designing Social Inquiry: Scientific Inference in Qualitative Research* (hereafter KKV), by Gary King, Robert O. Keohane, and Sidney Verba.

KKV'S CONTRIBUTION

KKV has emerged as one of the most influential statements ever published on the relationship between quantitative and qualitative methods. The book is based on the tacit assumption that quantitative, large-N researchers have superior tools for solving many problems of methodology and research design, compared to their qualitative counterparts. Accordingly, KKV seeks to make such tools accessible to qualitative analysts, so as to help them design better research. While the premise is, in effect, the superiority of quantitative methods, the goal is to build bridges. The authors take seriously the idea that we should seek a common language for framing issues that arise in all forms of inquiry, and their effort to articulate the shared concerns of quantitative and qualitative research is a valuable contribution.

KKV's wide influence also stems from the systematization of quantitative methods that it offers. Although framed as an extended set of recommendations for qualitative researchers, the book is based on ideas drawn from the mainstream quantitative framework. In the course of summarizing these ideas, KKV offers numerous specific recommendations about different steps in the research process: for example, defining the research problem, specifying the theory, selecting cases and observations, testing descriptive and causal arguments, and subsequently retesting and refining the theory. In sum, KKV's reach is broad and its practical advice abundant.

At the most general level, by focusing scholarly attention on problems of research design, KKV aims to improve the practice of social science, understood as a collective effort to describe and explain political and social phenomena. KKV characterizes this collective effort as being concerned with descriptive and causal inference, a term which may seem alien to some qualitative researchers. However, as Charles Ragin emphasizes (chap. 3 online), "there is no necessary wedge separating the goal of 'inference'—the key concern of quantitative approaches—from the goal of making sense of cases—a common concern of qualitative approaches." The term "inference" can thus be seen as one specific label for a shared objective that spans diverse traditions of research.

KKV has had as great an impact, in terms of encouraging analysts to think about research design, as any book in the history of political science. The book is widely read in other fields as well, and it has exercised a salutary influence on many different branches of qualitative research. Even qualitative analysts who strongly disagree with KKV have adopted terms and distinctions introduced in the book. In addition, the concern of qualitative analysts with defending their own approach vis-à-vis KKV has pushed these scholars toward a more complete systematization of qualitative methods. In this and other ways, KKV has been strikingly successful in achieving its basic goal of encouraging researchers to think more carefully about methodological issues.

Finally, the authors of KKV deserve praise for their willingness to participate in an ongoing dialogue that is helping to advance this methodological discussion. In their response (reprinted as chapter 7 below) to a 1995 symposium on their book in the *American Political Science Review,* they observe that, "although our book may be the latest word on research design in political science [as of its publication in 1994], it is surely not the last" (111 this volume).

WHERE DO WE GO FROM HERE?

The present volume extends this methodological debate. We take as a point of departure a number of basic concerns about KKV's framework.

In our view, KKV gives insufficient recognition to well-known limitations of mainstream quantitative methods. The book does present a useful discussion of assumptions that underlie regression analysis. Yet KKV does not devote adequate attention to a key statistical idea: Regression analysis depends on the model, and if the model is wrong, so is the analysis. For this reason, estimating a regression model with empirical data does not fully test the model. Relatedly, KKV places strong emphasis on evaluating uncertainty. Yet the book fails to acknowledge that significance tests are designed to evaluate *specific* kinds of uncertainty, and that the common practice of employing them as a *general-purpose* tool for estimating uncertainty extends these tests beyond the uses for which they were intended.

Against this backdrop, KKV goes too far in advocating the perspective of mainstream quantitative methods as a foundation for research design and qualitative inquiry. We are convinced that this perspective provides an excessively narrow understanding of the research process. More specifically, along with being too confident about the strengths of quantitative tools, the book gives insufficient recognition to the contributions of qualitative tools. KKV overemphasizes the strategy of increasing the number of observations, and it overlooks the different kinds of observations and the different ways that data are used in quantitative and qualitative research. The book is inattentive to the risk that increasing the N may push scholars toward an untenable level of generality and a loss of contextual knowledge. It overstates its warning against post hoc hypothesis formation and standard practices of disciplined inductive research. Relatedly, it neglects the fact that econometric writing on "specification searches" has sought to systematize inductive procedures. Finally, KKV occasionally refers to tradeoffs, yet the book does not acknowledge that they must be a basic concern in designing research.

We want to be clear about what these criticisms do and do not amount to. They do not amount to a rejection of the basic enterprise of striving for a shared vocabulary and framework for both quantitative and qualitative research. Indeed, we are strongly committed to the quest for a common framework. While we have great respect for scholars who explore epistemological issues, we worry that such concerns may sometimes unnecessarily lead researchers and students to take sides and to engage in polemics. Thus, we share KKV's (4–5) view that quantitative and qualitative methods are founded on essentially similar epistemologies.

Correspondingly, the present volume is certainly not meant to widen the gap between the qualitative and quantitative approaches by identifying profound and obdurate differences. Indeed, we would argue that the differences are less deep-seated than is sometimes believed. To the extent that differences do exist, however, we take the normative position that a basic goal in work on methodology is to overcome these differences. We should seek a shared framework allowing researchers using diverse analytic techniques to develop evidence that is convincing to analysts of differing methodological persuasions. This larger body of mutually accepted evidence can, in turn, contribute to finding better answers to the substantive questions that drive social research.

TOOLS AND STANDARDS

As we suggest in the subtitle of this book, while analysts have diverse tools for designing, executing, and evaluating research, it is meaningful to seek shared standards for employing such tools. These shared standards can facilitate recognition of common criteria for good research among scholars who use different tools. Methodological pluralism and analytic rigor can be combined.

By tools we mean the specific research procedures and practices employed by quantitative and qualitative researchers. Some tools are highly systematized and have elaborate technical underpinnings. Examples of such tools are regression analysis, structural equation modeling, factor analysis, tests of statistical significance, and probability theory. Increasing the number of observations is a research tool repeatedly advocated by KKV. Other tools include qualitative research practices such as within-case analysis, process tracing, procedures for avoiding conceptual stretching, qualitative validity assessment, and strategies for the comparison of matching and contrasting cases. Methods of data collection are also tools: for example, public opinion research, focus groups, participant observation, event scoring, archival research, content analysis, the construction of "unobtrusive measures," and the systematic compilation of secondary sources. At various points in the text, we have introduced summary tables that provide an overview of the different tools being discussed, and many tools are also discussed in the glossary.

The chapters in the present volume devote considerable attention to various methodological tools that KKV undervalues or overlooks. The following paragraphs enumerate four broad methodological literatures with which many of these tools are identified. Some correspond to standard practices of qualitative researchers; others are derived from statistical theory.

1. *Logical and Statistical Foundations of Causal Inference.* A large body of research on the logical and statistical foundations of causal inference expresses considerable skepticism about causal inference based on observational data. This literature points to the need for more robust approaches than those advocated in mainstream quantitative methodology.
2. *Concepts.* Research on concepts, concept formation, and the evolution of concepts in the course of research makes it clear that sustained attention to conceptual issues is an indispensable component of research design. The insights of this literature suggest that the limited advice that KKV does give on working with concepts in fact points in the wrong direction.
3. *Measurement.* A major literature located in the fields of mathematical measurement theory and psychometrics provides researchers with systematic guidance for measurement. This literature emphasizes, for example, the contextual specificity of measurement claims, reinforcing the conviction of many political scientists that knowledge of context and care in bounding the generality of research findings must be a central concern in research design. Such guidance is lacking in KKV.
4. *Causal Inference in Case Studies.* A long tradition of writing has explored tools and strategies of causal inference in case studies: for example, process tracing and other forms of within-case analysis; the deliberate selection of "most-likely," "least-likely," and "deviant" cases; and, in the comparative case-study tradition, the methods of agreement and difference. KKV seeks to subsume these tools within its own framework, based on the norms of large-N quantitative analysis. The case-study literature in effect turns KKV's argument on its head, suggesting that (a) the practice of causal inference in qualitative research is viable on its own terms, and (b) inference in quantitative research can sometimes be improved through the use of tools strongly identified with the qualitative tradition.

Through focusing on tools drawn from these diverse areas of methodology, as well as on more conventional quantitative tools, we seek to lay a stronger foundation for an integrated approach to the design and execution of research.

All research tools, both qualitative and quantitative, must be subject to critical evaluation. Correspondingly, scholars should seek shared standards for assessing and applying these tools. Relevant standards must include attention to basic trade-offs that arise in conducting research. Once we acknowledge that not all analytic goals can be achieved simultaneously—Przeworski and Teune's trade-offs among accuracy, generality, parsimony, and causality are a famous example (1970: 20–23)—then it is easier to move toward a recognition that alternative methodological tools are relevant and appropriate, depending on the goals and context of the research.

Neither qualitative nor quantitative analysts have a ready-made formula for producing good research. We are convinced that the wide influence exercised by KKV derives in part from the book's implicit claim that, if scholars follow the recommendations in the book, it is relatively straightforward to do good quantitative research; as well as the explicit argument that qualitative researchers, to the degree possible, should apply the quantitative template.[3]

3. KKV does briefly note the limitations of quantitative research. The book states that "[i]n both quantitative and qualitative research, we engage in the imperfect application of theoretical standards of inference to inherently imperfect research designs and empirical data" (7; see also 8–9). However, in the eyes of many critics, KKV does not follow through on these words of caution, instead going too far in extending the norms of quantitative analysis to qualitative research. Further, KKV's statements on the pages just cited are closely linked to its arguments about estimating error, and the authors are far more confident than we are about the viability of error estimates in quantitative research, not to mention in qualitative research. See,

In fact, it is difficult to make causal inferences from observational data, especially when research focuses on complex political processes. Behind the apparent precision of quantitative findings lie many potential problems concerning equivalence of cases, conceptualization and measurement, assumptions about the data, and choices about model specification such as which variables to include. The interpretability of quantitative findings is strongly constrained by the skill with which these problems are addressed. Thus, both qualitative and quantitative research are hard to do well. It is by recognizing the challenges faced in both research traditions that these two approaches can learn from one another.

Scholars who make particular choices about trade-offs that arise in the design of research should recognize the contributions of those who opt for different choices. For example, let us suppose that a scholar has decided, after careful consideration, to focus on a small N to carry out a fine-grained, contextually sensitive analysis that will facilitate operationalizing a difficult concept. A large-N researcher should, in principle, be willing to recognize this choice as legitimate.

At the same time, the small-N researcher should recognize that the advantages of focusing on few cases must be weighed against the costs. These costs include, for example, forgoing large-N tools for measurement validation and losing the generality that might be achieved if a wider range of cases is considered. In short, researchers should recognize the potential strengths and weaknesses of alternative approaches, and they should be prepared to justify the choices they have made.

TOWARD AN ALTERNATIVE VIEW OF METHODOLOGY

Building on these themes, the present volume develops alternative arguments about the appropriate balance between the quantitative and qualitative traditions, and about research design and methodology more broadly.[4] Here are some key steps in these arguments.

1. *In the social sciences, qualitative research is hard to do well. Quantitative research is also hard to do well. Each tradition can and should learn from the other.* One version of conventional wisdom holds that achieving

for example, Bartels's discussion of assessing measurement error (chap. 4, this volume), as well as the discussion in chapter 9 focused on the misuse of significance tests.

4. While issues of descriptive inference are a recurring theme in the following chapters (see, e.g. 34–37, 132–40 this volume), the focus here is primarily on causal inference.

analytic rigor is more difficult in qualitative than in quantitative research. Yet in quantitative research, making valid inferences about complex political processes on the basis of observational data is likewise extremely difficult. There are no quick and easy recipes for either qualitative or quantitative analysis. In the face of these shared challenges, the two traditions have developed distinctive and complementary tools.

   a. *A central reason why both qualitative and quantitative research are hard to do well is that any study based on observational (i.e., nonexperimental) data faces the fundamental inferential challenge of eliminating rival explanations.* Scholars must recognize the great divide between experiments and observational studies. Experiments eliminate rival explanations by randomly assigning the values of the explanatory variable to the units being analyzed. By contrast, in all observational studies, eliminating rival explanations is a daunting challenge. The key point, and a central concern of this book, is that quantitative and qualitative observational studies generally address this shared challenge in different ways.

