

Case Studies and Theory Development in the Social Sciences

Alexander L. George and Andrew Bennett

BCSIA Studies in International Security

MIT Press
Cambridge, Massachusetts
London, England

1 L SHORT

REGULAR

Long

JOB# 997

George

Copyright © 2004 by the Belfer Center for Science and
International Affairs
John F. Kennedy School of Government
Harvard University
Cambridge, Massachusetts 02138
(617) 495-1400

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, electrostatic, magnetic tape, mechanical, photocopying, recording, or otherwise—without permission in writing from the Belfer Center for Science and International Affairs, 79 John F. Kennedy Street, Cambridge, MA 02138.

Library of Congress Cataloging-in-Publication Data

p. cm.—(BCSIA studies in international security)
CIP

10 9 8 7 6 5 4 3 2 1

Printed in the United States of America

1 L SHORT

REGULAR

Long

JOB# 997

 George

Contents

Preface		ix
Part I	Case Studies and Social Science	1
Chapter 1	Case Studies and Theory Development	3
Chapter 2	Case Study Methods and Research on the Interdemocratic Peace	37
A Pedagogical Note to Parts II and III		61
Part II	How to Do Case Studies	65
Chapter 3	The Method of Structured, Focused Comparison	67
Chapter 4	Phase One: Designing Case Study Research	73
Chapter 5	Phase Two: Carrying Out the Case Studies	89
Chapter 6	Phase Three: Drawing the Implications of Case Findings for Theory	109
Part III	Alternative Methods and Select Issues	125
Chapter 7	Case Studies and the Philosophy of Science	127
Chapter 8	Comparative Methods: Controlled Comparison and Within-Case Analysis	151

Chapter 9	The Congruence Method	181
Chapter 10	Process-Tracing and Historical Explanation	205
Chapter 11	Integrating Comparative and Within-Case Analysis: Typological Theory	233
Chapter 12	Case Studies and Policy-Relevant Theory	263
Appendix		287
Studies That Illustrate Research Design		287

Preface

An extended methodological dialogue is bringing the comparative advantages of case study methods for theory development into sharper focus. Our own personal dialogue began with intermittent conversations in the 1990s on our independent work on case study methods. We both felt that the time was ripe to draw on the lessons learned from the widespread use of sophisticated case study methods developed in recent decades. These include Alexander George’s method of “structured, focused comparison of cases,” which outlines process-tracing and other within-case modes of analysis as key complements or alternatives to controlled comparison of cases, Arend Lijphart and Harry Eckstein’s extremely useful elaborations of different theory-building kinds of case studies, and Charles Ragin’s analysis of interactions effects and comparative methods of studying them.

This book draws on the work of many scholars over the past thirty years to raise the standards and explicate the procedures of theory-oriented case study methods. Further experience with theory-oriented case study research will no doubt lead to further refinements. This book seeks to advance earlier discussions of case study methods in three particular areas. First, in contrast to earlier discussions that focus on case comparisons, we emphasize that qualitative research usually involves a combination of cross-case comparisons and within-case analysis using the methods of congruence testing and process-tracing. Within-case methods of analysis can greatly reduce the well-known risks of inferential errors that can arise from using comparative methods alone. Second, we elaborate on the methods of congruence testing and process-tracing, discussing them in detail and providing examples from recent research. Third, we develop the concept of typological theorizing, which resembles

both Robert K. Merton's discussion of "middle-range theory" and Paul Lazarsfeld's notion of a "property space." We argue that typological theories involving several variables can better capture the complexity of social life than the two-variable typological theories that are common in the social sciences, and we offer methods for building typological theories in ways that keep this complexity manageable and clarify the task of selecting which cases to study.

In the process of writing this book, we have attempted through conferences, workshops, a web site, and other organized efforts to stimulate interest in improving and disseminating case study methods. In particular, together with Colin Elman of Arizona State University and David Collier of the University of California at Berkeley, we have established the Consortium on Qualitative Research Methods (CQRM), which sponsors an annual training institute in these methods for advanced graduate students and junior faculty. The institute convenes each January at Arizona State University, which generously funds the consortium with member departments and research institutes. In addition, we have assisted David Collier in the creation of a new Qualitative Methods section of the American Political Science Association. Information on both the Qualitative Methods Section and CQRM can be found on their shared web site (www.asu.edu/clas/polisci/cqrm/).

A note is in order regarding the development of our own interests in this subject and the division of labor in this book. Alexander George's interest in case study methodology developed in the 1960s while he was a researcher at the RAND Corporation working on generic problems of avoiding and managing interstate conflict during the Cold War. The first of these problems concerned extended deterrence on behalf of weaker U.S. allies; soon thereafter the research program extended to problems of using coercive diplomacy to reverse an adversary's action against an ally or friendly neutrals, and then to managing conflicts to avoid unwanted escalation.

George was interested in finding ways of studying historical instances of these generic problems that would permit valid, usable "lessons" to be drawn from case findings. These lessons should be formulated in ways that would help policy specialists diagnose accurately new cases of each of these phenomena so that informed judgments could be made in deciding whether and how to use one of these strategies in each new situation. George and his RAND colleagues found little in the academic literature that provided methods for studying historical experience from this perspective. Accordingly, it was necessary to devise a case study methodology to analyze past instances of each of these generic problems to identify conditions and procedures that were associated with successful or failed outcomes. The challenge was to find ways of doing

comparative analysis of a number of instances of each generic problem in ways that would draw analytical explanations of each case into a broader, more complex theory, one that would discourage reliance on a single historical analogy.¹

The aim was to identify more specific, differentiated causal patterns of successful and ineffective ways of employing each strategy. These patterns would initially consist of generalizations of quite limited scope. Such middle-range theories on deterrence, coercive diplomacy, and crisis management would consist of a variety of conditional, contingent generalizations (for a discussion of our use of middle-range theory, see Chapters 11 and 12).

For this purpose, George adapted methods of historical explanation to convert descriptive explanations of case outcomes into analytic explanations comprised of variables.² This procedure made use of an inductive approach for theory-building, but it was *analytic* induction, not raw empiricism. The black boxes of decision-making and strategic interaction were opened up and efforts were made to study actual processes of decision-making and of strategic interaction insofar as available data permitted.

In this research, George and his colleagues were not interested in—and indeed their methods did not permit—using the findings of a few cases that were not necessarily representative to project a probability distribution of different patterns discovered for the entire universe of instances of, for example, deterrence. Rather, contingent generalizations were intended to help policy specialists first to diagnose and then to prescribe for new situations, much as medical doctors do in clinical settings. This theme runs through all of the publications of George’s research program over the years and finds its latest, most detailed statement in Chap-

1. The key objective of the important book by Richard E. Neustadt and Ernest R. May, *Thinking in Time: The Uses of History for Decision Making* (New York: Free Press, 1986), is to suggest various ways in which policymakers can avoid relying on a single historical analogy. However, these authors do not address the question of how the lessons of a number of cases of a given phenomenon can be cumulated to provide a differentiated theory. For a more recent statement on the need to derive “lessons” from historical experience, see William W. Jarosz with Joseph S. Nye, Jr., “The Shadow of the Past: Learning From History in National Security Decision Making,” in Philip Tetlock et. al., *Behavior, Society, and International Conflict*, Volume 3 (New York: Oxford University Press, 1993), pp. 162–189.

2. In Harry Eckstein’s terminology, an ideographic atheoretical explanation was converted into a “disciplined configurative” study. An early explicit example of this procedure was contained in Gabriel Almond, Scott Flanagan, and Robert Mundt, eds., *Crisis, Choice, and Change: Historical Studies in Political Development* (Boston: Little, Brown, 1973), pp. 22–28.

ter 12 of his 1993 book on *Bridging the Gap* between scholarly research and policymaking.

Another early step in George's development of what he later termed the method of "structured, focused comparison" was his codification of Nathan Leites' concept of "operational code beliefs." George converted Leites' analysis into a set of *general questions* that could be asked in studying the operational code beliefs of other elites and individual leaders. He called attention to the potential use of the set of philosophical and instrumental beliefs embraced by an operational code in comparative studies of leaders.³ A large number of these types of studies were done after the publication of George's codification of operational code beliefs.

George's comparative work on deterrence led to the further development of the structured, focused method.⁴ He published an early version of this method in 1979, greatly elaborating on the brief description of it in his 1974 book on deterrence.⁵ Also in 1979, George published a companion piece that addressed more detailed aspects of the method.⁶ This second article provided the first detailed statement about process-tracing in case studies and the congruence method, both of which receive detailed treatment in the present book.

George also introduced the structured, focused method into a course he team-taught with several historians. This collaboration resulted in a book co-authored with Gordon Craig, *Force and Statecraft*, which has been updated several times since first published by Oxford University Press in 1983. George also taught a Ph.D. level seminar on structured, focused

3. Alexander L. George, "The 'Operational Code': A Neglected Approach to the Study of Political Leaders and Decision-Making," *International Studies Quarterly*, Vol. 13, No. 2 (June 1969), pp. 190-222. Ole Holsti contributed to the refinement of operational code research, and Stephen Walker has developed a detailed research program and many publications on operational codes.

4. See Alexander L. George and Richard Smoke, *Deterrence in American Foreign Policy: Theory and Practice* (New York: Columbia University Press, 1974), pp. 95-103. The earlier book, *The Limits of Coercive Diplomacy*, edited by Alexander L. George, David K. Hall, and William E. Simons (Boston: Little, Brown, 1971), was a comparative study of three cases, but did not explicitly follow the rubrics of structured, focused comparison. However, this was the research design of *U.S.-Soviet Security Cooperation: Achievements, Failures, Lessons*, coedited with Philip T. Farley and Alexander Dallin (New York: Oxford University Press, 1988).

5. Alexander L. George, "Case Studies and Theory Development: The Method of Structured, Focused Comparison," in Paul Gordon Lauren, ed., *Diplomatic History: New Approaches* (New York: Free Press, 1979).

6. Alexander L. George, "The Causal Nexus Between Cognitive Beliefs and Decision-Making Behavior: The 'Operational Code' Belief System," in Lawrence S. Falkowski, ed., *Psychological Models in International Politics* (Boulder, Colo.: Westview Press, 1979), pp. 95-124.

comparison through the 1980s that became a required course at Stanford for graduate students in comparative politics. Many international relations students took it as well, and it led to the completion of many theses and to the publication of numerous books using the structured, focused method.