2. *Mainstream quantitative methodologists sometimes advocate the quantitative approach as a general template for conducting research. By contrast, some statistical theorists question the general applicability of the conventional quantitative approach.* Strong advocacy of the quantitative template is found in many disciplinary subfields. Yet it is essential that political scientists—and scholars in other fields as well—take a broader view and reflect more deeply on the contributions and limitations of both qualitative and quantitative methods. A valuable component of this broader view draws on ideas from statistical theory.

   a. *One recurring issue regarding the tradition of advocacy based on the quantitative template concerns how much scholars can in fact learn from findings based on regression analysis, as well as their capacity to estimate the degree of uncertainty associated with these findings.* For regression results to be meaningful, analysts must assume, as noted earlier in this chapter, that they have begun with the correct statistical model. Empirical data analysis may provide some insight into the plausibility of this assumption, yet such analysis does not fully test the assumption. Another key idea identified with the quantitative template concerns the capacity to estimate uncertainty. Unfortunately, in some areas of research, standard practice in the use of significance tests extends their application to evaluating forms of uncertainty that they were not designed to assess.

   b. *Another issue regarding the quantitative template is the recurring recommendation that researchers can gain inferential leverage in addressing rival explanations by increasing the number of observations—in the con-*

ventional sense of increasing the N. Yet this advice is not always helpful, in part because it may push scholars to compare cases that are not analytically equivalent. Although adding new observations is frequently useful, adding observations from a different spatial or temporal context or at a different level of analysis can extend the research beyond the setting for which the investigator can make valid inferences. While some scholars might be concerned that this focus on context leads researchers toward a posture of excessive particularism, concern with context is in fact a prerequisite for achieving descriptive and causal inference that is valid and rigorous.

3. *In making choices about increasing leverage in causal inference, and to address the concerns just noted, scholars should recognize the contributions of different kinds of observations.* It is productive to distinguish between two quite distinct uses of the term "observation," one drawn from the quantitative tradition, the other from the qualitative tradition. Examples of these two types are presented in the appendix (see also 184–96 this volume).
   a. *Data-set observations.* These observations are collected as an array of scores on specific variables for a designated sample of cases, involving what is sometimes called a rectangular data set. Missing data are an obstacle to causal inference based on data-set observations; it is therefore valuable that the data set be complete. Data-set observations play a central role not only in quantitative research, but also in qualitative research that is based on cross-case analysis.
   b. *Causal-process observations.* These observations about context, process, or mechanism provide an alternative source of insight into the relationships among the explanatory variables, and between these variables and the dependent variable. Causal-process observations are sometimes less complete than data-set observations, in the sense that they routinely do not constitute a full set of scores across a given set of variables and cases. The strength of causal-process observations lies not in breadth of coverage, but depth of insight. Even one causal-process observation may be valuable in making inferences. Such observations are routinely used in qualitative research based on within-case analysis, and they can also be an important tool in quantitative analysis.
   c. *These two types of observations have contrasting implications for maintaining an appropriate scope of comparison.* A focus on increasing the number of data-set observations, either at the same level of analysis or in subunits at a lower level of analysis, can yield major analytic gains, but it can also push scholars toward shifts in the domain of analysis that may be counterproductive. By contrast, the search for additional causal-process observations may occur within the original domain.

4. *Methodological discussions could benefit from stronger advocacy from the side of the qualitative template, and all researchers should consider carefully some long-standing methodological priorities that derive from the qualitative perspective.* The qualitative template can make important contributions to broader methodological agendas. For example:
   a. *Knowledge of cases and context contributes to achieving valid inference.* To expand on the earlier argument (2b and 3c), analytic leverage can derive from a close knowledge of cases and context, which can directly contribute to more valid descriptive and causal inference. This knowledge sensitizes researchers to the impact of cultural, economic, and historical settings, and to the fact that subunits of a given case may be very different from the overall case. In other words, knowledge of context provides insight into potentially significant factors that are not among the variables being formally considered. In this sense, it helps us to know what is hidden behind the assumption "other things being equal," which is in turn crucial for the causal homogeneity assumption that is a requisite for valid causal inference. As discussed in this volume, such contextual knowledge is also crucial for measurement validity. Leverage derived from detailed knowledge of cases and context is closely connected to the idea of causal-process observations just discussed. Such knowledge is invaluable in both quantitative and qualitative research.
   b. *Inductive analysis can play a major role in achieving valid inference and generating new ideas. Induction is important in both qualitative and quantitative research.* Mainstream quantitative researchers are sometimes too quick in dismissing the contribution to scholarly knowledge of inductive analysis and of the retesting of hypotheses against the same set of cases, on occasion invoking the traditional mandate to avoid "post hoc" hypothesis reformulation and theory testing. Yet even in technically advanced forms of statistical estimation, quantitative researchers routinely test alternative specifications against a given set of data (i.e., specification searches) and on this basis seek to make complex judgments about which specification is best. This iterated refinement of models and hypotheses constitutes a point of similarity to the inductive practices that are perhaps more widely recognized in qualitative research. Inductive procedures play a role in both traditions, and developing norms that guide, systematize, and make explicit these procedures for causal inference should be a basic concern of methodology.
   c. *These arguments add up to a view of methodology in which qualitative*

research has a major role. The norms and practices of qualitative research deserve, in their own terms, serious attention in broader discussions of methodology. Further, ideas drawn from qualitative methodology can improve quantitative practices by addressing weaknesses in the quantitative approach.

5. *The contribution of qualitative methods can be justified both from within the qualitative tradition itself, and from the perspective of statistical theory.* Greater attention to qualitative methods can be justified, first of all, by the lessons that qualitative analysts learn from their own research. Many qualitative practices can also be justified on the basis of arguments drawn from statistical theory. Among the goals of this volume are to develop what may be thought of as a statistical rationale for qualitative research and to explore specific ways in which statistical theory can improve both qualitative and quantitative analysis. This perspective is very different from that of much writing in the tradition of mainstream quantitative methods, which seeks to subordinate qualitative research to the quantitative template.
6. *If both qualitative and quantitative methods are to play important roles as sources of norms and practices for good research, scholars must face the challenge of adjudicating between potentially conflicting methodological norms.* Such adjudication requires recognition of a basic fact and a basic priority.
   a. *Research design involves fundamental trade-offs.* Methodological advice needs to be framed in light of basic trade-offs among: (a) alternative goals of research, (b) the types of observations researchers utilize, and (c) the diverse tools they employ for descriptive and causal inference. A methodological framework that does not centrally consider trade-offs is incomplete.
   b. *Scholars should develop shared standards.* A basic goal of methodology should be to establish shared standards for managing these trade-offs. Shared standards can become the basis for combining the strengths of qualitative and quantitative tools.

These arguments form the basis for the ideas presented throughout this volume. The remainder of this introduction provides an overview of the chapters that follow.

OVERVIEW OF THE CHAPTERS

Part I of this book seeks to advance this methodological debate by building on the discussion stimulated by King, Keohane, and Verba's *Designing Social Inquiry*. We bring together a number of previously published statements in

this discussion—some presented basically in their original form, others extensively revised[5]—along with two introductory chapters, two concluding chapters that draw together different strands in this debate, and an appendix. The glossary defines basic terms, with a core definition presented in the first paragraph of each entry; for certain terms, subsequent paragraphs elaborate on the definition. Part I is divided into four sections: an Introduction (chaps. 1–2), Critiques of the Quantitative Template (chaps. 3–5), Linking the Quantitative and Qualitative Traditions (chaps. 6–7), and Diverse Tools, Shared Standards (chaps. 8–9).

INTRODUCTION

Following the present introductory chapter, David Collier, Jason Seawright, and Gerardo L. Munck (chap. 2) provide a detailed summary of the methodological recommendations offered by KKV, thereby framing the discussion developed later in the book. Chapter 2 focuses on the definition of scientific research, the treatment of descriptive and causal inference, and the assumptions that underlie causal inference. The chapter then synthesizes KKV's recommendations by formulating a series of guidelines for the design and execution of research. Although KKV does not present most of its methodological advice in terms of explicit rules, much of its argument can productively be summarized in this manner. Chapter 2 concludes by offering an initial assessment of KKV's framework.

CRITIQUES OF THE QUANTITATIVE TEMPLATE

How useful is the quantitative template as a guide for qualitative research? This question is addressed in chapters 3–5. It merits emphasis that these chapters praise KKV for presenting mainstream ideas of quantitative inference in a minimally technical manner; for offering many useful didactic arguments about how qualitative analysts can improve their research by applying simple lessons from statistics and econometrics; and for making genuine contributions to the field of methodology. At the same time, however, these chapters reconsider and challenge some of KKV's basic arguments.

"Doing Good and Doing Better: How Far Does the Quantitative Template Get Us?" by Henry E. Brady (chap. 3) argues that KKV does not adequately

5. The relationship of each chapter to previously published material is explained in the acknowledgment of permission to reprint copyrighted material at the end of this volume.

consider the foundations of causal inference in quantitative research, and that the book does not properly attend to conceptualization and measurement. Regarding causal inference, Brady suggests that KKV pays insufficient attention to the challenges faced in research based on observational, as opposed to experimental, data. Specifically, the book fails to discuss how theory and preexisting knowledge can justify a key assumption that underlies causal assessment with observational data, that is, the assumption that conclusions are not distorted by missing variables. Concerning the second theme, Brady finds that KKV ignores major issues of concept formation and basic ideas from the literature on measurement. This latter body of work shows that quantitative measurement is ultimately based on qualitative comparisons, suggesting a very different relation between quantitative and qualitative work than is advocated by KKV.

"Some Unfulfilled Promises of Quantitative Imperialism" by Larry M. Bartels (chap. 4) suggests that KKV's recommendations for qualitative researchers exaggerate the degree to which quantitative methodology offers a coherent, unified approach to problems of scientific inference. KKV classifies research activities that do not fit within its framework as prescientific, leading the authors to a false separation between (a) producing unstructured knowledge and "understanding," and (b) making scientific inferences. Bartels is convinced that unstructured knowledge and understanding are a necessary part of inference. Likewise, in Bartels's view, KKV claims to have solutions to several methodological problems that neither its authors nor anyone else can currently solve. These include the challenge of estimating the uncertainty of conclusions in qualitative (and even quantitative) research; distinguishing between the contribution made by qualitative evidence and quantitative evidence in analyses that employ both; assessing the impact of measurement error in multivariate analysis; and multiplying observations without violating the causal homogeneity assumption. According to Bartels, the fact that leading practitioners in political science cannot adequately address these problems suggests that they may be the most important issues currently pending for further research on methodology.

"How Inference in the Social (but Not the Physical) Sciences Neglects Theoretical Anomaly" by Ronald Rogowski (chap. 5) argues that KKV underestimates the importance of theory in the practice of research. KKV's rules about case selection and the number of cases needed to support or challenge a theory reflect this inattention. In fact, following KKV's rules would lead scholars to reject as bad science some of the most influential works in the recent history of comparative politics. Single-case studies are particularly useful in challenging already-existing theories, if these theories are precisely formulated; yet KKV claims that a single case cannot discredit a scientific theory. Rogowski suggests that if the analyst employs theory that

is both powerful and precise, carefully constructed studies that examine anomalous cases can be invaluable, notwithstanding KKV's warnings about selection bias.

QUALITATIVE TOOLS

The basic analytic tools of quantitative researchers are reasonably well understood. By contrast, qualitative tools are less well codified and recognized. What are these tools? This question was addressed in Chapters 7 to 9 of the first edition (as well as in Chapter 6), and for the second edition these chapters are now available on the Rowman & Littlefield website (as discussed in the Preface).

LINKING THE QUANTITATIVE AND QUALITATIVE TRADITIONS

Given that the qualitative and quantitative traditions have distinctive strengths, how can they best be combined? The third section offers two perspectives on this challenge. "Bridging the Quantitative-Qualitative Divide" by Sidney Tarrow (chap. 6) offers valuable suggestions for linking quantitative and qualitative research. Qualitative analysis is better suited than quantitative research for process tracing, for exploring the tipping points that play a critical role in shaping long-term processes of change, and for providing more nuanced insight into findings derived from quantitative investigation. Quantitative analysis, in turn, can frame and generalize the findings of qualitative studies. In Tarrow's view, the most valuable interaction between the two research traditions occurs when scholars "triangulate" among alternative methods and data sources in addressing a given research problem.

"The Importance of Research Design" (chap. 7), reprinted here with the kind permission of Gary King, Robert O. Keohane, and Sidney Verba, is from the 1995 symposium on *Designing Social Inquiry*, published in the *American Political Science Review*. This chapter should be understood as the authors' interim response to the ongoing debate about linking the quantitative and qualitative traditions. Because it was written in 1995, it obviously does not take into account all the arguments in the present volume, though it does make reference to ideas presented here by Rogowski and Tarrow (and also Collier, Mahoney, and Seawright, from the online posting), as well as to arguments advanced in some other chapters.

King, Keohane, and Verba underscore central themes in KKV and clarify certain key ideas. The authors argue that the fundamental challenge for both quantitative and qualitative analysis is good research design. King, Keohane, and Verba agree with Rogowski on the importance of theory, although they emphasize that telling people how to theorize is not their goal. Perhaps most significantly, they argue that "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" (chap. 7). All researchers, whether quantitative or qualitative, need to understand and utilize the same logic of inference.

King, Keohane, and Verba go on to explore and illustrate two related themes: the idea of science as a collective enterprise, which they discuss in relation to well-known books of Arend Lijphart and William Sheridan Allen; and problems of addressing selection bias, which they illustrate by reference to books by Peter Katzenstein and Robert Bates. Finally, the chapter proposes that Tarrow's arguments about "triangular conclusions" provide a valuable unifying idea that brings together the diverse perspectives on methodology under discussion.

DIVERSE TOOLS, SHARED STANDARDS

The final part of the book synthesizes and extends the debate on quantitative and qualitative methods. We argue that, precisely because researchers have a diverse set of methodological tools at their disposal, it is essential to seek shared standards for the application of these tools.

"Critiques, Responses, and Trade-Offs: Drawing Together the Debate," by David Collier, Henry E. Brady, and Jason Seawright (chap. 8), integrates and evaluates this methodological discussion. In a further effort to bridge the quantitative-qualitative divide, chapter 8 reviews the critiques of KKV offered in chapters 3–6 of the present volume and in the online chapters and formulates responses that draw on ideas derived from statistical theory. Two of the critiques concern the challenge of doing research that is important and the issue of probabilistic versus deterministic models of causation. For these topics, the statistical response calls for a synthesis that combines elements of KKV's position and the critique. For other parts of the debate—on conceptualization and measurement, and on selection bias—statistical arguments emerge that more strongly reinforce the critique of KKV. The final part of this chapter explores the idea that trade-offs are inherent in research design and develops the argument that the search for shared standards necessarily poses the challenge of managing these trade-offs.

The final chapter of Part I offers some broader conclusions about tools for causal inference. "Sources of Leverage in Causal Inference: Toward an Alternative View of Methodology," by David Collier, Henry E. Brady, and Jason Seawright (chap. 9), focuses on the fundamental challenge of eliminating rival explanations and making good causal inferences. This chapter formulates several methodological distinctions that help bring into sharper focus the relationship between the quantitative and qualitative traditions and, more specifically, the contrasts in how they deal with causal inference. A further goal of this discussion is to explore the implications of the distinction between data-set observations and causal-process observations. The chapter argues that this distinction offers a more realistic picture of the contributions to causal inference of both quantitative and qualitative tools—and of how these differing contributions can be integrated.

Taken together, the arguments developed in this volume lead us to reflect on the expanding influence in social science of increasingly technical approaches to method and theory. We advocate an eclectic position in response to this trend. While it is essential to recognize the powerful contribution of statistically and mathematically complex forms of method and theory, simpler tools are sometimes more economical and elegant, and potentially more rigorous. Scholars should carefully evaluate the strengths and weaknesses of these diverse tools in light of existing knowledge about the topic under study, and with reference to broader shared standards for descriptive and causal inference and for refining theory. This eclectic approach is the most promising avenue for productive decisions about research design.

2

The Quest for Standards: King, Keohane, and Verba's *Designing Social Inquiry*

David Collier, Jason Seawright, and Gerardo L. Munck

Scholars turn to methodology for guidance in conducting research that is systematic, rigorous, and cumulative. *Designing Social Inquiry: Scientific Inference in Qualitative Research*, by Gary King, Robert O. Keohane, and Sidney Verba (hereafter KKV), has commanded wide attention because it forcefully and articulately provides such guidance. With clarity of exposition and many examples, the book presents an extended set of practical recommendations for the design and execution of research. In conjunction with KKV's goal of providing a new framework for qualitative research, the book offers an important synthesis of what we will call mainstream quantitative methods. KKV therefore constitutes a general statement about methodology, and this fact helps account for the wide attention it has deservedly received.

The present chapter provides an overview of KKV. We first introduce three fundamental ideas in KKV's view of methodology: (1) the criteria for scientific research; (2) the concept of inference—a term used in the title of the book and central to KKV's exposition; and (3) the assumptions that justify causal inference.

The second part of this chapter adopts a different approach to summarizing KKV's framework by presenting it in terms of a set of guidelines for conducting research. KKV does not explicitly synthesize its recommendations as an over-arching set of rules,[1] yet we believe these guidelines provide a summary that plays a constructive role in focusing the discussion.

Finally, the conclusion to the chapter anticipates the debate in the remainder of the present volume, noting both points of convergence and areas of substantial divergence vis-à-vis the perspective presented by KKV (see table 2.2 toward the end of this chapter).

In this summary of KKV's arguments, we occasionally provide examples of our own. At certain points, as with the discussion of conditional independence, we offer a somewhat more elaborate presentation than KKV, given that these are topics to which we return later in the present volume. Nevertheless, the intent of the chapter, except for the conclusion, is to present KKV's framework.

SCIENTIFIC RESEARCH, INFERENCE, AND ASSUMPTIONS

Three central components of KKV are its treatment of scientific research, inference, and assumptions. In relation to prior discussions of these topics, KKV's goal is not primarily to present new ideas. However, as a set of recommendations designed specifically for qualitative researchers, KKV's treatment of these topics is innovative and deserves careful attention.

Scientific Research

KKV argues that social science ought to be good *science*. To that end, the book presents a careful definition of what makes research scientific. Some readers may find KKV's insistence on the idea of science jarring and this framing of goals too narrow. Yet these goals are in fact of broad relevance. How, then, does KKV define scientific research? First of all, such research always seeks to make *inferences*, "attempting to infer beyond the immediate data to something broader that is not directly observed" (8). The idea of inference is of such importance in KKV's methodological approach that it is explored in detail in the next section of this chapter.

Next, scientific research makes its procedures *public*. Researchers should report how they select cases, gather data, and perform analysis. This is necessary if the scholarly community is to judge the quality of the research and the plausibility of its conclusions. If analysts do not report how they conduct their research, then "[w]e cannot evaluate the principles of selection that were used to record observations, the ways in which observations were processed, and the logic by which conclusions were drawn" (8).

Moreover, researchers must view their conclusions as inherently *uncertain*. "A researcher who fails to face the issue of uncertainty directly is either asserting that he or she knows everything perfectly or that he or she has no idea how certain or uncertain the results are" (KKV 9). Neither measurement nor theory in the social sciences is ever perfect and complete. According to KKV, scientific research requires scholars to acknowledge this fact and to estimate the degree of uncertainty in their inferences.

The final characteristic of scientific research is that findings are judged in light of the *method* employed, because, as KKV (9) argues, the content of science is the method. In other words, scientific findings should not be accepted or rejected according to the authority of the researcher, or in light of whether they correspond to the particular results preferred by a given investigator. Rather, the credibility of the methods employed should be a central criterion in evaluating research findings.

These criteria present a simple, reasonably straightforward basis for distinguishing scientific research from other kinds of intellectual pursuits.

Inference

The idea of inference is a major component of KKV's methodological framework. Indeed, KKV views "inference"—in the sense of drawing larger conclusions on the basis of specific observations—as a foundation of social science. The book treats inference in broad terms, stating that "[i]nference, whether descriptive or causal, quantitative or qualitative, is the ultimate goal of all good social science" (34). KKV develops this idea in extended discussions of descriptive inference (chap. 2) and causal inference (chaps. 3–6).[2]

1. Munck's (1998) review essay on KKV was the first effort to summarize the book in terms of a complete set of rules. Subsequently, Epstein and King (2002) adopted this approach in their long essay, "The Rules for Inference." The recommendations in their essay are quite similar to those in KKV, except that they give more attention to the tasks of defining the universe of cases and building a tradition of publicly available data sets.

2. The relation between description and explanation is complex, as is clear in the discussion below of the contrast between the systematic and random components of phenomena. Even so, description versus explanation remains a fundamental heuristic distinction, both in KKV and in the present volume. At the simplest level, description addresses the question of "what?" and explanation addresses the question of "why?" Also, as noted in chapter 1 above (15–16 this volume), although the ideas of descriptive and causal "inference" may seem nonstandard to some readers, they can be viewed as convenient labels for the ubiquitous research task of moving from specific observations to more general ideas.

Descriptive Inference

In KKV's view, descriptive inference entails three tasks. First, it encompasses the idea of generalizing from a sample to a universe of cases, as routinely occurs in public opinion research. The researcher establishes the universe and the sample, analyzes the cases included in the sample, and makes inferences about the universe on the basis of the sample (e.g., KKV 70–71).

Second, descriptive inference encompasses inferences from observations to concepts. Analysts are rarely interested in reporting raw facts. Rather, they seek to describe political institutions, social structures, ideologies, and other complex phenomena. As conceptualized by social scientists, these phenomena are never directly observable: no one has ever *seen* an entire "social structure." Scholars observe certain facts, often at only one point in time, that are relevant to the complex idea of a social structure, that presumably persists over time. They must therefore make inferences from these particular facts to the broader idea of a social structure. Hence, "[d]escriptive inference is the process of understanding an unobserved phenomenon on the basis of a set of observations" (KKV 55).

A third aspect of descriptive inference, which is strongly emphasized by KKV, is the more complex issue of separating the "systematic" and the "random" components of any phenomenon. KKV (43) argues that descriptive inference inherently involves simplification, and one productive form of simplification can be to focus description on the systematic component of the phenomenon that the researcher seeks to explain.

Although in practice the separation of the systematic and random components may be difficult to achieve, it is important to see why this can be a useful idea. The rationale for this distinction depends on making a link between descriptive inference and causal inference. The systematic component of a phenomenon is understood as that which is explained by an accepted causal model; the random component is that which is not (60, 63).[3]

KKV points to alternative views of this random component. In one view, the world is inherently probabilistic. Thus, "[r]andom variation exists in nature and [in] the social and political worlds and can never be eliminated" (59). Another view rejects the idea that the world is inherently probabilistic, contending instead that what appears to be random "is only that portion of the world for which we have no explanation" (59). In other words,

3. KKV presents this idea by taking as a point of departure the supposition that the researcher lacks any prior knowledge of causal patterns: "[W]e begin any analysis with all observations being the result of 'nonsystematic' forces. Our job is then to provide evidence that particular events or processes are the result of systematic forces" (60).

causation is deterministic, and what appears to be random is simply the facet of reality that is explained by variables not yet included in the relevant model, or is due to measurement error.

KKV illustrates this distinction with the example of fluctuations in the vote for a given party within a particular electoral district (55). The vote for this party may vary over time in part due to factors that are truly random. Alternatively, it might vary due to specific events that are outside the conventional explanatory concerns of political scientists—for example, variations in the weather, or some accidental occurrence such as the use of ballots that voters find confusing. In either case, an analyst may wish to generate a description of the party's vote share from which these fluctuations are removed. A common way of accomplishing this is to take an average of the party's vote share across several elections, on the assumption that the random fluctuations will cancel one another out (58).

Of course, variation that falls outside the focus of one explanatory framework or theory may be a central concern for another theory. Correspondingly, a description based on a careful separation of systematic and random components that is well suited to one theory may be less appropriate to another theory. Notwithstanding this limitation, the possibility of such separation raises the important idea that analytically productive description may isolate that part of a phenomenon that we really seek to explain. More broadly, it serves as a useful reminder to researchers that the facts do not "speak for themselves." Rather, they are interpreted from some theoretical perspective.

KKV considers description a fundamental part of the social scientific enterprise, and the book warns that in research contexts where causal inference is unusually difficult, analysts should sometimes be satisfied with careful descriptive inference (44–45; also 34, 75 n. 1). Nonetheless, KKV pays greater attention to causal inference, arguing that the best description is organized as a collection of evidence that evaluates a causal claim (46–49). It is therefore hardly surprising that the larger part of KKV's focus is on research designed to test causal hypotheses.

Causal Inference

KKV's treatment of causation follows in the tradition of Neyman (1990 [1923]), Hodges and Lehmann (1964), Rubin (1974, 1978), and Holland (1986), who developed a counterfactual understanding of causation.[4] According to this account, the idea that "*X* causes *Y*" in any given unit of analysis raises the hypothetical question of how the outcome on *Y* would have differed if *X* had not occurred in that unit. Given that it is impossible

4. This approach is reviewed in more detail on 44–49 below, in the discussion of conditional independence.

to observe both the occurrence and nonoccurrence of X for any given unit at one point in time, causal inference involves comparing something that did occur with something that did not occur. This is the source of what Holland and KKV (79, 82) call the "fundamental problem of causal inference," that is, the problem that causal inference implicitly depends on a comparison with something that did not occur.