Andrew Bennett's interest and training in case study methods began when he was one of George's undergraduate students at Stanford University in the early 1980s. Bennett then used qualitative methods in books on Soviet and Russian military interventions and burden-sharing in the 1991 Persian Gulf War.⁷ Bennett has taught a graduate seminar in case study methods at Georgetown since 1997.

This book is very much a product of close co-authorship, and each of us has contributed to every chapter, but it is worth noting which author is primarily responsible for each chapter. Alex George is the primary author of Chapters 3, 4, 5, 8, 9, 12, and the Appendix, while Andy Bennett is the primary author of Chapters 2, 6, and 11. Chapters 1, 7, and 10 were truly joint efforts with equal contributions by both authors.

Organization of the Book

It may be helpful to steer readers toward the chapters that are likely to meet their interests.

Chapter 1 surveys the developments over a period of years that have improved the direction and quality of case study research and its contributions to theory development. Readers will note that our objective in this book, as in our previous works, is to raise the standards for case studies and explicate procedures for improving their value.

Chapter 2 provides a concrete demonstration of how case studies combined with quantitative methods have contributed to the development of research on democratic peace theory. It illustrates one of the major themes of the book, namely the purposes best served by different research methods, and how knowledge cumulates within a research agenda.

These two chapters should satisfy general readers who want to understand the role and contribution of case studies for the development of theories but have no plans for doing such research themselves.

For readers who are undertaking Ph.D. dissertations and for instructors who offer course work and guidance on case study methods, we

7. Andrew Bennett, *Condemned to Repetition? The Rise, Fall, and Reprise of Soviet-Russian Military Interventionism, 1973–1996* (Cambridge, Mass.: MIT Press, 1999); and Andrew Bennett, Joseph Lepgold, and Danny Unger, eds., *Friends in Need: Burden-Sharing in the Gulf War* (New York: St. Martin's Press, 1997).

present a manual in Part II and Part III of the book. A detailed Note to Parts Two and Three provides additional information on the development of the manual for doing case study research. We have also included an Appendix, “Studies That Illustrate Research Design,” which briefly reviews the research designs of numerous books. We expect these ingenious and varied research designs to be helpful to Ph.D. students contemplating such research and to professors in designing instruction on case methods.

Acknowledgments

This book would not have been possible without the support of many programs and individuals. We wish to thank the MacArthur Foundation for funding a conference that helped launch the book project; the Carnegie Corporation of New York for additional research funding; and the Stanford University Center for International Security and Cooperation and Columbia University’s History and Political Science Departments for sponsoring workshops on our book.

We want to give special thanks to David Dessler, Jack Levy, David Collier, Colin Elman, Miriam Fendius Elman, James Mahoney, Gary Goertz, and Bear Braumoeller for insightful suggestions on major portions of the manuscript. We want to thank many other colleagues as well for their useful suggestions on various chapters, including Hayward Alker, Robert Art, Pierre Atlas, Aaron Belkin, Aaron Boosehecker, Henry Brady, Lynn Eden, Leslie Eliason, Mary Jane Fox, David Friedman, John Lewis Gaddis, John Gerring, Emily Goldman, Jack Goldstone, Stuart Gottlieb, Thomas Homer-Dixon, Ronald Jepperson, Chaim Kaufmann, Jane Kellett-Cramer, Charles Kiami, Deborah Larson, Jeff Legro, Roy Licklider, Dan Lindley, Daniel Little, Andy Loomis, Timothy McKeown, Ron Mitchell, Andrew Moravcsik, Gerry Munck, Dan Nexon, Charles Ragin, Volker Rittberger, Scott Sagan, Steve Saideman, Daniel Schwartz, Jack Snyder, Detlef Sprinz, Brian Taylor, Charles Tilly, Stephen Van Evera, David Waldner, Steve Walt, and Yael Wolinsky.

Michael Boyle provided excellent suggestions for improving the entire manuscript to make it more accessible to Ph.D. students contemplating case study research. Eliana Vasquez provided much essential support in a cheerful and efficient manner.

We wish to thank Sean Lynn-Jones and the Belfer Center for Science and International Affairs and MIT Press for hosting a conference on our book and for outstanding help at every stage of the publishing process. Miriam Avins did an excellent job as copyeditor in streamlining and clarifying a complex manuscript. Alex George expresses deep appreciation to Belinda Yeomans for her years of invaluable research assistance and ad-

ministrative services, and Andy Bennett thanks her for her exceptional work in organizing a manuscript written on two coasts by authors incompletely socialized to e-mail.

Finally, we want to express particular thanks to Bob Keohane and Gary King for insightful and very useful suggestions on an earlier draft of our manuscript. Their constructive approach to our work has been especially valuable in view of the fact that we express important disagreements with their book, *Designing Social Inquiry* (co-authored with Sidney Verba, who lacked time to give us comments). Our disagreements have been intellectual, not personal, and they stem from a shared interest in improving research methods of all kinds. One of our central themes is that statistical methods, case studies, and formal models should be regarded as complementary, rather than competitive. Research can progress more effectively through diverse methods than it can through any one method alone.

—Alexander L. George
Stanford, California

—Andrew Bennett
Washington, D.C.
September 2004

1 L SHORT

REGULAR

Long

JOB# 997

 George

Part I
Case Studies and Social Science

1 L SHORT

REGULAR

Long

JOB# 997

George



1 L SHORT

REGULAR

Long

JOB# 997

 George



Chapter 1

Case Studies and Theory Development

After decades of rapid and contentious change, social science research methods are entering a new phase of development conducive to cross-method collaboration and multi-method work. The changes in these methods over the past four decades have been truly revolutionary. With improvements in computing capabilities, databases, and software, statistical methods and formal models increased rapidly in their sophistication and their prevalence in published research in the 1960s and 1970s. While qualitative and case study methods have also become more sophisticated, the proportion of published research using these methods declined sharply in the 1960s and 1970s, as these methods had already been prominent or even dominant in the social sciences, and their share of the social science market naturally declined as the more novel statistical and formal methods of research grew. To take one example from a leading journal in our own field of political science, between 1965 and 1975, the proportion of articles in the *American Political Science Review* using statistics rose from 40 percent to over 70 percent; that using formal models rose from zero to over 40 percent; and the proportion using case studies plunged from 70 percent to under 10 percent, with about 20 percent of the articles using more than one method.¹ Other social science disciplines, including sociology, history, and economics, have undergone methodological changes as well, each in its own way and at its own pace.

These rapid and far-reaching shifts in research methods in earlier decades were naturally contentious, as they affected opportunities for re-

1. Andrew Bennett, Aharon Barth, and Ken Rutherford, "Do We Preach What We Practice? A Survey of Methods in Political Science Journals and Curricula," *P.S.: Political Science and Politics*, Vol. 36, No. 3 (July 2003), p. 375.

search funds, teaching positions, and publication outlets. Even scholars with similar substantive interests have formed into largely separate communities along methodological lines. To take another example from our field, of two journals that cover similar theoretical issues and policy concerns, *The Journal of Conflict Resolution* publishes almost no case studies, and *International Security* publishes almost no statistical or formal work. Such methodological specialization is not in itself counterproductive, as every journal needs to establish its own niche. In this instance, however, there is troubling evidence of a lack of cross-method communication, as each of these journals frequently cites its own articles and very rarely cites those published by the other.²

More recently, however, a variety of developments has made possible an increasingly sophisticated and collaborative discourse on research methods in the social sciences that focuses upon the essential complementarity of alternative methodological approaches.³ Over the past few decades, proponents of case study methods, statistics, and formal modeling have each scaled back their most ambitious goals regarding the kinds of knowledge and theories that they aspire to produce. Practitioners of each approach have improved and codified their techniques, reducing some of the problems identified by their critics but also gaining renewed appreciation for the remaining limits of their methods. The mix of methods has become fairly stable, at least in our own field, with each method secure in its ability to contribute to theoretical progress. In contrast to the sharp changes in methods used in journal articles in the 1960s and 1970s, the mix of methods used in articles in the top political science journals has been fairly stable since the mid-1980s, and in recent years

2. Ibid., p. 376.

3. A useful commentary on developments in case study research is provided by Jack Levy, "Qualitative Methods in International Relations," in Michael Brecher and Frank P. Harvey, eds., *Millennial Reflections on International Studies* (Ann Arbor: University of Michigan Press, 2002), pp. 432–454. See also the excellent treatment of these issues in Stephen Van Evera, *Guide to Methods for Students of Political Science* (Ithaca, N.Y.: Cornell University Press, 1997). In their 1996 review of the state of political science, Robert Goodin and Hans-Dieter Klingemann argue that in the "Jacobin" behavioral revolution of the 1960s and the "Thermidorian" reaction that followed, contending factions "heaped Olympian scorn" on one another. This scenario was then replayed in the "Manichean" controversy over rational choice theory. More recently, they argue, there has been a "rapprochement," fostered by the rise of the "new institutionalism," and "political scientists no longer think in the either/or terms of agency or structure, interests or institutions . . . realism or idealism, interests or ideas . . . science or story-telling . . . mono-causality or hopeless complexity." They do not see this rapprochement as a sloppy "'live and let live' pluralism," but as a sign that the present generation of political scientists are "equipped with a richer toolkit than their predecessors." Robert Goodin and Hans-Dieter Klingemann, eds., *A New Handbook of Political Science* (New York: Oxford University Press, 1996), pp. 10–13.

roughly half of these articles used statistics, about the same proportion used case studies, slightly fewer than a quarter used formal models, and about one in five used more than one method.⁴

Moreover, a new generation of scholars has emerged with training in or at least exposure to more than one methodology, allowing easier translation among the different forms through which fundamental epistemological limits are embodied in different methods. Developments in the philosophy of science have also clarified the philosophical foundations of alternative approaches. Finally, the various fields in the social sciences have, at different speeds and to different degrees, addressed the historical, sociological, and postmodernist “turns” by focusing on norms, institutions, and actors’ identities and preferences, but doing so through largely neopositivist means. As a result, scholars are increasingly working collaboratively across methodological divides to advance shared substantive research programs. Most of this cross-method collaboration has taken place sequentially, as different researchers have used the methods in which they are most adept but have also drawn on the findings of those using other methods. Because cross-method collaboration in the social sciences has until recently rarely involved one or more individuals working on the same publication with different methods, it has been underappreciated.

A prerequisite for this revitalized methodological dialogue is a clear understanding of the comparative strengths and limits of various methods, and how they complement each other. This book contributes to this dialogue by focusing on the comparative advantages of case study methods and on these methods’ ability to contribute to the development of theories that can accommodate various forms of complex causality.