Using this counterfactual view of causation, KKV (76–82) hypothetically posits the existence of two parallel universes, exactly alike in every way except for one. Taking the example of a dichotomous independent variable, we might find that in one of these two universes, the unit being studied has a positive score on the hypothesized cause and thus receives the "treatment." In the other universe, the hypothesized cause does not occur in the unit being studied: it is a "control." The causal effect of the explanatory variable is the difference in the outcome between the two parallel universes.

This definition helps researchers in reasoning about causation as an abstract concept. It serves to clarify why scholars do indeed face a fundamental problem of causal inference: out of the two observations of a given case needed to directly assess a causal effect, researchers can, in the real world, only make one. Either a case gets the treatment, or it does not. In observational studies, analysts cannot even choose which of these two universes to observe, because they cannot manipulate the independent variable. Some kind of inference is necessary to overcome this fundamental problem; hence, causal inference is the only way to appraise causation. When this understanding of causation is applied in observational studies, analysts seek to approximate these hypothetical comparisons through real-world comparisons among observed cases. A central component of KKV's advice focuses on how to carry out these real-world comparisons.

Making Inferences: Quantitative Tools and Analytic Goals

KKV's recommendations can usefully be summarized in terms of the tools the book proposes, and in light of the goals it seeks to pursue with these tools. KKV draws heavily on regression analysis, econometrics, and other standard techniques of quantitative methodology (table 2.1). These include basic methods for describing quantitative data, such as means and variances, and, very crucially, the use of regression analysis for causal assessment. Regression analysis in the social sciences relies on quantitative tools of parameter estimation (i.e., estimating the coefficients associated with each independent variable), and generally also on significance tests (which address uncertainty due to sampling error or other forms of randomness in the model). In discussing causal inference from a regression perspective, KKV implicitly draws on these statistical techniques. Increasing the number of observations is frequently recommended as a basic tool for enhancing inferential leverage in empirical tests (i.e., achieving higher levels of statistical significance). Finally, KKV employs tools of probability theory, such as expected value and variance of the estimator. KKV's tools are designed for use with quantitative data, and the book's fundamental advice to qualitative analysts is to use procedures in their own research that make a parallel contribution to valid inference. Although the chapters below debate whether it is in fact possible to implement this recommendation, there is not the slightest question that this advice has extended the analytic horizon of qualitative researchers.

Table 2.1. Quantitative Tools Employed in *Designing Social Inquiry*

Tools	*Comments*
Means and Variances	Means and variances are the basis for other tools discussed below.
Regression Analysis	Regression analysis is KKV's basic tool for causal inference from empirical data (e.g., 95–97, 121–22, 130–32, 168–72). Parameter estimation and significance tests, as used in regression analysis, provide a major part of the statistical basis for KKV's discussion of causal inference.
Increasing the N	KKV repeatedly advocates increasing the number of observations as the best way to enhance the inferential leverage of empirical tests (e.g., 19, 23–24, 29–31, 46–49, 52, 67, 99, 117–18, 120–21, 123, chap. 6).
Probability Theory	Many of KKV's "Formal Analysis" text boxes (e.g., 97–99, 166–68, 184–85) evaluate the variance and bias of different estimators by applying tools of probability theory.

With regard to KKV's broader analytic agenda, within the framework of what we will call the book's "overarching goals" of achieving valid descriptive and causal inference, a central focus is on "intermediate goals," which provide a justification for the use of these quantitative tools in pursuit of the overarching goals. Two major intermediate goals are avoiding bias and minimizing the variance of estimators in order to achieve higher levels of statistical significance.[5] Analysts should seek to avoid bias, potential sources of which include systematic measurement error (155–57), selection procedures that are correlated with the dependent variable—including proce-

5. KKV uses the term "efficiency" to refer to the goal of minimizing estimator variance. However, the technical definition of efficiency in statistics is somewhat different, so we have used this more general phrase in the text. KKV does not explicitly defend its preference for lower-variance estimators in terms of statistical significance, but this is the most obvious interpretation.

dures that may cause selection bias (128–37), missing explanatory variables (168–76), and endogeneity, that is, the problem that the outcome variable or the error term influences the explanatory variables (185–96). Researchers should also minimize the variance of their estimators by excluding irrelevant explanatory variables (182–85) and by reducing nonsystematic measurement error (157–68). In addition to reducing variance, which maximizes the precision of the inferences that can be drawn from a given data set, KKV recommends increasing leverage by creating data sets that have greater inferential power. Additional intermediate goals are summarized in the guidelines below. KKV thus builds on the tools of mainstream quantitative methods to propose a series of procedures for achieving valid inference in qualitative research.

KKV does not simply present these tools and goals in a mechanical fashion, but at various points considers how some of them intersect with concerns that derive from the qualitative tradition. For example, although researchers can avoid some types of selection bias through random sampling, the book recognizes that in small-N research, random sampling may create as many problems as it solves (124–28). Within the framework of nonrandom sampling, KKV is careful to avoid a piece of clichéd advice that is often invoked in discussions of selection bias—that is, "do *not* select on the dependent variable." Instead, KKV argues that scholars who, for good reason, avoid random sampling and *do* select on the dependent variable should choose cases to reflect the full range of variation on that variable (141).[6]

Assumptions

KKV discusses the assumptions routinely employed to justify causal inference. Some scholars may think of these as "quantitative" or "statistical" assumptions. However, KKV (93) argues that these assumptions should not be understood narrowly as relevant only for quantitative analysis. Rather, assumptions are important for any study, whether quantitative or qualitative, that seeks to make the kind of inferences discussed in the previous section.

KKV urges researchers to "make the substantive implications of [their assumptions] extremely clear and visible to readers" (91). This advice is valuable because inferences depend on the assumptions that produce them, and a somewhat different set of assumptions can generate radically divergent inferences. This is one of the reasons why—as noted in chapter 1 above—it is hard to do really good quantitative research, just as it is hard

6. This corresponds to the second meaning of "selecting on the dependent variable" discussed in the glossary.

to do really good qualitative research. KKV consequently advises researchers to justify their assumptions with theory and empirical evidence to the greatest extent possible (91). Yet KKV recognizes that it is often difficult to establish such justifications (93, 95).

Causal homogeneity, independence of observations,[7] and conditional independence are three major assumptions that KKV's authors view as essential for causal inference.[8] These assumptions focus researchers' attention on three interrelated tasks: analyzing an appropriate set of cases; considering how cases and observations can influence each other in a way that may affect causal inference; and selecting variables appropriately and modeling the relations among them.

Causal Homogeneity

The assumption of causal homogeneity[9] states that "all units with the same value of the explanatory variables have the same expected value of the dependent variable" (KKV 91). In other words, the outcomes for all the cases in the analysis must be produced by one causal model; after controlling for the values of the included independent variables, every case must have the same expected value on the dependent variable.[10]

Discussions of causal homogeneity are motivated by the concern that a given form of a causal model may only be appropriate to a particular domain of cases. If the model is extended to further cases, the researcher may have to make it more complex to accommodate distinctive causal features of those cases. Hence, this assumption is concerned with the relation between our causal ideas and the cases on which we focus.

In the statistical literature on causation (e.g., Rubin 1974; Holland 1986), a stronger version of the causal homogeneity assumption is presented, which Rubin and Holland call "unit homogeneity." According to this version of the assumption, different units are presumed to be *fully iden-*

7. This assumption is not treated in the same pages as the other two (KKV 91–97), yet it is likewise important (222–23).

8. We would add that somewhat modified versions of these assumptions do also permit causal inference. For example, independence of observations can be weakened, as in time-series analysis, where autocorrelation often arises. However, even the modified assumptions must, in fact, have the same basic properties as the assumptions discussed here.

9. KKV refers to this assumption as "unit homogeneity," as we explain below.

10. Two points should be made here. First, the "expected value" refers not to the value that one should anticipate for every case being analyzed, but rather to the average value across many hypothetical replications of each case. Second, KKV notes that one way to meet the causal homogeneity assumption is through the related assumption of "constant causal effects" (92–93).

tical to each other in all relevant respects except for the values of the main independent variable. This strong version is sufficient to allow causal inference without the assumption of conditional independence discussed below, but it is extremely unlikely that this strong homogeneity assumption will ever hold in the social sciences.

However, the weaker version of causal homogeneity that we discuss in this section, which allows units to differ from each other but requires that the causal parameters in the analyst's model be constant across all units, is more plausible and plays an important role in causal inference.

Though KKV occasionally makes reference to the stronger version of this assumption,[11] much of its discussion invokes the weaker version.[12] KKV refers to both versions of this assumption as "unit homogeneity." However, in labeling the weaker version of the assumption, which is much more central to KKV's overall framework, we find the term "causal homogeneity" more useful, both because it distinguishes this concept from the more rigorous standard of unit homogeneity and because it calls more explicit attention to the need for all cases to share the same causal model.

Specifically, if the causal homogeneity assumption is not met, and a researcher analyzes the data as if it were, the inference will be a misleading average that lumps together differences among subgroups of cases. This average may not adequately represent the pattern of causation in any given case. For example, it has been argued that among advanced industrial countries, in some national contexts the more highly paid workers are more class conscious, whereas in other national contexts they are less class conscious

11. KKV (91) defines unit homogeneity as being met if "the expected values of the dependent variables from each unit are the same when our explanatory variable takes on a particular value" (italics omitted). In this quote, the reference to multiple dependent variables for each unit invokes the Rubin-Holland framework for causality, and this clearly should be read as a reference to the strong version of unit homogeneity.

12. KKV (91) alternatively defines unit homogeneity as "the assumption that all units with the same value of the explanatory variables have the same expected value of the dependent variable." This statement, which refers only to the observed value of the dependent variable for each unit, does not invoke more complex statistical ideas of causation. Therefore, it would seem that it should be read as referring to the weaker version of unit homogeneity, involving constancy of causal parameters. This weaker version is also more compatible with KKV's (93) claim that "[t]he notion of unit homogeneity . . . lies at the base of all scientific research." In the Rubin-Holland framework, much scientific research specifically does not employ the unit homogeneity assumption, turning instead to alternatives such as randomization, conditional independence, and "ignorable treatment assignment." Hence, KKV's statement should be read as referring to the weaker assumption, and we therefore use the label "causal homogeneity" in discussing their arguments.

(Przeworski and Teune 1970: 26). If researchers simply average these two findings, they may find no relationship, resulting in a misleading conclusion. The appropriate solution would be to analyze the two groups of countries separately. Researchers would thus address causal heterogeneity by recognizing that causal processes are different between the two groups of countries, and by assuming that they are similar within each group. In regression analysis, this can sometimes be accomplished by introducing an interaction term that includes a dummy variable. In qualitative comparison, separate comparisons can be employed for the two groups. The fact that causal heterogeneity can thus be overcome by using a more complex model underscores a key point: causal homogeneity is not simply a property of the data, but of the data in relation to a particular causal model.

Independence of Observations

Another assumption concerns the independence of observations, that is, the idea that for each observation, the value of a particular variable is not influenced by its value in other observations and therefore provides new information about the phenomenon in question (222–23).[13] If independence of observations is not met, this does not necessarily bias the causal inference. However, it does reduce the amount of new evidence gained from each additional observation, thereby increasing the variance associated with an inference.

For some readers, a familiar alternative label for this assumption, which is appropriate for discussing cross-sectional analysis, is "independence of cases." However, this same assumption plays a major role in time-series analysis, in which the researcher analyzes multiple observations over time for each "case." Hence, the broader idea of independence of multiple observations for the same case becomes a central issue, and it is therefore useful to employ this more general label.

An example of this problem in time-series analysis is found in the literature on advanced industrial countries that explores the impact of corporatism and partisan control of government on economic growth. Scholars who had been working with an N of twelve to fifteen countries sought to achieve a major increase in the N by combining cross-sectional and time-series analysis, focusing on the period 1967–1984 (Alvarez, Garrett, and Lange 1991). However, subsequent research argued that prior results had been based on an incorrect assumption about the independence of observations. Consequently, the estimates of standard errors were too low, yielding excessive confidence in the conclusions. Revised estimates, based on a recognition of interdependence among observations—both among coun-

13. Unlike the other two assumptions discussed in this chapter, the assumption of independence of observations is also important for descriptive inference.

tries and within countries over time—supported some of the findings of the 1991 study, but cast doubt on others (Beck et al. 1993; Beck and Katz 1995; Kittel 1999).