The case study approach—the detailed examination of an aspect of a historical episode to develop or test historical explanations that may be generalizable to other events—has come in and out of favor over the past five decades as researchers have explored the possibilities of statistical methods (which excel at estimating the generalized causal weight or causal effects of variables) and formal models (in which rigorous deductive logic is used to develop both intuitive and counterintuitive hypotheses about the dynamics of causal mechanisms). Perhaps because case study methods are somewhat intuitive—they have in some sense been around as long as recorded history—the systematic development of case study methods for the cumulative building of social science theories is a comparatively recent phenomenon (notwithstanding notable contributions to these methods by John Stuart Mill). Only in the past three de-

4. Bennett, Barth, and Rutherford, p. 374.

ades have scholars formalized case study methods more completely and linked them to underlying arguments in the philosophy of science.

Indeed, statistical methods have been so prominent in recent decades that scholars' understanding of case studies is often distorted by critiques based on the assumptions of statistical methods. We argue that while case studies share a similar epistemological logic with statistical methods and with formal modeling that is coupled with empirical research, these methods have different methodological logics. Epistemologically, all three approaches attempt to develop logically consistent models or theories, they derive observable implications from these theories, they test these implications against empirical observations or measurements, and they use the results of these tests to make inferences on how best to modify the theories tested.⁵ Methodologically, these three methods use very different kinds of reasoning regarding fundamental issues such as case selection, operationalization of variables, and the use of inductive and deductive logic. These differences give the three methods complementary comparative advantages. Researchers should use each method for the research tasks for which it is best suited and use alternative methods to compensate for the limitations of each method.

In addition to clarifying the comparative advantages of case studies, this book codifies the best practices in the use of case studies; examines their relationship to debates in the philosophy of science; and refines the concept of middle-range or typological theories and the procedures through which case studies can contribute to them. Our focus extends to all aspects of theory development, including the generation of new hypotheses as well as the testing of existing ones.

Throughout the book, we have paid special attention to the method of process-tracing, which attempts to trace the links between possible causes and observed outcomes. In process-tracing, the researcher examines histories, archival documents, interview transcripts, and other sources to see whether the causal process a theory hypothesizes or implies in a case is in fact evident in the sequence and values of the intervening variables in that case. Process-tracing might be used to test

5. While at this general level the epistemologies of alternative research methods are quite similar, significant differences remain, as these methods are optimized for different epistemic aims. These aims include the estimation of measures of correlation for populations of cases and the establishment of probabilistic levels of confidence that these correlations are not due to chance (tasks at which statistical methods are effective when the assumptions necessary for these methods are met), the development and testing of historical explanations and the detailed exploration of hypothesized causal mechanisms in the context of particular cases (where case studies have comparative advantages), and the deductive development of logically complete and consistent theories (the forte of formal modeling).

whether the residual differences between two similar cases were causal or spurious in producing a difference in these cases' outcomes. Or the intensive study of one deviant case, a case that fails to fit existing theories, may provide significant theoretical insights. Process-tracing can perform a heuristic function as well, generating new variables or hypotheses on the basis of sequences of events observed inductively in case studies.

Typological theories also receive more attention than their one-chapter allotment in the book would suggest. Such theories provide one way of modeling complex contingent generalizations. They frequently draw together in one framework the research of many social scientists, cumulating their individual efforts into a larger body of knowledge. The procedures we recommend for developing typological theories also foster the integration of within-case analyses and cross-case comparisons, and they help researchers opportunistically match up the types of case studies needed for alternative research designs and the extant cases that history provides. This helps to resolve the problem of case selection, one of the most challenging aspects of case study research designs. In addition, typological theories can guide researchers toward questions and research designs whose results will be pertinent to problems faced by policymakers. One of the chief goals of political science, as noted in Chapter 12, is to provide policymakers with "generic knowledge" that will help them form effective strategies.

Highly general and abstract theories ("covering laws," in Carl Hempel's term), which set aside intervening processes and focus on correlations between the "start" and "finish" of a phenomenon, are too general to make sharp theoretical predictions or to guide policy.⁶ For example, Kenneth Waltz's structural-realist theory, which posits that the material structure of the international system—the number of great powers, the balance of material power among them, the nature of contemporary military and economic technologies, and the geography of the system—creates structural incentives (such as the incentive to balance against other powerful states) that states can defy only at their peril. Though this theory dominated the field of international relations for some time, it is not a theory of foreign policy, as Waltz himself emphasizes, but a theory of constraints on foreign policy and of the predicted price to be paid for ignoring them.⁷

Theories that describe independent, stable causal mechanisms that

6. Here and in the next few paragraphs we draw directly from distinctions among covering laws, causal mechanisms, and typological theories suggested by David Dessler (private communication, January 7, 1998).

7. Kenneth Waltz, *Theory of International Politics* (New York: McGraw-Hill, 1979). For additional discussion, see Chapter 12.

under certain conditions link causes to effects also fail to provide specific guidance to those in search of policy guidance. For example, a theory may address the contribution a specific democratic norm makes to the fact that democracies have rarely fought one another—but without contingent generalizations on the conditions under which the norm is actualized and those under which it is overridden by other mechanisms, such a theory cannot tell policymakers whether they should, say, promote the adoption of this norm in newly democratic states. In contrast, middle-range typological theories, which identify recurring conjunctions of mechanisms and provide hypotheses on the pathways through which they produce results, provide more contingent and specific generalizations for policymakers and allow researchers to contribute to more nuanced theories. For example, one typological theory identifies subtypes of ways in which deterrence might fail: through a *fait accompli* or series of limited probes by a challenger, through a misperception of the adversary's will or capabilities, through the intrusion of domestic politics into decision-making, and so on.

The next section of this introduction discusses six reasons why we have undertaken the task of codifying case study practices and theory. We then offer a definition of case studies, outline their advantages and limitations, and conclude with a short discussion of the plan of the book.

Advances in Case Study Methods

The time seemed ripe to offer a book that would allow readers to view and assimilate advances and debates in case study methods and that might help these methods find wider use and acceptance. First, interest in theory-oriented case studies has increased substantially in recent years, not only in political science and sociology, but even in economics—arguably the most ambitious social science in its epistemological aspirations. Scholars in these and other disciplines have called for a “return to history,” arousing new interest in the methods of historical research and the logic of historical explanation, discussed in Chapter 10.⁸

Second, several developments in the philosophy of science in the past three decades, discussed in Chapter 7, have provided a firmer foundation for case study methods. In particular, the “scientific realist” school of thought has emphasized that causal mechanisms—independent stable factors that under certain conditions link causes to effects—are central to causal explanation. This has resonated with case study researchers’ use of process-tracing either to uncover evidence of causal mechanisms at work

8. See, for example, Terrence J. McDonald, ed., *The Historic Turn in the Human Sciences* (Ann Arbor: University of Michigan Press, 1996).

or to explain outcomes. We also find Bayesian logic useful in assessing how “tough” a test a particular case poses to a theory, and how generalizable the results are from a given case. This logic helps to refine Harry Eckstein’s discussion of using crucial, most-likely, and least-likely cases to test theories. A crucial case is one in which a theory that passes empirical testing is strongly supported and one that fails is strongly impugned. Since cases suitable for such doubly discriminating tests are rare, Eckstein emphasized the inferential value of instances where a theory fails to fit a case in which it is most likely to be true, and hence the theory is strongly undermined, or fits a case in which it is least likely to be true, and thus is convincingly supported.⁹

Third, we wished to engage contemporary debates among rational choice theorists, structuralists, historical institutionalists, social constructivists, cognitive theorists, postmodernists, and others, who at times may see themselves as having a stake in debates over case study or other methods. We argue that theoretical arguments are for the most part separable from methodological debates and that case study methods have wide applicability. For example, much of the early political science research on rational choice theories relied on formal models and statistical tests, but a growing number of rational choice theorists are realizing that case study methods can also be used in conjunction with or to test rational choice theories.¹⁰ Social constructivists, cognitive theorists, and historical institutionalists may welcome the comparative advantages of case studies for addressing qualitative variables, individual actors, decision-making processes, historical and social contexts, and path dependencies. Meanwhile, structuralists may worry that case studies are more amenable to these social and institutional theories than to materialist theories. We maintain, however, that case studies (as well as statistical and formal methods) are useful for theory development across all these schools of thought and that they can incorporate both material and ideational variables. Postmodernists will be skeptical of our aspiration to cumulative theoretical knowledge, but even they may find our version of case study methods useful in studying discourses, identities, and interactions systematically.

Fourth, there is growing interest across the social and physical sciences in modeling and assessing complex causal relations, such as path dependence, tipping points, multiple interactions effects, selection ef-

9. Harry Eckstein, “Case Studies and Theory in Political Science,” in Fred Greenstein and Nelson Polsby, eds., *Handbook of Political Science*, Vol. 7 (Reading, Mass.: Addison-Wesley, 1975), pp. 79–138. For further discussion, see Chapter 6.

10. An important example is Robert Bates et al., *Analytic Narratives* (Princeton, N.J.: Princeton University Press, 1998).

fects, disproportionate feedback loops, equifinality (many alternative causal paths to the same outcome), and multifinality (many outcomes consistent with a particular value of one variable). Case study methods, particularly when used in the development of typological theories, are good at exploring many of these aspects of complex causality.

Fifth, we found it necessary to address an imbalance in our field, and perhaps in others, between the mix of methods that we and our colleagues use in our own research and that which we teach to our students. Although almost half the articles published in the top political science journals in recent years used case studies, only about two-thirds of the thirty top-ranked graduate programs in political science offer a dedicated graduate course in qualitative or case study methods, and only two of these departments require such a course.¹¹ In contrast, all of the top thirty departments offer courses in statistics, and almost all of these departments require some training in statistics, often several courses. We believe that graduate students should be trained to produce cutting-edge research in their method of choice (which requires more courses for statistical methods than for qualitative methods) and to be critically aware consumers of research using the other two methods.¹² In this regard, this book is designed as a text for teaching students cutting-edge qualitative methods.

Finally, the publication of *Designing Social Inquiry: Scientific Inference in Qualitative Research (DSI)* by Gary King, Robert O. Keohane, and Sidney Verba has greatly influenced our field and usefully forced us to clarify our thinking on case study methods.¹³ We find much to agree with in this important work.¹⁴ At the same time, we find it necessary to qualify

11. Bennett, Barth, and Rutherford, "Do We Preach What We Practice?"

12. We do not necessarily expect individuals to do state-of-the-art work using more than one method in a single research project. There are examples in our field of exceptional and well-trained individuals doing excellent multi-method work, but while we want to encourage this practice, we do not want to set it as the standard expectation for Ph.D. theses, books, or articles. Since it is sufficiently difficult to do cutting-edge work with one method, we suspect most multi-method work will involve collaboration between researchers who are expert at different methods, a practice that deserves encouragement.

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

14. Some points of agreement involve fairly standard methodological admonitions: leave a clear and replicable record of your research methods, generate a list of observable implications for alternative hypotheses under consideration, specify what empirical findings would call into question each of these hypotheses, and keep in mind that

DSI's central argument that there is one "logic of inference." If this logic of inference refers in a broad sense to the epistemological logic of deriving testable implications from alternative theories, testing these implications against quantitative or case study data, and modifying theories or our confidence in them in accordance with the results, then perhaps on a very general level there is one logic that is the modern successor of the still-evolving positivist tradition, although many disagreements remain about particular aspects of this logic.¹⁵

If, however, the logic of inference refers to specific methodological injunctions on such issues as the value of single-case studies, the procedures for choosing which cases to study, the role of process-tracing, and the relative importance of causal effects (the expected change in the dependent variable given a unit change in an independent variable) and causal mechanisms as bases for inference and explanation, as *DSI* appears to argue, then we disagree with the overall argument as well as some of the methodological advice *DSI* provides to case study researchers on these issues. *DSI* risks conflating these epistemological and methodological logics by stating that "the same underlying logic provides the framework for each research approach. This logic tends to be explicated and formalized clearly in discussions of quantitative research methods."¹⁶

We take up our disagreements with *DSI* (in this chapter and in Chapter 8); here, for reference, we merely list them, starting with our epistemological differences and proceeding to our methodological ones. One critique is that although *DSI* disavows being "a work in the philosophy of the social sciences," it implicitly makes many important philosophical assumptions regarding highly contested issues in the philosophy of science.¹⁷ For example, *DSI* suggests that causal mechanisms are in some sense less fundamental to causal explanation than what *DSI*

science is a social enterprise in which no research is perfect and diversity of belief serves as a useful check on individual misperceptions and biases. We also agree that counterfactual analysis can serve as a useful cross-check on theorizing, that reconfiguring one's theory after seeing some of the data is defensible as long as it leads to new predictions on other data that hold up to additional empirical tests, and that parsimonious theories are desirable but should not be pursued at the cost of oversimplifying a complex world and reducing our ability to produce rich explanations. Most of these points are raised in King, Keohane, and Verba, *Designing Social Inquiry*, pp. 7–33.

15. Postmodernists would of course disagree with us and with *Designing Social Inquiry* on the applicability of positivist logic, even broadly construed, to the social sciences.

16. King, Keohane, and Verba, *Designing Social Inquiry*, p. 3.

17. *Ibid.*, p. 3.

1 L SHORT

REGULAR

Long

JOB# 997

Georg

defines as “causal effects.”¹⁸ This runs counter to our view that causal mechanisms and causal effects are equally important to causal explanation. More generally, in our view *DSI*’s treatment of causal mechanisms is unsatisfactory, as we detail in Chapter 8. Robert Keohane has given a clearer exposition of the nature and importance of causal mechanisms for explanation in his later publications.¹⁹

We also critique *DSI* for emphasizing almost exclusively the epistemic goal of hypothesis testing (sometimes known as the “logic of confirmation”), neglecting other aspects of theory development, such as the formation of new hypotheses or the choice of new questions to study. *DSI* relegates these goals, the “logic of discovery,” to a quotation from Karl Popper that “there is no such thing as a logical method of having new ideas. . . . Discovery contains ‘an irrational element,’ or a ‘creative intuition.’”²⁰

We agree that there is no linear logic of discovery, but we emphasize theory development, focusing on hypothesis formation and the historical explanation of individual cases, as well as the testing of general hypotheses. We outline procedures that are conducive to the generation of new hypotheses, such as the study of deviant or outlier cases.

Another concern is that *DSI* pays little attention to problems of causal complexity, particularly equifinality and multiple interactions effects. It addresses these subjects very briefly, discussing only the simple case of two-variable interactions, and it tends to be optimistic on how easily statistical models can address complex interactions within a realistic sample size.²¹ We emphasize that various kinds of complex causal relations are central concerns of the social sciences, including not only equifinality and multiple interactions effects, but also disproportionate feedback loops, path dependencies, tipping points, selection effects, expectations effects,

18. Ibid., pp. 85–87.

19. Robert Keohane, “Problematic Lucidity: Stephen Krasner’s ‘State Power and the Structure of International Trade,’” *World Politics*, Vol. 50, No. 1 (October 1997), pp. 150–170; and his unpublished paper that focuses on the importance of causal mechanisms in efforts to explain the extinction of dinosaurs: “Dinosaurs, Detectives and Causal Mechanisms: Coping With Uniqueness in Social Science Research,” paper presented at American Political Science Association Annual Meeting, September 4, 1999, Atlanta, Georgia. In the latter paper, Keohane also concedes that “[Ronald] Rogowski was right [in his *American Political Science Review* (June 1995) critique] to criticize *Designing Social Inquiry* for not emphasizing sufficiently the importance of elaboration of models and the deduction of implications from them.” Keohane also notes that he and his co-authors of *Designing Social Inquiry* “implicitly recognized the importance of theory . . . but we certainly did not emphasize it enough.”

20. King, Keohane, and Verba, *Designing Social Inquiry*, p. 14.

21. Ibid., pp. 85–87.

and sequential interactions between individual agents and social structures. Our approach to the problem of complexity is to recommend process-tracing as a means of examining complexity in detail and to suggest typological theorizing as a way to model complexity; *DSI* does not distinguish between typological theories, which model causal relations of equifinality, and mere taxonomical typologies.²²

On the methodological level, we take issue with *DSI*'s arguments on case selection criteria, the value of single-case studies and "no variance" research designs, the costs and benefits of increasing the number of cases studied, and the role of process-tracing. On case selection criteria, *DSI* gives the standard statistical warnings about selection on the dependent variable and argues that single-case research designs are seldom valuable.²³ This advice overlooks the opportunities for studying deviant cases and the dangers of certain forms of selection bias in case studies that can be more severe than those in statistical studies.

DSI also argues for increasing the number of observable implications of a theory both within cases and across them. While we agree that increasing the number and diversity of observable implications of alternative theories is generally extremely useful, *DSI* tends to understate the dangers of "conceptual stretching" that can arise if the means of increasing observations include applying theories to new cases, changing the measures of variables, or both. *DSI* acknowledges, for example, that additional cases to be studied must be "units within which the process entailed by the hypothesis can take place," but it does not cite here or elsewhere Giovanni Sartori's well-known article on the subject of conceptual stretching.²⁴

We also disagree with *DSI*'s treatment of process-tracing as simply another means of increasing the number of observable implications of a theory. In fact, process-tracing is fundamentally different from statistical analysis because it focuses on sequential processes within a particular historical case, not on correlations of data across cases. This has important implications for theory testing: a single unexpected piece of process-tracing evidence can require altering the historical interpretation and theoretical significance of a case, whereas several such cases may not greatly alter the findings concerning statistical estimates of parameters for a large population.

DSI's arguments on all these methodological issues may be appropri-

22. Ibid., p. 48.

23. Ibid., pp. 129–132, 210–211.

24. Ibid., p. 221; and Giovanni Sartori, "Concept Misformation in Comparative Politics," *American Political Science Review*, Vol. 64, No. 4 (December 1970), pp. 1033–1053.

ate to statistical methods, but in our view they are ill-suited or even counterproductive in case study research. We differ, finally, with *DSI* on a presentational issue that is primarily pedagogical but has important implications. This is the fact that there is an unresolved tension in *DSI* between the authors' emphasis on research objectives that address important theoretical and policy-relevant problems and the fact that many of the examples used to illustrate various points in *DSI* are either hypothetical or entail research objectives of a simple character not likely to be of interest to sophisticated research specialists.

This gap is aggravated by the fact that many of the hypothetical and actual examples are of quantitative, not qualitative research. *DSI* cites very few qualitative research studies that in its authors' view fully or largely meet the requirements of its methods or deserve emulation, nor do the authors cite their own works in this regard.²⁵ This is not surprising

25. Five earlier publications by Gary King are listed in the bibliographical references, but the single reference to "King" in the index refers to one of his large-N statistical studies, described on page 189. Only one article by Sidney Verba is listed in the bibliographical references ("Some Dilemmas of Political Research," *World Politics*, Vol. 20, No. 1 (October 1967), pp. 111–127), and the only references under "Verba" in the index refer to a large-N statistical study later published as Sidney Verba, Kay Lehman Schlozman, and Harry Brady, *Voices and Equality: Civic Voluntarism in American Politics* (Cambridge, Mass.: Harvard University Press, 1995). It is referred to briefly in the text to indicate that parsing an explanatory variable can avoid the problem of bias due to endogeneity (pp. 193–195), and later to provide "an example of seeking additional observable implications of one's hypotheses" by working with subunits of a national state (pp. 220–221).

Five earlier publications by Robert Keohane are listed in the bibliographical references at the end of the book, but there is no discussion of these works in the *DSI* text as examples of the methods recommended in the book. Keohane's detailed introductory essay for a subsequent collaborative small-n study he coedited (Robert O. Keohane and Marc A. Levy, eds., *Institutions for Environmental Aid: Pitfalls and Promise* (Cambridge, Mass.: MIT Press, 1996)), published two years after *Designing Social Inquiry*, makes no reference to identifying "observable implications" of the theories examined in the book. This small-n study employs procedures closely resembling those of the method of structured, focused comparison and process-tracing. Thus, Keohane writes, the case studies in the book "are written according to a common analytical format to ensure consistency and a comparability across cases. . . . We have insisted on such a systematic approach for two reasons: (1) to ensure that each chapter [reporting a case study] systematically considers the *sequence of action* relevant to the effectiveness of financial transfers, from explanatory and evaluative standpoints as well as descriptively, and (2) to facilitate a process of drawing out *generalizations across cases, about conditions for success and failure* of financial transfers and mechanisms" (pp. 16–17; emphasis added).