Of course, the nonindependence of observations can also be viewed *not* as a methodological problem, but as a substantive topic—that is, as causation that occurs through processes of diffusion. However, within the framework of most work in regression analysis, it is indeed a methodological problem.

Conditional Independence

KKV's final major prerequisite for causal inference with observational data is the assumption of conditional independence, or, to give it a more complete name, conditional independence of assignment and outcome. We present this assumption by first returning to the counterfactual definition of causation noted above, from which the idea of conditional independence emerges, and then by offering two examples to make clear the importance of this assumption. Our presentation here will be more detailed than for the other two assumptions, given that this third assumption is particularly important to the discussion later in the present volume (172–177).

According to the counterfactual understanding of causation, causal inference consists of comparing (a) the value of the outcome variable (Y_t with "*t*" for treatment) for a particular case when that case is exposed to a treatment, with (b) the value of the outcome variable (Y_c with "*c*" for control)[14] for the same case when that case is not exposed to the treatment. Y_t and Y_c are thus two different variables that reflect the outcomes a case will experience on the dependent variable, according to whether the independent variable, conceptualized as an experimental treatment, is present or absent.[15]

The causal effect of the treatment for a given case is the difference between the two variables for the case: $Y_t - Y_c$. However, to restate the fundamental problem of causal inference discussed above, it is impossible to simultaneously observe Y_t and Y_c for any particular case. The value of one variable may be observed, but the value of the other is necessarily hypothetical. Consequently, it is impossible to compute $Y_t - Y_c$. Hence, in practice, causal inference seeks to replicate this hypothetical comparison by making real-world comparisons across (hopefully) similar units, some of which are exposed to the treatment and some of which are not.

When a real-world comparison is employed, the quality of the resulting causal inference depends on how cases are "assigned" to the treatment group and to the control group. Two issues are important here. First, a question of terminology: In observational studies, researchers do not actually assign cases to treatment and control groups. However, what we refer to as assignment does take place; it is carried out by social and political processes over which the researcher usually has no control.

The second issue, which is vital to the quality of causal inference, concerns the relationship between the assignment process and the outcome variables, Y_t and Y_c. The key question here is whether the cases are assigned in such a way that those in the treatment category have the same average values on both Y_t and Y_c as the cases in the control category. In other words, is the average of Y_t across the cases exposed to the treatment equal to the average of Y_t across the cases in the control group? Is this also true for Y_c?

If the answers to these questions are "yes," then the standard of independence has been met,[16] and the researcher will be able to make a good inference about the causal effect of the treatment by comparing the observed Y_t among the cases given the treatment with the observed Y_c among the cases assigned to the control. The underlying logic here is that, if independence of assignment holds, any difference between the treatment group and the control group must be due to the treatment—because all other relevant factors are balanced between the two groups. If, on the other hand, cases are assigned in such a way that those in the treatment group tend to have a different Y_t or Y_c than the cases in the control group, then causal inference will be biased. For example, if cases with a high value of Y_t are more likely to enter the treatment group than cases with a lower value of Y_t, the researcher will probably overestimate the causal effect of the treatment.

Independence of assignment is a strong condition, and it is rarely plausible in an observational study. Observational studies often employ an assumption of *conditional* independence, which serves to justify causal inference even though the treatment and control groups initially do not

14. We follow here the Rubin-Holland notation of "*t*" and "*c*," which is also employed in chapter 13 below. In chapter 3 below, where Brady presents his direct commentary on KKV, he follows the book's notation, which is based on KKV's running example: "*i*" for "incumbent" and "*n*" for "nonincumbent."

15. In this discussion, the independent variable may be dichotomous; alternatively, the treatment and control may reflect two different values on a continuous variable.

16. To be more precise, what is discussed here as independence is *mean* independence. Likewise, conditional independence as discussed here is actually *mean* conditional independence. For a discussion of these distinctions, see Stone (1993). Finally, the text above neglects two important, although somewhat narrow, technical issues: (a) whether there is a broader population from which the cases under investigation are a sample; and (b) whether the *expected* means of Y_t and Y_c, rather than the observed means, are in fact equal. The equality of the expected means is actually the key condition for mean independence and, if control variables have been introduced, for mean conditional independence.

have the same hypothetical average values on Y_t and Y_c. Suppose a variable (which we shall call Z) identifies subgroups of cases, within which independence of assignment does hold, and among which it does not hold. Then controlling for Z by comparing Y_t and Y_c within subgroups allows researchers to make unbiased inferences from the observational data. By stratifying in this manner, the standard of conditional independence is met.[17] In fact, because Y_t and Y_c cannot both be directly observed, the researcher never knows with certainty that their average values are equal. But in principle, the introduction of the appropriate control can make them equal, and hence yield conditional independence. In practice, achieving appropriate statistical control may involve more than one control variable (Z_1 to Z_n), and multivariate techniques are needed to introduce these multiple controls. For convenience, we will use the label Z to refer to one or more controls.

Given the importance of introducing control variables, the two words in this label, "conditional independence," thus bring together two essential ideas. (a) It is best for inference that assignment to the treatment and control groups be independent of the two outcome variables Y_t and Y_c. Correspondingly, the full name of the assumption is "conditional independence of assignment and outcome." (b) When independence does not hold, researchers can, in principle if not in practice, make inferences *as if* assignment were independent of Y_t and Y_c by statistically controlling for, or "conditioning" on, Z.

Conditional independence can be established if the appropriate statistical controls are introduced, removing the effect of an assignment process that does not meet the standard of independence. The assumption of conditional independence is thus addressed by employing with observational data the procedure of *statistical* control, as a substitute for the *experimental* control that is achieved through random assignment.

The effort by scholars to satisfy conditional independence by introducing the appropriate control can be illustrated with a well-known example of spurious correlation. In the United States, political participation is lower for African Americans and Latinos than for whites. In other words, if we hypothetically think of "nonwhite" as the treatment condition, and "white" as the control condition, individuals "assigned" to be African American and Latino have an average rate of participation, or average Y_t, that is lower than

17. Regression analysis depends on related assumptions about causation, such as the specification assumption discussed in chapter 9. For most purposes, these assumptions may be seen as similar, in that they both focus attention on the potential problem of missing variable bias. However, it is important to remember that alternative analytic tools (e.g., regression versus stratification) depend on assumptions that sometimes differ in important ways.

the average rate of participation, or average Y_c, among people "assigned" to be white. The lower participation rate of the first two groups provides an appropriate basis for descriptive inference (i.e., describing their levels of participation), but it is problematic as a basis for causal inference. It does not necessarily follow that being African American or Latino causes citizens to participate less. Rather, membership in these two groups is correlated with other factors, such as education and income, that could explain lower participation rates. These other factors serve the role of identifying salient subgroups among the cases; hence these other factors may be equivalent to the variable Z in the discussion above. When these other factors are controlled for, thus making it more plausible that conditional independence is satisfied, "neither being African American nor being Latino has a direct impact" on participation (Verba, Schlozman, and Brady 1995: 442).

In other words, after conditioning on—that is, controlling for—Z, these authors conclude that the average value of Y_t is in fact about the same as the average value of Y_c. It is not being African American or Latino that reduces the political activity of individuals within these groups. That apparent causal relation is spurious, and other factors such as low education or low income account for the lower rate of participation. Once the effect of these other factors is removed statistically, the underlying causal relationship emerges.

A second example illustrates the point that the conditional independence assumption is hard to meet when analysts cannot identify, or cannot measure, the variable or set of variables that must be controlled for. Consider the question of whether the *size* of revolutionary movements (independent variable) affects their *success* in overthrowing an existing regime (dependent variable). As Goldstone (1991: 137) emphasizes, because the personal cost of participating in an unsuccessful revolutionary movement can be high, many individuals will only join revolutionary movements that are seen as having at least some probability of defeating the regime. This evaluation obviously depends on the perceived strength of both the revolutionary movement and the regime. Specifically, the probability that a revolutionary movement will grow in size (which corresponds to the treatment) depends in part on the particular characteristics of the national regime that individuals evaluate in judging the relative strength of that regime. Yet the strength of the regime *also* plays a key, direct role in influencing the likelihood that the regime will fall, which is the outcome being explained.

Thus, due to these regime characteristics, those countries most susceptible to revolution may be most likely to face large revolutionary movements, and are in effect assigned to the treatment group. In this discussion, characteristics of the national regime are an instance of the variable Z above. Contrasts in these characteristics group together regimes that differ in the degree to which they are perceived as weak. Perceptions of weakness are, in turn, correlated: (a) with the likelihood of regime collapse, given a strong insurgent movement, or Y_t; and (b) with potential insurgents' decisions to rebel, which, when aggregated, constitutes the treatment. Unless these regime characteristics are included in the analysis and controlled for, researchers will overestimate the importance of popular participation in revolutionary opposition movements for causing regime collapse—given that greater popular participation is more likely when the chance of regime collapse is high.[18]

To meet the assumption of conditional independence, the researcher would need to collect data on these characteristics that adequately capture their role in influencing both the size of revolutionary movements and the likelihood of regime collapse. Yet collecting these variables and adequately controlling for them is doubtless more difficult than it is for the education variable in the prior example. The researcher would have to collect enough information about regime characteristics to arrive at the same evaluations and judgments that potential revolutionaries make about the strength of the regime. Hence, the idea of conditional independence is crucial here, but it is difficult to meet this assumption.

Overall, the idea of conditional independence uses the counterfactual definition of causation to provide a logical framework for reasoning about the critical task of controlling for rival explanations in causal inference.

To summarize the discussion of these three assumptions, KKV's goal is to underscore the idea that they are important to all researchers, and not just quantitative analysts. In all observational studies, causal inference never relies exclusively on the actual data, but also on assumptions about the political and social processes we are studying. It is evident that not only

18. This problem can arise regardless of whether the researcher takes a more structural or a more actor-centered view of revolution. One interpretation of this causal pattern could be that the perception of these revolutionary actors is an intervening variable that links these regime characteristics to the revolutionary outcome, involving an actor-centered and potentially "agental" explanatory perspective. Another interpretation views regime characteristics as direct, structural causes of revolution. For example, according to Chehabi and Linz (1998), under sultanistic regimes a poorly institutionalized, personalistic military is a critical structural factor in regime breakdown. Although the perception of the military on the part of revolutionary and regime actors may have some importance, this weakness of the military is seen, in its own right, as a critical causal factor. The point here is not to adjudicate between a structural and an actor-centered perspective, but rather to show that, from either perspective, failure to satisfy conditional independence may interfere with causal inference. Whether the structural weakness in the military causes revolution directly, or primarily through the perceptions of state and popular actors, varying degrees of regime strength can still confound our attempts to estimate the impact of popular participation on revolution.

KKV's discussion of these assumptions, but also the book's treatment of inference and the definition of scientific research, involve a perspective that is far more familiar to quantitative than to qualitative researchers. However, KKV is strongly committed to the idea that these issues are of equal relevance to both traditions. Even a scholar who disagrees with KKV must recognize that the book makes a fundamental contribution by pushing a broader range of researchers to grapple with these questions.

GUIDELINES: SUMMARIZING KKV'S FRAMEWORK

This section adopts a different approach to synthesizing KKV by presenting many of the book's more specific methodological recommendations as a set of guidelines. These guidelines are largely concerned with what we refer to in chapter 1 as intermediate goals, focusing on procedures for linking specific quantitative tools to the overarching goals of valid descriptive and causal inference. The guidelines help to make clear how KKV's broad ideas, summarized in the present chapter, inform the book's treatment of specific decisions about research design.

We organize the guidelines in terms of a research cycle (figure 2.1): defining the problem, specifying the theory, selecting cases and observations, carrying out descriptive and causal inference, and retesting and reformulating the theory. The final step completes this cycle by bringing the researcher back to the step of theory specification, and potentially also to redefining the research problem (see dashed arrow in the figure). Although research routinely moves through a series of ordered steps such as this, what is learned at each step certainly may lead to revisiting prior steps or jumping forward to subsequent steps. Hence, one could in fact place many more arrows in the diagram.

These guidelines are, of course, our summary of KKV's arguments. KKV makes periodic reference to "rules" for research (e.g., 6–7, 9), and the book presents five specific rules for constructing causal theories (99–114). However, the book does not synthesize its recommendations in terms of an overall set of rules or guidelines.[19] Each of the guidelines presented below is introduced as a brief, self-explanatory phrase. For some of the guidelines, we spell out the idea in greater detail, often drawing on quotations from KKV. In all cases, specific page references are provided.