In correspondence with Alexander L. George (April 8, 2003), Robert Keohane acknowledged that two students whose dissertations he supervised, Vinod Aggarwal and Lisa Martin, both employed process-tracing to establish the possibility of a causal

since both Gary King and Sidney Verba are quantitatively oriented researchers. On the other hand, Robert Keohane's voluminous research is largely of a qualitative character and, surprisingly, none of his previous studies are cited in *Designing Social Inquiry* as examples of the methods advocated therein.²⁶

In contrast, in the present volume we present numerous examples of qualitative research on important policy-relevant problems, including research we ourselves have done. We do so not to imply that our own or others' work is methodologically flawless or worthy of emulation in every respect, but because the hardest methodological choices arise in actual research. Illustrating how such choices are made is vitally important in teaching students how to proceed in their own work. In addition, understanding methodological choices often requires sophisticated familiarity with the theories and cases in question, which reinforces the usefulness of using one's own research for examples.

Certainly King, Keohane, and Verba deserve the fullest praise and appreciation for their effort to improve qualitative research. *DSI*, despite our many disagreements with it, remains a landmark contribution. It is not alone in viewing the goals, methods, and requirements of case studies partly from the viewpoint of statistical methods. We choose to critique *DSI* in such detail not because it is the starkest example of this phenomenon, but because its clarity, comprehensiveness, and familiarity to many scholars make it an excellent vehicle for presenting our contrasting view of the differences, similarities, and comparative advantages of case study and statistical methods. In the next sections we define case studies and outline their advantages, limitations, and trade-offs, distinguishing between criticisms that in our view misapply statistical concepts and critiques that have real merit regarding the limits of case studies.

chain linking independent and dependent variables. (Aggarwal's use of process-tracing is described in Chapter 9; Lisa Martin's in the Appendix, "Studies That Illustrate Research Design.") Keohane graciously added that he recognized the importance of process-tracing used for this purpose.

26. The impact *DSI* has had on qualitative research in the social sciences since its publication in 1994 has not, so far as we know, been systematically assessed. Certainly the book has been widely read and consulted. In response to a question concerning *DSI*'s impact (letter to Alexander L. George, April 27, 2003), Robert Keohane notes that the book's advice about observable implications has caught on in much of the field. He cites a number of specific papers by Mark Pollock and Erica Gould who have cited *DSI*'s emphasis on observable implications. Dan Nielson and Michael Tierney, "Delegation to International Organizations: Agency Theory and World Bank Environmental Reform," *International Organization*, Vol. 57, No. 2 (April 2003), pp. 241–276, does the same. The clearest published reference to *DSI* on this score appears in Lisa Martin's *Democratic Commitments* (Princeton, N.J.: Princeton University Press, 2000), p. 9.

1 L SHORT

REGULAR

Long

JOB# 997

George

A major new reassessment of *Designing Social Inquiry* has been provided by a team of specialists in a book edited by Henry Brady and David Collier,²⁷ in which the editors integrate their respective specializations in quantitative survey research and qualitative comparative studies. Their book provides a major scholarly statement on the relationship between quantitative and qualitative methods. While generous and specific in its praise for contributions made in *DSI*, Brady, Collier, and the contributors to their volume express major misgivings: First, *DSI*, “does not adequately address the basic weaknesses within the mainstream quantitative approach it advocates.” Second, *DSI*’s “treatment of concepts, operationalizations, and measurement” is regarded as “seriously incomplete.” Third, Brady and Collier “disagree with *DSI*’s claims that it provides a general framework for ‘specific inference in qualitative research.’” They emphasize, as others have, *DSI*’s “failure to recognize the distinctive strengths of qualitative methods,” which leads its authors to “inappropriately view qualitative analysis almost exclusively through the optic of mainstream quantitative methods.”

The present book has much in common with Brady and Collier’s book. They emphasize, as we do, the need to “rethink the contributions” of quantitative and qualitative approaches and to indicate how scholars can most effectively draw on the respective strengths of each. Considerable attention is given to our emphasis on the importance of within-case analysis and process-tracing. Brady and Collier, and other distinguished scholars contributing to their book, share our criticism of *DSI*’s almost exclusive focus on increasing the number of observations in order to increase “leverage.” In their conclusion, Brady, Collier, and Jason Seawright develop a “multi-faceted approach to evaluating sources of leverage for addressing rival explanations.”

Despite these important reservations and criticisms, Brady and Collier, as do we, regard *DSI* as a major contribution that has usefully stimulated important new work on the relation between quantitative and qualitative methods.

27. Henry E. Brady and David Collier, eds., *Rethinking Social Inquiry: Diverse Tools, Shared Standards* (Lanham, Md.: Rowman and Littlefield, 2004). The quotations that follow are from the uncorrected proof of the Preface, Chapter 1, and Chapter 13. Their book reprints important articles that have expressed criticism of *Designing Social Inquiry* by Larry Bartels, Ronald Rogowski, David Collier, James Mahoney, James Seawright; Gerardo Munck, Charles C. Ragin, Timothy J. McKeown, and Sidney Tarrow. It also reprints the reply made by Gary King, Robert O. Keohane, and Sidney Verba to comments on *Designing Social Inquiry* made in the symposium on their book in the *American Political Science Review*, Volume 89, Number 2 (June 1995), pp. 475–481.

Advantages and Limitations of Case Studies: Casting Off the Prism of Statistical Methods

In the 1960s and 1970s, definitions of case studies relied on distinctions between the study of a small versus a large number of instances of a phenomenon. Case studies were characterized as “small-n” studies, in contrast to “large-N” statistical studies. This distinction suggests that the difference in the number of cases studied is the most salient difference between statistical and case study methods; in our “bigger is better” culture, this language implies that large-N methods are always preferable when sufficient data is available for study, as Arend Lijphart implied in a 1971 article.²⁸ In fact, case studies and other methods each have particular advantages in answering certain kinds of questions.

One early definition, still widely used, states that a case is a “phenomenon for which we report and interpret only a single measure on any pertinent variable.”²⁹ This definition, which case study researchers have increasingly rejected, has sometimes led scholars trained in statistical methods to misapply the “degrees of freedom problem” (which we discuss below) and to conclude that case studies provide no basis for evaluating competing explanations of a case.

We define a case as an instance of a class of events.³⁰ The term “class of events” refers here to a phenomenon of scientific interest, such as revolutions, types of governmental regimes, kinds of economic systems, or

28. Arend Lijphart, “Comparative Politics and the Comparative Method,” *American Political Science Review*, Vol. 65 (September 1971), pp. 682–693.

29. Eckstein, “Case Studies and Theory in Political Science,” p. 85. King, Keohane, and Verba reject the term “case” as subject to too many uses and substitute “observations” for “cases” (*Designing Social Inquiry*, p. 52), but this leads to ambiguity as well. See, for example, *DSI*’s discussion of whether Eckstein viewed cases as having single or multiple observations, pp. 210–211. Our reading is that Eckstein envisioned multiple process-tracing observations in each case study despite his definition of a case as having one measure of the dependent variable. (In our view it is more precise to speak of one instance of the dependent variable, which may have several qualitative measures.) A full discussion of *DSI*’s advice on how to generate additional observable implications of a theory is presented in Chapter 8.

30. Alexander L. George, “Case Studies and Theory Development,” paper presented at Carnegie-Mellon University, October 15–16, 1982, p. 45. For a similar definition by sociologists, see Charles Ragin’s, “Introduction” in Charles Ragin and Howard Becker, eds., *What is a Case? Exploring the Foundations of Social Inquiry* (New York: Cambridge University Press, 1992), pp. 1–3. In the concluding chapter (“‘Casing’ and the Process of Social Inquiry”), Ragin emphasizes the importance for theory development of focusing research on specific subclasses of a phenomena, which he calls “casing” (pp. 217–226).

personality types that the investigator chooses to study with the aim of developing theory (or “generic knowledge”) regarding the causes of similarities or differences among instances (cases) of that class of events. A case study is thus a well-defined aspect of a historical episode that the investigator selects for analysis, rather than a historical event itself. The Cuban Missile Crisis, for example, is a historical instance of many different classes of events: deterrence, coercive diplomacy, crisis management, and so on.³¹ A researcher’s decision about which class of events to study and which theories to use determines what data from the Cuban Missile Crisis are relevant to her or his case study of it.³² Questions such as “what is this event a case of?” and “is this event a designated phenomenon?” are integral to selecting cases for study and designing and implementing research of these cases.³³

There is potential for confusion among the terms “comparative methods,” “case study methods,” and “qualitative methods.” In one view, the comparative method (the use of comparisons among a small number of cases) is distinct from the case study method, which in this view involves the internal examination of single cases. However, we define case study methods to include both within-case analysis of single cases and comparisons of a small number of cases, since there is a growing consensus that the strongest means of drawing inferences from case studies is the use of a combination of within-case analysis and cross-case comparisons within a single study or research program (although single-case studies can also play a role in theory development). The term “qualitative methods” is sometimes used to encompass both case studies carried out with a relatively positivist view of the philosophy of science and those implemented with a postmodern or interpretive view. We exclude postmodern narra-

31. The Cuban Missile Crisis is treated as a case of deterrence failure in Alexander L. George and Richard Smoke, *Deterrence in American Foreign Policy: Theory and Practice* (New York: Columbia University Press, 1974); as a case of coercive diplomacy in Alexander L. George, David K. Hall, and William E. Simons, *The Limits of Coercive Diplomacy* (Boston: Little, Brown, 1971); and as a case of crisis management in Ole R. Holsti, *Crisis, Escalation, War* (Montreal: McGill-Queen’s University Press, 1972).

32. It is important to note that the definition of which independent variables are relevant to the class of events remains open to revision as the research proceeds. In conducting interviews, reading secondary accounts, or reviewing historical documents, the researcher may inductively discover independent variables that previous theories may have overlooked. This inductive side to identifying variables is open also to statistical researchers who are constructing their own data sets from primary and secondary sources, but it is closed to statistical studies that rely on existing data sets, as well as to the purely deductive development of formal models.

33. Ragin and Becker, eds., *What is a Case?*; and David Collier, “Translating Quantitative Methods for Qualitative Researchers: The Case of Selection Bias,” *American Political Science Review*, Vol. 89, No. 2 (June 1995), pp. 461, 465.

tives from our view of case studies, though some of the more disciplined forms of discourse analysis approach our view of case studies. This book therefore hews to the traditional terminology in focusing on case studies as the subset of qualitative methods that aspires to cumulative and progressive generalizations about social life and seeks to develop and apply clear standards for judging whether some generalizations fit the social world better than others.

Strengths of Case Study Methods

Case studies are generally strong precisely where statistical methods and formal models are weak. We identify four strong advantages of case methods that make them valuable in testing hypotheses and particularly useful for theory development: their potential for achieving high conceptual validity; their strong procedures for fostering new hypotheses; their value as a useful means to closely examine the hypothesized role of causal mechanisms in the context of individual cases; and their capacity for addressing causal complexity.

CONCEPTUAL VALIDITY

Case studies allow a researcher to achieve high levels of conceptual validity, or to identify and measure the indicators that best represent the theoretical concepts the researcher intends to measure. Many of the variables that interest social scientists, such as democracy, power, political culture, state strength, and so on are notoriously difficult to measure. For example, a procedure that is “democratic” in one cultural context might be profoundly undemocratic in another. Thus, researchers must carry out “contextualized comparison,” which “self-consciously seeks to address the issue of equivalence by searching for *analytically equivalent* phenomena—even if expressed in substantively different terms—across different contexts.”³⁴ This requires a detailed consideration of contextual factors, which is extremely difficult to do in statistical studies but is common in case studies.

Whereas statistical studies run the risk of “conceptual stretching” by lumping together dissimilar cases to get a larger sample, case studies allow for conceptual refinements with a higher level of validity over a smaller number of cases. Research in comparative politics on democratic systems, for example, has proceeded in part through the conceptual development of “democracy with adjectives,” where each adjective, such as

34. Richard Locke and Kathleen Thelen, “Problems of Equivalence in Comparative Politics: Apples and Oranges, Again,” *American Political Science Association: Comparative Politics Newsletter*, No. 8 (Winter 1998), p. 11.

a “federal,” “parliamentary,” “presidential,” or “authoritarian” democracy, denotes a subtype or subclass with a smaller number of cases that are presumably more similar than those under the overall concept of “democracy.”³⁵ A common path of theoretical development has been from broad generalizations, such as the “democratic peace” theory (which argues that democracies are less war-prone) into more contingent generalizations (such as the “interdemocratic peace” theory, which holds that democracies rarely fight other democracies; see Chapter 2). Often, when such phenomena are examined in more detail, they prove to exhibit “equifinality”; that is, they involve several explanatory paths, combinations, or sequences leading to the same outcome, and these paths may or may not have one or more variables in common.

Consequently, statistical research is frequently preceded by case study research to identify relevant variables and followed by case study work that focuses on deviant cases and further refines concepts.³⁶ For example, after a range of statistical studies suggested that democracies do not fight other democracies, case study researchers started to explore which aspects of democracy—democratic values, democratic institutions, the transparency of decision-making in democracies, and so on—might be responsible for this apparent “democratic peace.” Should these case studies indicate, say, that transparency is an important causal factor whereas universal suffrage is not, then revised and new statistical tests are performed.

DERIVING NEW HYPOTHESES

Case studies have powerful advantages in the heuristic identification of new variables and hypotheses through the study of deviant or outlier cases and in the course of field work—such as archival research and interviews with participants, area experts, and historians. When a case study researcher asks a participant “were you thinking X when you did Y,” and gets the answer, “No, I was thinking Z,” then if the researcher had not thought of Z as a causally relevant variable, she may have a new variable demanding to be heard. The popular refrain that observations are

35. David Collier and Steven Levitsky, “Democracy with Adjectives: Conceptual Innovation in Comparative Research,” *World Politics*, Vol. 49, No. 3 (April 1997), pp. 430–451. Collier and Levitsky also note the use of “diminished subtypes,” or cases that lack a few attributes of the overall concept, such as “limited suffrage democracies” (pp. 437–442).

36. David Collier, “Comparative Historical Analysis: Where Do We Stand?” *American Political Science Association: Comparative Politics Newsletter*, No. 10 (Winter 1999), pp. 1–6.

theory-laden does not mean that they are theory-determined. If we ask one question of individuals or documents but get an entirely different answer, we may move to develop new theories that can be tested through previously unexamined evidence.

Statistical methods can identify deviant cases that may lead to new hypotheses, but in and of themselves these methods lack any clear means of actually identifying new hypotheses. This is true of all studies that use existing databases or that modify such databases only slightly or without recourse to primary sources. Unless statistical researchers do their own archival work, interviews, or face-to-face surveys with open-ended questions in order to measure the values of the variables in their model, they have no unproblematic inductive means of identifying left-out variables. Even statistical methods of “data mining” necessarily include only those variables that a researcher has already thought to code into a data base. Deductive theorizing can also identify new variables, but with the exception of purely deductive theories, inductive field research methods typically lie behind every newly identified variable.

EXPLORING CAUSAL MECHANISMS

Case studies examine the operation of causal mechanisms in individual cases in detail. Within a single case, we can look at a large number of intervening variables and inductively observe any unexpected aspects of the operation of a particular causal mechanism or help identify what conditions present in a case activate the causal mechanism. Our definition of causal mechanism (see Chapter 7) notes that such mechanisms operate only under certain conditions. Statistical studies, which omit all contextual factors except those codified in the variables selected for measurement or used for constituting a population of cases, necessarily leave out many contextual and intervening variables.

Researchers can also use theories on causal mechanisms to give historical explanations of cases. Historical explanation is quite different from the development and testing of variable-centered theories from the statistical study of a large number of cases. As statistical researchers frequently point out, correlation does not imply causation. If a prosecutor knows on the basis of criminological studies that 90 percent of acts of arson are perpetrated by the owner of the property that is burned down, this is not sufficient to convict a particular property owner of arson. The prosecutor needs to empirically establish that means, motive, and opportunity existed in this particular case. Ideally, the prosecutor will construct a complete and uninterrupted chain of evidence to establish how the specific crime may have been done by the particular individual accused, using forensic theories to bolster each point in the chain.

MODELING AND ASSESSING COMPLEX CAUSAL RELATIONS

A final advantage of case studies is their ability to accommodate complex causal relations such as equifinality, complex interactions effects, and path dependency.³⁷ This advantage is relative rather than absolute. Case studies can allow for equifinality, but to do so they produce generalizations that are narrower or more contingent. We find great value in such middle-range theories, but others may prefer theories that are more general even if this necessarily means they are more vague or more prone to counterexamples. Case studies also require substantial process-tracing evidence to document complex interactions. Analogously, statistical methods can model several kinds of interactions effects, but only at the cost of requiring a large sample size, and models of nonlinear interactions rapidly become complex and difficult to interpret. New statistical methods may be able to improve upon the statistical treatment of equifinality and interactions effects.³⁸

Trade-offs, Limitations, and Potential Pitfalls of Case Studies

It is important to distinguish among the recurrent trade-offs, inherent limits, and examples of poor implementation of case study methods and not to misinterpret these aspects through the prism of statistical methods, as has been done in the past. Recurrent trade-offs include the problem of case selection; the trade-off between parsimony and richness; and the related tension between achieving high internal validity and good historical explanations of particular cases versus making generalizations that apply to broad populations. The inherent limitations include a relative inability to render judgments on the frequency or representativeness of particular cases and a weak capability for estimating the average “causal effect” of variables for a sample. Potential limitations can include indeterminacy and lack of independence of cases.

CASE SELECTION BIAS

One of the most common critiques of case study methods is that they are particularly prone to versions of “selection bias” that concern statistical researchers.³⁹ Selection biases are indeed a potentially severe problem in case study research, but not in the same ways as in statistical research.

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

38. Bear F. Braumoeller, “Causal Complexity and the Study of Politics,” (unpublished manuscript, Harvard University, Cambridge, Mass., 2002).

39. Christopher H. Achen and Duncan Snidal, “Rational Deterrence Theory and Comparative Case Studies,” *World Politics*, Vol. 41, No. 2 (January 1989), p. 160; and

Selection bias, in statistical terminology, “is commonly understood as occurring when some form of selection process in either the design of the study or the real-world phenomena under investigation results in inferences that suffer from systematic error.”⁴⁰ Such biases can occur when cases or subjects are self-selected or when the researcher unwittingly selects cases that represent a truncated sample along the dependent variable of the relevant population of cases.⁴¹ If for some reason a statistical researcher has unwittingly truncated the sample of cases to be studied to include only those whose dependent variable is above or below an extreme value, then an estimate of the regression slope for this truncated sample will be biased toward zero. In other words, in statistical studies selection bias always understates the strength of the relationship between the independent and dependent variables. This is why statistical researchers are admonished not to select cases on the dependent variable.⁴²

In contrast, case study researchers sometimes deliberately choose cases that share a particular outcome. Practitioners and analysts of case study methods have argued that selection on the dependent variable should not be rejected out of hand. Selection of cases on the basis of the value of their dependent variables is appropriate for some purposes, but not for others. Cases selected on the dependent variable, including single-case studies, can help identify which variables are not necessary or sufficient conditions for the selected outcome.⁴³

In addition, in the early stages of a research program, selection on the dependent variable can serve the heuristic purpose of identifying the potential causal paths and variables leading to the dependent variable of interest. Later, the resulting causal model can be tested against cases in which there is variation on the dependent variable.⁴⁴ Ideally, researchers

Barbara Geddes, “How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics,” *Political Analysis*, Vol. 2 (1990), pp. 131–150.

40. David Collier and James Mahoney, “Insights and Pitfalls: Selection Bias in Qualitative Work,” *World Politics*, Vol. 49, No. 1 (October 1996) p. 59.

41. Ibid., p. 60; and King, Keohane, and Verba, *Designing Social Inquiry*, pp. 128–132.

42. Collier and Mahoney, “Insights and Pitfalls,” p. 60.

43. Douglas Dion, “Evidence and Inference in the Comparative Case Study,” in Gary Goertz and Harvey Starr, eds., *Necessary Conditions: Theory, Methodology, and Applications* (Lanham, Md.: Rowman and Littlefield, 2003), pp. 95–112; and Collier, “Translating Quantitative Methods for Qualitative Researchers,” p. 464.

44. Case study researchers in many instances should make comparisons between the subset of cases or types studied and the larger population, where there is more variance on the dependent variable (Collier and Mahoney, “Insights and Pitfalls,” p. 63). Sometimes, such comparisons can be made to existing case studies in the literature, or the researcher might include “mini-case” studies, or less in-depth studies, of a wide number of cases in addition to full studies of the cases of greatest interest. To say

would like to have the functional equivalent of a controlled experiment, with controlled variation in independent variables and resulting variation in dependent variables, but the requisite cases for such research designs seldom exist.