KKV states that "[a]ny meaningful rules admit of exceptions. . . . We seek not dogma, but disciplined thought" (7). Correspondingly, we do not want to give the impression that KKV's framework consists of rigid rules. Rather,

19. See note 1 above.

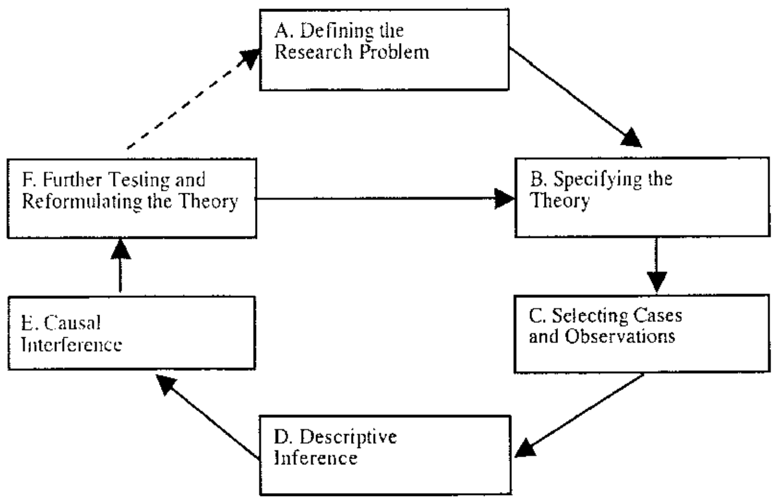

Figure 2.1 Steps in the Research Cycle: A Framework for Summarizing *Designing Social Inquiry*

Note: Solid arrows show the main links among steps in the cycle. Choices made at any one step can, of course, potentially affect any other step. This is reflected, for example, by the placement of a dashed line from F to A, in addition to the solid line from F to B.

we seek to bring together systematically the large number of specific recommendations offered by the book, as a means of demonstrating both the scope of these recommendations, and KKV's relative emphasis on different methodological issues.

A. Defining the Research Problem

1. *Address a problem that is important in the real world* (15).
2. *Contribute to a scholarly literature.* Contribute to "an identifiable scholarly literature by increasing the collective ability to construct verified scientific explanations of some aspect of the world" (15, 16–17).[20]
3. *Modify or abandon a topic that cannot be refined into a research project that permits valid inference* (18).

B. Specifying the Theory

4. *Construct falsifiable theories.* "[C]hoose theories that could be wrong" (19; also 100).
   a. *Strengthen falsifiability by choosing a theory that maximizes observable implications* (19).

20. The italics in many quotations have been omitted.

   b. *Strengthen falsifiability by being concrete.* "Theories that are stated precisely and make specific predictions can be shown more easily to be wrong and are therefore better" (20, 109–12).
5. *Build theories that are logically consistent.* "[I]f two or more parts of a theory generate hypotheses that contradict one another, then no evidence from the empirical world can uphold the theory" (105).
6. *Increase leverage by explaining more with less.* Explain "as much as possible with as little as possible" (29).
   a. *Increase leverage through parsimony.* "[M]aximize leverage by limiting the number of explanatory variables" (123).
   b. *Increase leverage by explaining more observable outcomes.* "State theories in as encompassing [a way] as feasible" (113), and "list all possible observable implications of [the main] hypothesis that might be observed in [the] data or in other data" (30).

C. Selecting Cases and Observations

7. *Distinguish between cases and observations.* "Cases" are understood as the broader units, that is, the broader research settings or sites within which analysis is conducted; "observations" are pieces of data, drawn from those research sites, that form the direct basis for descriptive and causal inference (52–53, 117–18, 217–18).
8. *Focus on the range of variation relevant to the theory.* Select cases among which the dependent variable in fact exhibits "the variation [researchers] wish to explain" (108). It is thus important not merely to have variation on the dependent variable, but that this variation capture the contrasts addressed by the theory.
9. *Construct a determinate, rather than an indeterminate, research design by including a sufficient number of observations.*[21] Avoid an indeterminate research design from which "virtually nothing can be learned about the causal hypotheses" because the researcher has "more inferences to make than implications observed" (118, 119; also 116, 120, 178–79, 213–17, 228). In the face of an insufficient number of observations, scholars can:
   a. *Address indeterminacy by increasing the number of observations—either through changing the dependent variable, or through focusing on subunits* (24, 47, 120, 217–28).
   b. *Address indeterminacy by gaining leverage from strong theory.* If the number of observations is insufficient, "limited progress in understanding causal issues is nevertheless possible, *if* the theo-

21. A determinate research design also requires the absence of perfect multicollinearity. This likewise involves the issue of having enough observations, in that a sufficiently large N can help overcome multicollinearity. See no. 30 below.

retical issues with which [researchers] are concerned are posed with sufficient clarity and linked to appropriate observable implications" (179).

c. *Address indeterminacy by situating observations within a larger research program.* Even "a single observation can be useful for evaluating causal explanations if it is part of a research program. If there are other single observations, perhaps gathered by other researchers, against which it can be compared, it is no longer a single observation" (211, 129 n. 6).

10. *Seek causal homogeneity.* Causal homogeneity[22] is "the assumption that all units with the same value of the explanatory variables have the same expected value of the dependent variable" (91, 116).

11. *Avoid selection bias.* Selection bias poses important "dangers" (116), in that it can invalidate both causal inference (129–32) and descriptive inference (135). One important source of such bias is the failure of the sample to reflect the full range of variation on the dependent variable. The random selection of cases is a standard means for avoiding important forms of selection bias, yet in small-N research this may not be appropriate (126).

12. *Select cases nonrandomly in small-N analysis.* Random selection in small-N research can too easily fail to capture the full range of variation on the variables of interest. "Usually, selection must be done in an *intentional* fashion, consistent with . . . research objectives and strategy" (139). This recommendation is relevant both for descriptive (135) and causal (129–32) inference. With reference to causal inference, KKV suggests the following standards for nonrandom selection:

a. *Avoid selecting a set of observations in which either the independent or dependent variable is constant.* "[T]he causal effect of an explanatory variable that does not vary cannot be assessed . . ." (146). Researchers "can also learn nothing about a causal effect from a study which selects observations so that the dependent variable does not vary" (147; also 108–9, 129, 148–49). "The cases of extreme selection bias—where there is by design no variation on the dependent variable—are easy to deal with: avoid them!" (130).

i. *In selecting observations on either the independent or dependent variable, ensure that these observations encompass sufficient variation on this variable.* For example, when selecting on the dependent variable, "select observations with particularly high and particularly low values . . . " (129, 141, 147–49).

22. Regarding definitions of causal homogeneity versus unit homogeneity, see the glossary.

ii. *To address the problem of a no-variance design, seek variance by situating observations within a larger research program* (146–47).

b. *Selecting simultaneously on both the independent and dependent variables can pose a grave problem.* "The most egregious error is to select observations in which the explanatory and dependent variables vary together in ways that are known to be consistent with the hypothesis that the research purports to test" (142).

13. *If observations are not independent from one another, recognize that this reduces the certainty of the findings; researchers may also address the causes of this interdependence.* When observations are not fully independent of each other, "each new [observation] does not bring as much new information to bear on the problem as it would if the observations were independent of one another. . . . [W]hen dealing with partially dependent observations . . . be careful not to overstate the certainty of the conclusions. . . . [C]arefully analyze the reasons for the dependence among the observations" (222).

D. Descriptive Inference

14. *Description requires inference.* Description in social science research must be understood not as the process of collecting unmediated facts, but rather as involving inferences from observations to the broader ideas and comparisons around which the research is organized (chap. 2).

15. *Recognize the similarity between quantitative or formal work and "interpretation," as compared to the full complexity of reality.* "[T]he difference between the amount of complexity in the world and that in the thickest of descriptions is still vastly larger than the difference between this thickest of descriptions and the most abstract quantitative or formal analysis" (43).

16. *Extract analytically relevant features from the uniqueness of cases* (42). "All phenomena, all events, are in some sense unique. . . . The real question . . . [is] whether the key features of social reality that we want to understand can be abstracted from a mass of facts" (42).

17. *Know the context.* "Where possible, analysts should simplify their descriptions only after they attain an understanding of the richness of history and culture. . . . [R]ich, unstructured knowledge of the historical and cultural context of the phenomena with which they want to deal in a simplified and scientific way is usually a requisite for avoiding simplifications that are simply wrong" (43).

18. *Good description is better than bad explanation.* In research contexts in which good causal inference is difficult, it may be preferable to stick

to carefully executed descriptive inference (44; also 34, 45, 75 n. 1, 178–79).

19. *Study observable concepts.* "[C]hoose observable, rather than unobservable, concepts wherever possible" (109). "Attempting to find empirical evidence of abstract, unmeasurable, and unobservable concepts will necessarily prove more difficult and less successful than for many imperfectly conceived specific and concrete concepts" (110).
20. *In general, avoid typologies and classifications, except as preliminary heuristic devices.* "[C]onstructs such as typologies, frameworks, and all manner of classifications, are useful as temporary devices [for] collecting data. . . . However, in general, we encourage researchers *not* to organize their data in this way" (48).
21. *Use valid indicators.* "Validity refers to measuring what we think we are measuring" (25). Among the issues that arise in striving for validity is the need to "use the measure that is most appropriate to [the researcher's] theoretical purposes" (153).
22. *Use reliable data-collection procedures that, if applied again, would produce the same data* (25).
23. *Estimate measurement error.* "Since all observation and measurement . . . is imprecise," researchers should "estimate the amount of [measurement] error . . . " (151); "qualitative researchers should offer uncertainty estimates in the form of carefully worded judgments about their observations" (152).
24. *Separate the systematic and random components of phenomena.* "[O]ne of the fundamental goals of [descriptive] inference is to distinguish the systematic component from the nonsystematic component of the phenomena" being studied (56). Thus, analytically productive description may seek to isolate the systematic component, as it is this component that researchers really seek to explain.

E. Causal Inference

25. *Causal assessment requires inference.* Causation is not observed directly. Rather, causation is inferred on the basis of data and assumptions (chap. 3).
26. *Demonstrate, to the extent possible, that the assumptions underlying causal inference are met in a given context of research.* Assumptions such as causal homogeneity, conditional independence, and the independence of observations "can and should be justified" to the greatest extent possible on the basis of insights derived from prior research and knowledge of the research setting (91).
27. *Use theory to select appropriate explanatory variables and avoid "data mining.*" "Without a theoretical model, [researchers] cannot decide which potential explanatory variables should be included in [the] analysis." "[W]ork toward a theoretically motivated model rather than 'data mining'. . . ." In other words, researchers should not simply run "regressions or qualitative analyses with whatever explanatory variables [they] can think of" (174).

28. *Avoid missing variable bias by including all relevant explanatory variables.* "[S]ystematically look for omitted control variables and consider whether they should be included in the analysis" (172). If a given variable is correlated with both the dependent variable and an explanatory variable, then failure to include it will bias the causal inference (170). The following three steps can help avoid missing variable bias:
    a. *First, list potentially relevant explanatory variables* (174).
    b. *Second, control for relevant explanatory variables* (174).
    c. *Third, in estimating the main causal effect, do not control for intervening variables.* "[I]n general, [researchers] should not control for an explanatory variable that is in part a consequence of [the] key explanatory variable" (174).
29. *Minimize the variance of estimators by excluding irrelevant variables.* Do not "collect information on every possible causal influence . . . " (182, italics omitted) because "[t]he inclusion of irrelevant variables can be very costly" (183). While the best solution to the problem of "many variables, small N" is to collect more observations, "if this is not possible, researchers are well-advised to identify irrelevant variables" (184) and exclude them from the analysis.
30. *Avoid an indeterminate research design due to multicollinearity.*[23] Avoid a research design in which two or more of the explanatory variables are so highly correlated that it is impossible to separate their causal effects (119). The proposed solution to this problem is to:
    a. *Address multicollinearity by collecting additional observations.* "[S]earch for observable implications at some other level of analysis" (123), which can give more leverage in differentiating the causal effects of highly correlated explanatory variables.
31. *Avoid endogeneity.* "A very common mistake is to choose a dependent variable which in fact causes changes in [the] explanatory variables. . . . [T]he easiest way to avoid [this mistake] is to choose explanatory variables that are clearly exogenous and dependent variables that are endogenous" (107–8; also 94, 185). Five solutions to endogeneity are:
    a. *Address endogeneity by careful selection of observations.* "[W]e can

23. A determinate research design also requires a sufficient number of observations. See guideline 9 above.

first translate a general concern about endogeneity into [a concern about] specific potential sources of omitted variable bias and then search for a subset of observations in which these sources of bias could not apply" (193).

b. *Address endogeneity by transforming it into an omitted variable problem.* "By transforming [a research] problem in this way, scholars [can] get a better handle on the problem since they [can] explicitly measure this omitted variable and control for it . . . " (190).

c. *Address endogeneity by disaggregating the dependent variable.* "[R]econceptualize the dependent variable as itself containing a dependent and an explanatory component. . . . The goal of this method of avoiding endogeneity bias is to identify and measure only the dependent component of [the] dependent variable" (188–89).

d. *Address endogeneity by disaggregating the explanatory variable.* "[D]ivide a potentially endogenous explanatory variable into two components: one that is clearly exogenous and one that is at least partly endogenous. . . ." Then use "only the exogenous portion of the explanatory variable in a causal analysis" (193).

e. *Address endogeneity by correcting the biased inference.* "[E]ven if [researchers] cannot avoid endogeneity bias, [they] can sometimes improve . . . inferences after the fact by estimating the degree of bias. At a minimum, this enables [them] to determine the direction of bias, perhaps providing an upper or lower bound on the correct estimate" (188).

32. *Estimate and, if possible, correct for selection bias.* "[I]f selection bias is unavoidable, [researchers] should analyze the problem and ascertain the direction and, if possible, the magnitude of the bias, then use this information to adjust [their] original estimates in the right direction" (133). If they "know there is bias but cannot determine its direction or magnitude . . . [researchers should] at least increase the level of uncertainty [they] use in describing [their] results" (199; also 128–37, 168–82).

F. Further Testing and Reformulating the Theory

33. *Report research procedures, thereby allowing other analysts to evaluate and replicate the findings.* "Only by reporting the study in sufficient detail so that it can be replicated is it possible to evaluate the procedures followed and methods used" (26; also 8, 23, 51).

34. *Test the theory with data other than that used to generate the theory* (46). The original data can be used to test a new implication of a theory, "as long as the implication does not 'come out of' the data but is a hypothesis independently suggested by the theory or a different data set" (30).

35. *The theory should generally not be reformulated after analyzing the data.* "Ad hoc adjustments in a theory that does not fit existing data must be used rarely . . . " (21).

a. *If the theory is reformulated by making it more restrictive, retest it with new data.* If a theory is modified after analyzing the data, researchers "can make the theory less restrictive (so that it covers a broader range of phenomena and is exposed to more opportunities for falsification), but [they] should not make it more restrictive without collecting new data to test the new version of the theory" (22, italics omitted).

ANTICIPATING THE DISCUSSION OF KKV'S FRAMEWORK

Subsequent chapters in the present volume provide alternative perspectives on quantitative and qualitative methods, making central reference to the framework offered by KKV. This final section of chapter 2 anticipates the assessment presented in the following chapters.[24] As can be seen in table 2.2, we organize the discussion with reference to specific guidelines. Some aspects of KKV's framework evoke agreement, whereas for others there is disagreement.

I. Areas of Convergence

a. *Broad Convergence.* The chapters in this volume strongly endorse the overall goal of developing shared standards for descriptive and causal inference. This convergence once again calls attention to the contribution made by KKV in focusing scholarly attention on such standards.

b. *Specific Points of Convergence.* Many of KKV's suggestions are not challenged or reevaluated. The recommendation to move beyond the uniqueness of cases by extracting analytically relevant features (guideline no. 16 above) articulates a fundamental priority in social science research. KKV's suggestion to distinguish between cases and observations (no. 7) and the discussion of descriptive and causal inference (nos. 14, 25) have given some qualitative researchers a useful new vocabulary. As noted earlier in

24. Whereas the last section in chapter 1 above summarizes the arguments chapter by chapter, the organization here is thematic.

Table 2.2. Anticipating the Debate on *Designing Social Inquiry*

Evaluation of KKV's Contribution and Selected Examples Drawn from Guidelines Presented in Chapter 2	*Relevant Chapters in RSI (includes online chapters)*
I. Areas of Convergence	
a. Broad Convergence. Consensus on importance of standards for good descriptive and causal inference.	All Chapters
b. Specific Points of Convergence. Consensus that KKV offers much valuable advice with direct practical application in social science research (2, 3, 4, 5, 7, 12, 12b, 14, 16, 25, 31b/c/d, 33).	All Chapters
II. Areas of Divergence	
a. Extensive Treatment of Causal Inference, but Insufficient Attention to Its Logical Foundations. Greater attention needed to adequately address the obstacles faced in causal inference based on observational data (10, 26, 28, 29, 31, 31a/b/c/d/e, 32).	Brady (chap. 3); Bartels; Collier, Brady, and Seawright (chap. 9); Ragin (online); McKeown (online)
b. Important Issues Are Noted, but Seriously Neglected. Valuable advice is discussed briefly, but this advice must play a far more central role in research design (8, 9b/c, 12a–ii, 17, 21, 22).	Brady (chap. 3); Rogowski; Collier, Mahoney, and Seawright (online); Ragin (online); McKeown (online)
c. Regarding Key Advice, Practical Application May Not Be Feasible. Some advice may be hard to apply, not only in qualitative, but even in quantitative, research (13, 18, 23, 26, 28c, 31).	Brady (chap. 3); Bartels; Collier, Brady, and Seawright (chap. 9); Munck (online); McKeown (online)
d. Idea of Trade-Offs Is Mentioned, but Not Recognized as a Central Issue. Trade-offs among methodological goals must be a central concern in designing research (4a, 6b, 9, 9a, 11, 12a, 19, 27, 30a, 31, 34, 35, 35a).	Brady (chap. 3); Bartels; Rogowski; Tarrow; Collier, Brady, and Seawright (chaps. 8 and 9); Collier, Mahoney, and Seawright (online), Munck (online), Ragin (online)
e. Independent Contribution of Qualitative Tools Is Undervalued. Qualitative analysts have developed valuable tools that must to a greater degree be taken seriously on their own terms (1, 10, 13, 15, 17, 21, 22, 24, 30, 31).	Rogowski; Tarrow; Collier, Brady, and Seawright (chap. 9); Brady (chap. 12); Collier, Mahoney, and Seawright (online); Munck (online); Ragin (online); McKeown (online)

this chapter, part of the advice about selection bias is quite nuanced, in that KKV recognizes the importance of nonrandom sampling in the context of small-N research. Rather than offering the excessively limiting recommendation that scholars should not select on the dependent variable, the book suggests how sampling on the dependent variable is best carried out (no. 12). Replicability (no. 33) is certainly a widely held goal in the social sciences,[25] and other areas of agreement likewise emerge, as indicated in the table.

II. Areas of Divergence

In a number of other areas, the authors in the present volume raise questions about KKV's recommendations.

a. *Extensive Treatment of Causal Inference, but Insufficient Attention to Its Logical Foundations.* KKV is on the right track in pushing analysts to consider the assumptions that constitute the logical foundations of inference. However, the book's presentation of methodological norms falls short in helping scholars include the right variables, exclude the wrong ones, and more generally design their research and specify their models appropriately.

KKV's suggestion that researchers systematically search for and include relevant omitted variables (no. 28) usefully raises the issue of confounding variables, but does not say enough about which kinds of omitted variables ought to be included and which should be excluded. The recommendation that researchers exclude irrelevant explanatory variables (no. 29) leaves the same kinds of questions unanswered: How, exactly, should analysts distinguish between relevant and irrelevant explanatory variables before making a causal inference? Likewise, the advice that analysts should avoid endogeneity (no. 31) does too little to help researchers understand the substantive and theoretical reasons that endogeneity might or might not be a problem in a particular context. The specific techniques for addressing problems of endogeneity (nos. 31a–e) are valuable in pushing analysts to seek solutions to these problems, but much more needs to be said about the rather stringent assumptions behind these techniques.

Overall, KKV appears to embrace the proposition that these key problems of causal inference have been largely solved in mainstream quantitative research, and that, by extension, qualitative researchers should come as close as they can to adopting these solutions. By contrast, as argued by Brady, Bartels, and Seawright (chaps. 3, 4, 13, this volume), we are convinced that causal inference—not only in qualitative but also in quantitative research—is often problematic. Related issues of the logical founda-

25. Gary King has played a central role in subsequent debate on this issue. See *PS: Political Science and Politics* (1995) and *APSA-CP* (1996).

tions of inference are addressed by Ragin and McKeown in their online chapters.

KKV simply does not confront these difficulties squarely. The book does not give adequate recognition to problems of causal inference created by omitted variables and endogeneity. These issues are not easily resolved, even with advanced quantitative techniques. Consequently, causal inference, even with a large N, is often problematic. Hence, the applicability of KKV's methodological framework for causal inference in qualitative research remains doubtful.

b. *Important Issues Are Noted, but Seriously Neglected.* KKV mentions some key issues once or perhaps twice, yet some authors in the present volume consider them to be fundamental problems in the design of research that require far more attention. For example, KKV does cite Lieberson's (1985: chap. 5) incisive discussion of the need to focus empirical analysis on the range of variation relevant to the theory (no. 8); KKV also refers to using strong theory to address the problem of indeterminacy (no. 9b). Likewise, KKV notes that situating observations within a larger research program can help address the small-N problem (indeterminacy) and the problem of no-variance designs (nos. 9c, 12a–ii). Further, the book does mention the importance of knowing the context of research and of seeking validity and reliability in measurement (nos. 17, 21, 22). However, although these topics are noted briefly, they require much greater attention, given that KKV aims to provide a balanced set of recommendations for research design. These themes are explored below in the chapters by Brady and Rogowski. See also the online chapters by Collier, Mahoney, and Seawright; Ragin; and McKeown.

c. *Regarding Key Advice, Practical Application May Not Be Feasible.* Many of KKV's guidelines offer potentially useful methodological recommendations, yet authors in the present volume are concerned that it sometimes may not be feasible to apply this advice. For example, KKV usefully suggests that researchers pay close attention to the implications of measurement error for causal inference (no. 23). However, as Bartels argues, current statistical knowledge suggests that it can be difficult to know what those consequences are, even in quantitative research. Likewise, it is probably good advice to suggest that, in contexts where good causal inference is difficult, it is preferable to stick to good descriptive inference (no. 18). Yet this advice runs against the prevailing intellectual orientation within political science (and in KKV), where causal inference is strongly privileged over descriptive inference. As Brady (chapter 3) and McKeown (online chapter 4) argue, more reflection is needed on the proper relation between descriptive and causal inference.

Returning to the topic of endogeneity (no. 31), we find it useful to raise this issue, but it is also valuable to be candid about the fact that it can be exceedingly hard to address this problem, in either qualitative or quantitative research. Finally, the priority of demonstrating that the assumptions underlying causal inference are met in a given context of research (no. 26) is obviously important—as discussed in chapter 9 and in the online chapter by Munck—but little attention is devoted to exploring how this is to be done. In many contexts, it is simply not possible to demonstrate that these assumptions are met.

d. *Idea of Trade-Offs Is Mentioned, but Not Recognized as a Central Issue.* KKV pays insufficient attention to trade-offs, failing to recognize that they are an overarching issue in research design. Trade-offs are a central theme in the chapters below. As discussed in this volume by Brady (chap. 3) and Bartels, and in chapters 8 and 9, the mandate to increase the number of observations—for the purpose of strengthening falsifiability, increasing leverage, and addressing indeterminacy and multicollinearity (nos. 4a, 6b, 9a, 30a)—may make it harder to achieve other important goals, such as maintaining independence of observations, measurement validity, and causal homogeneity.

Next, as emphasized by Brady (chap. 3) and in chapter 8 of this volume, while working with concrete and observable concepts (no. 19) certainly makes measurement easier, many theories depend on abstract concepts that are well worth measuring, even if it is not easy to do so. An obvious example is the concept of causation. KKV (76, 79) in fact recognizes it as an abstract, theoretical concept, and much of the book is devoted to discussing how best to measure it. Many other indispensable concepts are likewise hard to measure.

Additionally, the idea of a determinate versus indeterminate research design (no. 9) raises the important issue of having a sufficient number of observations to adjudicate among rival explanations; yet, as chapter 9 in the present volume argues, this distinction creates the misleading impression that research designs based on observational, as opposed to experimental, data can really be determinate—which is not the case. Indeed, causation can generally only be inferred in observational studies if the researcher imposes several restrictive assumptions, which may be difficult to test or even to defend.

Finally, as argued by Rogowski, and by Collier, Mahoney, and Seawright, the warning against designs that lack variance on the dependent variable (no. 12a) must be weighed against the analytic gains that can derive from closely analyzing positive cases of a given phenomenon, especially if little is known about it.