A related issue is whether researchers' foreknowledge of the values of variables in cases—and perhaps their cognitive biases in favor of particular hypotheses—necessarily bias the selection of case studies.⁴⁵ Selection with some preliminary knowledge of cases, however, allows much stronger research designs; cases can be selected with a view toward whether they are most-likely, least-likely, or crucial for a theory, making the process-tracing test of a theory more severe. Also, within-case analysis often leads to the finding that the researcher's (or the literature's) preliminary knowledge of the values of the independent and dependent variables was incomplete or simply wrong, and case study researchers sometimes conclude that none of the proposed theories adequately explains a case. In addition, researchers selecting cases can benefit from knowledge of the findings of existing studies, and be guided by estimations of whether the theories of interest are strong and previously tested or new and relatively weak.⁴⁶ There are also methodological safeguards against investigator-induced bias in case studies, such as careful congruence testing and process-tracing.

Interestingly, statistical views of selection bias understate both the most severe and the most common kinds of selection biases in qualitative research. The most damaging consequences arise from selecting only cases whose independent *and* dependent variables vary as the favored hypothesis suggests, ignoring cases that appear to contradict the theory, and overgeneralizing from these cases to wider populations. This type of selection bias can occur even when there is variation in both independent and dependent variables and this variation covers the full range of values that these variables can assume.

Rather than *understating* the relationship between independent and dependent variables, as in the statistical view of selection bias, this selection bias can *understate or overstate* the relationship.⁴⁷ While this form of

that such comparisons are often useful for many research goals, however, is very different from arguing that they are always necessary for all research goals.

45. The standard protection against this bias in statistical studies is random selection, but as King, Keohane, and Verba note (*Designing Social Inquiry*, pp. 124–127), in studies of a small number of cases, random selection can be more likely to result in bias than intentional selection.

46. David Laitin, "Disciplining Political Science," *American Political Science Review*, Vol. 89, No. 2 (June 1995), p. 456.

47. Collier and Mahoney, "Insights and Pitfalls," pp. 71–72.

selection bias seems too obvious to require a warning to social scientists, case researchers may fail to realize that by implicitly or explicitly limiting their sample of cases (say, to history that is contemporary, Western, specific to one country, or easily researchable), they may bias their sample with regard to a wider set of cases about which they are trying to make inferences—unless they carefully define and limit the *scope* of their findings to a well-specified population that shares the same key characteristics as the cases studied.

This form of selection bias is far more common in political argumentation than in social science case studies. Several other case selection biases, however, are quite common in case study research and deserve increased attention. These include selection of a cases based on their “intrinsic” historical importance or on the accessibility of evidence.

IDENTIFYING SCOPE CONDITIONS AND “NECESSITY”

A limitation of case studies is that they can make only tentative conclusions on *how much* gradations of a particular variable affect the outcome in a particular case or how much they generally contribute to the outcomes in a class or type of cases. Case studies are much stronger at identifying the scope conditions of theories and assessing arguments about causal necessity or sufficiency in particular cases than they are at estimating the generalized causal effects or causal weight of variables across a range of cases. More confident estimates of causal effects, the equivalent of beta coefficients in statistical studies, are possible in case studies only when there is a very well-controlled before-after case comparison in which only one independent variable changes, or more generally when extremely similar cases differ only in one independent variable. Otherwise, case studies remain much stronger at assessing *whether* and *how* a variable mattered to the outcome than at assessing *how much* it mattered.

Methodologists are working to reduce this limitation, however. Douglas Dion, for example, has focused on the role of case studies in testing theoretical claims that a variable is a necessary or sufficient condition for a certain outcome.⁴⁸ Dion convincingly argues that selection bias is not a problem in tests of necessity or sufficiency, that single counterexamples can falsify deterministic claims of necessity or sufficiency (if

48. The reader should note that any necessary condition can be inverted and stated as a sufficient condition, and vice versa. To say that “A is necessary for B” (for a specified population or set of scope conditions) is the same as saying “the absence of A is sufficient for the absence of B” (for the specified population or scope conditions). Thus from a methodological point of view any discussion of testing a necessary condition can be restated in terms of testing a sufficient condition, and vice versa.

measurement error can be ruled out), and that only small numbers of cases are required to test even probabilistic claims that a condition is almost always necessary or sufficient for an outcome.⁴⁹ These factors make case studies a powerful means of assessing claims of necessity or sufficiency.

It is important to distinguish carefully, however, among three kinds of claims of necessity or sufficiency. The most general claim would be that a single variable is necessary or sufficient for an outcome with respect to an entire population of cases. Unfortunately, few nontrivial single-variable relationships of necessity or sufficiency have been found to hold for large populations or wide-scope conditions in the social world. A second kind of claim is that a variable was either necessary or sufficient in a particular historical context or case for a specific historical outcome to have occurred. This kind of claim can only be tested counterfactually, and there is no infallible means of making such counterfactual tests.

The third and in our view most useful kind of assertion of necessity or sufficiency concerns the relationship of a variable to *conjunctions* of variables that are themselves necessary and/or sufficient for an outcome. Consider the following example. Let us assume that the variable *A* causes *Y* only in conjunction with *B* and *C*. Assume further that the conjunction *ABC* is sufficient for *Y*, and the conjunction *BC* cannot cause *Y* in the absence of *A*. In this instance, *A* is a necessary part of a conjunction that is sufficient for the outcome *Y*. Many different possible combinations of conjunctive necessity and sufficiency are possible. If equifinality is present, for example, the conjunction *ABC* itself may not be necessary for the outcome, which might arise through other causal paths that have little or nothing in common with *ABC*.⁵⁰

Three caveats are in order regarding inferences of necessity or sufficiency. First, it is often not possible to resolve whether a causal condition identified as contributing to the explanation of a case is a *necessary*

49. See Dion, "Evidence and Inference," pp. 95–112.

50. One further variation on methods for assessing necessity and sufficiency is Charles Ragin's suggestion for using "fuzzy set" techniques to examine theories that make probabilistic assertions about conditions that are "almost always" or "usually" necessary or sufficient. Such relationships might be more commonly observed than deterministic relationships of necessity or sufficiency because measurement error, as well as the possibility of an irreducibly random element in human affairs, can never be conclusively eliminated. Charles Ragin, *Fuzzy-Set Social Science* (Chicago, Ill.: University of Chicago Press, 2000). We address Ragin's "fuzzy set methods" in Chapter 8 on comparative methods. We explore further in Chapter 11 some of the different types and implications of conjunctive conditions.

condition for that case, for the type of case that it represents, or for the outcome in general. It is often more appropriate to settle for a defensible claim that the presence of a variable “favors” an outcome, or is what historians often term a “contributing cause,” which may or may not be a necessary condition. When a complex explanation identifies a number of contributing causes, it may be difficult, even with the help of counterfactual analysis, to offer a convincing argument that one condition or another was necessary to the outcome.

Second, whether a factor is necessary to an outcome in a case is a separate issue from *how much* it contributed to the magnitude of the outcome. One “last straw” may be necessary to break a camel’s back, but it does not contribute as much to the outcome as the bales of straw that preceded it. As noted above, determining such relative causal weights for variables can be difficult to do with any precision in a single case or a small number of cases, but process-tracing evidence and congruence tests can provide useful evidence on this question.

Third, even when a plausible argument can be made that a factor is necessary to the outcome in a particular case, this does not automatically translate into a general claim for its causal role in other cases. If equifinality is present, the factor’s necessity and causal weight may vary considerably across cases or types of cases.⁵¹

51. A complex debate has emerged over whether, as Dion argues, claims of necessity should be tested only against cases that are positive on the outcome of interest, and claims of sufficiency should be tested only against cases that are positive on the independent variable of interest. Jason Seawright has used sophisticated Bayesian reasoning to argue, contra Dion, that studying diverse cases, in addition to looking for single cases that could disprove necessity or sufficiency, can yield stronger and more efficient tests of necessity and sufficiency; others dispute this. See Jason Seawright, “Testing for Necessary and/or Sufficient Causation: Which Cases Are Relevant?” *Political Analysis*, Vol. 10, No. 2 (Spring 2002), pp. 178–193; for the critiques of Seawright and his response see Kevin Clarke, “The Reverend and the Ravens,” *Political Analysis*, Vol. 10, No. 2 (Spring 2002), pp. 194–197; Bear F. Braumoeller and Gary Goertz, “Watching Your Posterior: Sampling Assumptions, Falsification, and Necessary Conditions,” pp. 198–203; and Seawright’s rejoinder, “What Counts as Evidence? Prior Probabilities, Posterior Distributions, and Causal Inference,” pp. 204–207. Without attempting to resolve this debate, we merely note that apart from Seawright’s arguments, there may be reasons for selecting diverse cases to test theories, including those of necessity or sufficiency. Researchers are often testing not just one necessary condition hypothesis, for example, but also alternative hypotheses that may require different cases to test. Also, the causal mechanisms being tested may be different from the independent and dependent variables used to identify and select cases. We return to this question of case selection for testing claims of necessity and sufficiency in Chapter 2, which looks at the literature on whether it is a sufficient condition for peace between two countries if both of them are democratic.

THE 'DEGREES OF FREEDOM PROBLEM' AND CASE STUDIES:

MISAPPLICATION OF A STATISTICAL VERSION OF UNDERDETERMINATION

Analysts have occasionally criticized case studies for having a “degrees of freedom problem.” This is the statistical term for the broader issue of underdetermination, or the potential inability to discriminate between competing explanations on the basis of the evidence. In our view, the statistical concept and nomenclature of “degrees of freedom” has often led to a misunderstanding of how the more generic problem of underdetermination can pose a challenge to case study methods.

In statistical methods—we focus for purposes of illustration on the example of multiple regression analysis—the term “degrees of freedom” refers to the number of observations minus the number of estimated parameters or characteristics of the population being studied (such as mean or variance). In a multiple regression analysis, the number of observations is taken as the number of cases (or the sample size) and the number of parameters is the number of independent variables and one additional parameter for the value of the intercept (the point at which the estimated regression line intercepts the axis on a graph). Thus, a study with 100 cases and 6 variables would have $100 - (6+1)$ or 93 degrees of freedom.

In a statistical study, degrees of freedom are crucial because they determine the power of a particular research design or the probability of detecting whether a specified level of explained variance is statistically significant at a specified significance level. In other words, as the sample size increases or the number of variables decreases—either of which would increase the degrees of freedom—lower and lower levels of explained variance are necessary to conclude with some confidence that the relationship being studied is unlikely to have been brought about by chance.