Other recommendations made by KKV also involve trade-offs. These recommendations involve issues of inductive analysis, endogeneity, and complexity. From one point of view, the injunctions against the post hoc reformulation and testing of hypotheses (nos. 34, 35, 35a) make good sense, in that it weakens the power of statistical tests. However, as Ragin (online chapter 3), Munck (online chapter 2), and Tarrow argue, for qualitative researchers the refinement of theory and hypotheses through the iterated analysis of a given set of data is an essential research tool, and researchers lose other aspects of analytic leverage by not employing it.[26] Indeed, quantitative studies regularly follow a similar path. When quantitative researchers analyze observational data, they almost never conduct one test of the initially hypothesized statistical model and then stop. Rather, they routinely carry out elaborate specification searches, involving iterated attempts to find an appropriate fit between models and data. For this reason, a major literature within econometrics has discussed procedures and tools that help quantitative researchers conduct their specification searches in a disciplined manner. This literature recognizes that the quantitative analysis of observational data routinely involves an iterated, partly inductive, mode of research.

A closely related point concerns data mining. Indiscriminate data mining is a bad idea, and the statement that selecting relevant explanatory variables requires theory is uncontroversial (no. 27). However, as just noted, all research has an inductive component, and we should not foreclose the possibility of accidental discoveries. The challenge is to be open to such discoveries that are not anticipated by our theory; yet at the same time to avoid the atheoretical, indiscriminate pursuit of new hypotheses, which may lead to findings that are not analytically meaningful.

Finally, returning to the issue of endogeneity (no. 31), selecting cases so as to avoid this problem makes sense in that it facilitates causal inference. Yet this priority absolutely should not preclude, for example, looking at processes of change over time, where endogeneity is commonly present. Given the larger intellectual movement in recent decades toward the historicization of the social sciences, scholars who study causal processes over a long time horizon must routinely treat endogeneity as a problem to be confronted, rather than avoided.

e. *Independent Contribution of Qualitative Tools Is Undervalued.* KKV pays insufficient attention to the independent contributions of qualitative tools, sometimes too quickly subordinating them to a quantitative template. KKV makes an interesting argument that quantitative/formal work and interpretation are *similar* in an important respect: both simplify drastically, compared to the full complexity of reality (no. 15). While this is true, for the researcher trying to learn about the distinctive strengths of alternative methodological approaches, the dissimilarity of interpretation and quantitative/formal analysis is a far more central concern, a theme that arises in chapter 13 below. KKV's framing inappropriately deemphasizes the contributions of interpretive work, and of other qualitative approaches, to goals that a regression-oriented framework addresses much less successfully—including concept formation and fine-grained description.

26. KKV does discuss the interaction between theory and data, but within the framework of arguing that any further test of the theory should be undertaken with *new* data (KKV 21, 46).

Qualitative researchers also have distinctive perspectives on causal heterogeneity (no. 10). It is a central component within Ragin's framework, and Tarrow shows how qualitative methods provide valuable tools for explaining transitions and nonlinearity that have been discovered through quantitative analysis. With reference to separating the systematic and the random components of phenomena (no. 24), Munck suggests that qualitative researchers may approach this issue by employing insights about causal mechanisms and the larger research context. Isolating the systematic components can, in turn, provide a substitute for statistical control by eliminating the variance on the dependent variable caused by factors outside the focus of the analysis.

Finally, and most importantly, KKV's arguments about strengthening causal inference through increasing the number of observations can be refined by recognizing the importance of different kinds of observations: that is, data-set observations and causal-process observations, a distinction introduced in chapter 1 above and explored at length in chapter 13 and in the appendix. Utilizing this distinction makes it easier to recognize the valuable leverage in causal inference that derives from within-case analysis—which has been a long-standing focus in discussions of qualitative methods and is an important concern in the chapter below by Rogowski, the online chapters by Collier, Mahoney, and Seawright, Munck, and McKeown, as well as in Tarrow's discussion of triangulation. KKV notes these procedures, but the book prematurely seeks to subordinate them to the standard tools of quantitative inference (KKV 85–87, 226–28).

To conclude, KKV articulates a clear summary of the mainstream quantitative framework in social science. At the same time, the book seeks to impose this framework on other kinds of research. In the process, KKV loses sight both of major weaknesses in the quantitative template and of many strengths that have made other tools worth developing in the first place. KKV's arguments have stimulated scholars to rethink both the quantitative and qualitative traditions. Based on this rethinking, the chapters below seek to present a more balanced view of methodology and research design.

6

Bridging the Quantitative-Qualitative Divide

Sidney Tarrow

In *Designing Social Inquiry* (hereafter KKV), Gary King, Robert O. Keohane, and Sidney Verba 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—in deference to 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" (105), I will praise the book no more, but focus on an important weakness in the book: KKV's 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 (3). If this is so, then the authors 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. While they offer a good deal of generous (if 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 practitioners often illustrate their theories with narrative, there is a need for a set of ground rules on how to make intelligent use of qualitative data.

KKV does not address this issue. Rather, it uses 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 its cardinal operational rule.) But in today's social science world, how many social scientists can simply be 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. In what follows, I will discuss some of the problems of combining qualitative and quantitative data, as well as some solutions to these problems.

CHALLENGES OF COMBINING QUALITATIVE AND QUANTITATIVE DATA

A recent work that KKV warmly praises illustrates both that its 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. 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 twenty-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 (quoting Putnam) recommends as "an intimate knowledge of the internal political maneuvering and personalities that have animated regional politics over the last two decades" (5) and what Putnam calls "marinating yourself in the data" (KKV: 5; Putnam 1993: 190). KKV (38) uses *Making Democracy Work* to praise the virtues of "soaking and poking," in the best Fenno (1977: 884) tradition.

But Putnam's debt to qualitative approaches is much deeper and more problematic than this; 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 northcentral and southern regions. In his effort to find them, his quantitative evidence offered only indirect help, 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 northern Italian regional governments vis-à-vis southern ones. This he based on the civic traditions of the (northern) Renaissance city-states, which, according to him, provided "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 question in melding quantitative and qualitative approaches that KKV's canons of good scientific practice do not help to resolve. 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? 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 by what rules of inference Putnam chose the Renaissance as determining the Italian North's late twentieth-century 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 doesn't tell us; nor does KKV.

To generalize from the problem of Putnam's book, qualitative researchers have much to learn from the model of quantitative research. But 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 tools for bridging the quantitative-qualitative divide from recent examples of comparative and international research (see table 6.1).

TOOLS FOR BRIDGING THE DIVIDE

Tracing Processes to Interpret Decisions

One such tool that KKV cites 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'" (226; quoting George and McKeown 1985: 35). KKV interprets the advantages of process tracing narrowly, assimilating it to their favorite goal of increasing the number of theoretically relevant observations (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 cites so favorably.

Systematic and Nonsystematic Variable Discrimination

KKV gives us a second example of the uses of qualitative data but, once again, underestimates its particularity. The authors 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 (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," the

Table 6.1. Tools for Bridging the Qualitative-Quantitative Divide

Tool	*Contribution to Bridging the Divide*
Process Tracing	Qualitative analysis focused on processes of change within cases may uncover the causal mechanisms that underlie quantitative findings.
Focus on Tipping Points	Qualitative analysis can explain turning points in quantitative time series and changes over time in causal patterns established with quantitative data.
Typicality of Qualitative Inferences Established by Quantitative Comparison	Close qualitative analysis of a given set of cases provides leverage for causal inference, and quantitative analysis then serves to establish the representativeness of these cases.
Quantitative Data as Point of Departure for Qualitative Research	A quantitative data set serves as the starting point for framing a study that is primarily qualitative.
Sequencing of Qualitative and Quantitative Studies	Across multiple research projects in a given literature, researchers move between qualitative and quantitative analysis, retesting and expanding on previous findings.
Triangulation	Within a single research project, the combination of qualitative and quantitative data increases inferential leverage.

authors conclude, "our observations might be rather poor measures of underlying Conservative strength" (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 proposes, 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 strike incidence and industrial relations.

To put this point more abstractly, distinct historical events often serve as the tipping points that explain the shifts 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 seems to argue, to peel away layers of unsystematic fluff from the hard core of systematic variables; but also to assist researchers in understanding shifts in the values of the systematic variables.

Framing Qualitative Research within Quantitative Profiles

The uses of qualitative data described in the two previous sections 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 database as a frame within which the qualitative analysis is carried out. Case studies have been validly criticized as often being based on dramatic but frequently unrepresentative cases. Studies of successful social revolutions often focus on 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 representativeness 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 database 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 700 revolutionary events in over 500 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. However, he did not believe the concept of *revolution* had the monolithic quality that other social scientists had assigned to it (1993: chap. 1). Therefore, he resisted the temptation for quantification, using his database, 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 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

An American sociologist, Doug McAdam, has shown how social science can be enriched by carrying out a sustained qualitative analysis of what is initially a quantitative database. 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. 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 re-specifying hypotheses to allow them to be tested on different units!" (see chap. 7). But would they argue, as I do, 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 workers who were 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 demonstrated 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, which had been emphasized in a series of books by their foreign friends.

The response of those who had formulated the intellectualist interpretation of Solidarity was predictably indignant. 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 her *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 popular 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. However, as the worker and student mobilization 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 by triangulating on the same problem. For a larger study, I had gathered 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 that in Florence throughout the country (Sciubba and Pace 1976). By reoperationalizing the hypothesis cross-sectionally, I was able to show a reasonably high positive correlation (.426) between the presence of Catholic base communities in various cities and the magnitude of general collective action in each city (Tarrow 1989: 200). Triangulation demonstrated that the findings of my longitudinal, local, and qualitative case study coincided with the results of cross-sectional, national, and quantitative correlations. My inductive hunch that Italy in the 1960s underwent an integrated cycle of protest became a more strongly supported hypothesis.

KKV does not take the position that quantification is the answer to all the problems of social science research. But the book's 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 42–43; and of the second, see KKV 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.

CONCLUSION

KKV gives 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 its authors. But the book's theoretical effort is marred by the narrowness of its empirical specification of qualitative research and by its 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 argue here.

7

The Importance of Research Design

Gary King, Robert O. Keohane, and Sidney Verba

Receiving five serious reviews in this symposium[1] 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 raise 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

1. Editors' note: This chapter is reprinted from the 1995 symposium on *Designing Social Inquiry*, published in the *American Political Science Review*. In this chapter, the authors respond to arguments developed in three additional articles in the *APSR* symposium that are reprinted in the present volume: those by Rogowski, Tarrow, and (reprinted in part) Collier. King, Keohane, and Verba likewise respond here to the two other articles in the symposium—by Laitin (1995) and Caporaso (1995)—to which reference is made in the present volume, but which are not included here. The full original citation for this chapter is Gary King, Robert O. Keohane, and Sidney Verba (1995) "The Importance of Research Design in Political Science." *American Political Science Review* 89, no. 2 (June): 475–81. The table of contents, preface, and chapter 1 of *Designing Social Inquiry* are available at pup.princeton.edu/titles/5458.html.

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" (KKV 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 (95 this volume). 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" (1995: 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 (KKV chap. 2). We discuss causality in relation to counterfactual analysis and what Paul Holland (1986) calls the "fundamental problem of causal inference" and consider possible complications introduced by thinking about causal mechanisms and multiple causality (KKV chap. 3). Our discussion of counterfactual reasoning is, we believe, consistent with Donald Campbell's "quasi-experimental" emphasis (Campbell and Stanley 1963); and we thank James Caporaso for clarifying this.[2]

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

2. 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 emphasizes also the value of random assignment, which is desirable in some situations (but not in others, see KKV 124–28) 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. It is a fine idea, much of which we have adopted; but it is an unnecessary category.

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 single-mindedness 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).[3]

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

3. 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."

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 entailed in 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 (1975 [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" (KKV 211; see also sections 1.2.1 and 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 (1975 [1968]) sought to estimate this causal effect.[4] 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 is 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*" (95 this

4. 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" (97 this volume) is surely the first time it has received such accolades.

volume)—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. 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" (KKV 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" (1995: 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 chapter 4).[5]

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 forty 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" (1985: 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 (1985: chap. 4).

Katzenstein seeks to test the first of these causal claims by comparing economic openness in small and large states (1985: 86, table 1). 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, and 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

5. 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" (1995: 460). 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.

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" (1985: 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" (KKV 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 argue in *Designing Social Inquiry*; see also Collier 1995) 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 commitment 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 analysis 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 wrote 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.[6]

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 num-

6. 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 did exactly what we recommend: having developed his theory in one domain, he extracted its observable implications and moved to other domains to see whether he observes what the theory would lead him to expect.

ber 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; 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, refers to the practice of increasing the amount of information brought to bear on a theory or hypothesis, and that is what our book is about.

D. DIVERSE TOOLS, SHARED STANDARDS