It is easy to see why this important consideration in the design of statistical research might seem directly applicable to case study research, which also uses the terms “case” and “variables.” In a strictly literal sense, any study of a single case using one or more variables might seem to have zero or even negative degrees of freedom and be hopelessly indeterminate apart from simple tests of necessity or sufficiency. This is a fundamentally mistaken interpretation.

We have criticized above the definition of a case as a phenomenon in which we report only one measure on any pertinent variable. It is this definition that leads to the conclusion that case studies suffer from an inherent degrees of freedom problem. In fact, each qualitative variable has many different attributes that might be measured. Statistical researchers tend to aggregate variables together into single indices to get fewer independent variables and more degrees of freedom, but case study researchers do the reverse: they treat variables qualitatively, in many of their rele-

vant dimensions. Statistical databases, for example, have created indices for “democracy,” while qualitative researchers have been more active in measuring different attributes of or types of democracy, or what has been called “democracy with adjectives.”⁵²

In addition, within a single case there are many possible process-tracing observations along the hypothesized causal paths between independent and dependent variables. A causal path may include many necessary steps, and they may have to occur in a particular order (*other* causal paths, when equifinality is present, might involve different steps in a different order.) Some analysts emphasize that defining and observing the steps along the hypothesized causal path can lead to “a plethora of new observable implications for a theory” and circumvent the degrees of freedom problem.⁵³ Donald Campbell noted this in setting out to “correct some of my own prior excesses in describing the case study approach,” arguing that:

I have overlooked a major source of discipline (i.e., degrees of freedom if I persist in using this statistical concept for the analogous problem in nonstatistical settings). In a case study done by an alert social scientist who has thorough local acquaintance, the theory he uses to explain the focal difference also generates predictions or expectations on dozens of other aspects of the culture, and he does not retain the theory unless most of these are also confirmed. In some sense, he has tested the theory with degrees of freedom coming from the multiple implications of any one theory.⁵⁴

Thus, as long as competing theories make different predictions on the causal processes thought to have taken place in a case—and sufficient evidence is accessible for process-tracing and congruence testing—case study researchers may have the means to reject many of the possible alternative explanations of a case.⁵⁵

We would go even further than Campbell on this issue. While Campbell states that “most” predictions or expectations a theory makes regarding a case must be confirmed in order for the theory to be retained, we would distinguish retaining a theory that has general utility in many cases from retaining a historical explanation of a particular case. A satisfactory historical explanation of a particular case needs to address and

52. Collier and Levitsky, “Democracy with Adjectives,” pp. 430–451.

53. King, Keohane, and Verba, *Designing Social Inquiry*, p. 225.

54. Donald Campbell, “Degrees of Freedom and the Case Study,” *Comparative Political Studies*, Vol. 8, No. 8 (July 1975), pp. 179, 181–182.

55. King, Keohane, and Verba, *Designing Social Inquiry*, pp. 119–120. King, Keohane, and Verba also acknowledge here that generating process-tracing observations can mitigate the problem of indeterminacy.

explain each of the significant steps in the sequence that led to the outcome of that case. If even one step in the hypothesized causal process in a particular case is not as predicted, then the historical explanation of the case needs to be modified, perhaps in a trivial way that is consistent with the original theory, or perhaps in a crucial way that calls into question the theory's general utility and its applicability to other cases. It is this insistence on providing a continuous and theoretically based historical explanation of a case, in which each significant step toward the outcome is explained by reference to a theory, that makes process-tracing a powerful method of inference (a point that we take up in detail in Chapter 10).

The misguided focus on case studies' supposed "degrees of freedom problem" has diverted attention from a more fundamental problem of indeterminacy that affects all research methods, even experimental methods. This is the problem that evidence, whether from a case or a database, can be equally consistent with a large or even infinite number of alternative theories. The pragmatic (but necessarily incomplete) approach we and others suggest to this problem is that researchers limit themselves to testing alternative theories, which individuals have proposed, rather than worrying over the infinite number of potential theories that lack any proponent. Even so, a particular database or case might not be able to discriminate between which of two or more competing explanations fits best. This is more a matter of how the evidence in a particular case matches up with competing hypotheses than a mechanical issue of the number of cases and the number of variables. This is why case study researchers seek crucial cases in order to be able to definitively test which of several theories fits best and, when such cases are not available, why they look for instances where a theory fails to fit a most-likely case or fits a least-likely one. When more than one competing explanation fits a case equally well, it may still be possible to narrow the number of plausible explanations, and it is also important to indicate as clearly as possible the extent to which the remaining hypotheses appear to be complementary, competing, or incommensurate in explaining the case.⁵⁶

LACK OF REPRESENTATIVENESS

Case researchers do *not* aspire to select cases that are directly "representative" of diverse populations and they usually do not and should not make claims that their findings are applicable to such populations except

56. Olav Njølstad, "Learning From History? Case Studies and the Limits to Theory-Building," in Olav Njølstad, ed., *Arms Races: Technological and Political Dynamics* (Newbury Park, Calif.: Sage Publications, 1990), pp. 220–246. See also our discussion of this problem in Chapter 2.

in contingent ways.⁵⁷ Statistical methods require a large sample of cases that is representative of and allows inferences about a larger population of cases from which the sample is drawn. Statistical researchers thus devote much effort to trying to make the sample as representative as possible. While useful and necessary in statistical studies, these practices are inappropriate and sometimes counterproductive when extended to case study methods or used to judge these methods, as some methodologists have urged.⁵⁸

Case study methods involve a trade-off among the goals of attaining theoretical parsimony, establishing explanatory richness, and keeping the number the cases to be studied manageable. Parsimonious theories rarely offer rich explanations of particular cases, and such theories must be stated in highly general terms to be applicable across different types of cases.⁵⁹ Greater explanatory richness within a type of case usually leads to less explanatory power across other types of cases. In order to explain in rich detail different types of cases, it is usually necessary to give up theoretical parsimony and to study many cases. Case studies may uncover or refine a theory about a particular causal mechanism—such as collective action dynamics—that is applicable to vast populations of cases, but usually the effects of such mechanisms differ from one case or context to another.

In view of these trade-offs, case study researchers generally sacrifice the parsimony and broad applicability of their theories to develop cumulatively contingent generalizations that apply to well-defined types or subtypes of cases with a high degree of explanatory richness.⁶⁰ Case study researchers are more interested in finding the conditions under which specified outcomes occur, and the mechanisms through which they occur, rather than uncovering the frequency with which those conditions and their outcomes arise. Researchers often select cases with the

57. Timothy McKeown, “Case Studies and the Statistical World View,” *International Organization*, Vol. 53, No. 1 (Winter 1999), pp. 161–190.

58. Stanley Lieberman, “Small N’s and Big Conclusions,” in Ragin and Becker, eds., *What is a Case? Exploring the Foundations of Social Inquiry*, pp. 108–109, 113, 116; Achen and Snidal, “Rational Deterrence Theory and Comparative Case Studies,” pp. 160–161.

59. Useful here is the reminder by King, Keohane, and Verba (*Designing Social Inquiry*, p. 20), that parsimony is not an unalloyed goal, and that “theory should be just as complicated as all our evidence suggests.”

60. Alexander L. George and Timothy McKeown, “Case Studies and Theories of Organizational Decision Making,” in Robert Coulam and Richard Smith, eds., *Advances in Information Processing in Organizations*, Vol. 2 (Greenwich, Conn.: JAI Press, 1985), pp. 43–68; McKeown, “Case Studies and the Statistical World View.”

goal of providing the strongest possible inferences on particular theories—most-likely or least-likely cases for a theory, or perhaps cases where the variables are at extreme values and the causal mechanisms are starkly evident. Researchers can also use deviant cases to help identify left-out variables.

In any of these research designs, the cases are necessarily unrepresentative of wider populations. Of course, in such research designs researchers must be careful to point out that they seek only contingent generalizations that apply to the subclass of cases that are similar to those under study, or that they seek to uncover causal mechanisms that may be in operation in a less extreme form in cases that have less extreme values on the pertinent variables. To the extent that there is a representativeness problem or a selection bias problem in a particular case study, it is often better described as the problem of “overgeneralizing” findings to types or subclasses of cases unlike those actually studied.⁶¹

SINGLE-CASE RESEARCH DESIGNS

Several of the above critiques of case study methods have converged into skepticism of the value of single-case studies. For example, *DSI* discourages research designs in which there is no variance on the dependent variable, and it also criticizes “single-observation” research designs.⁶² As *DSI* argues, studies involving only a single observation are at great risk of indeterminacy in the face of more than one possible explanation, and they can lead to incorrect inferences if there is measurement error. This same text notes that a single case study can involve many observations, however, and in our view this greatly reduces these two problems.⁶³ Thus, in our view, several kinds of no-variance research designs can be quite useful in theory development and testing using multiple observa-

61. In some instances, critiques of particular case studies have overstated the problems of representativeness and selection bias by assuming that these studies have purported to offer generalizations that cover broad populations, whereas in fact these studies carefully circumscribed their claims to apply them only to cases similar to those studied. Collier and Mahoney (“Insights and Pitfalls,” pp. 80–87) make this critique of Barbara Geddes’s review of case studies and selection bias (Barbara Geddes, “How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics,” *Political Analysis*, Vol. 2 (1990), pp. 131–150).

62. King, Keohane, and Verba, *Designing Social Inquiry*, pp. 108, 208–211.

63. King, Keohane, and Verba, *Designing Social Inquiry*, pp. 208–211. A third potential problem *DSI* cites, the possibility of omitted variables or of some form of inherent probabilism, cannot be ruled out, regardless of methods, even when one has multiple observations.

DSI acknowledges that Harry Eckstein may also have intended for research designs of crucial, least-likely, and most-likely cases to use multiple observations from the same case to test alternative explanations (footnote, p. 210).

tions from a single case. These include the deviant, crucial, most-likely, and least-likely research designs, as well as single-case study tests of claims of necessity and sufficiency. Several influential works in comparative politics have used such single-case research designs to good effect.⁶⁴

POTENTIAL LACK OF INDEPENDENCE OF CASES

One research design issue concerns whether cases are “independent” of one another. Here again, the statistical version of this problem does not apply to case studies, but a more fundamental concern does. In a statistical study, if a correlation is the result not of the hypotheses under consideration but of learning or diffusion from one case to the others, then the additional cases do not provide substantially new information and there are fewer degrees of freedom than the researcher thought (this is sometimes referred to as “Galton’s Problem”).⁶⁵ In case studies, as in large-N research, there is a danger that the researcher will fail to identify a lack of independence between cases and will consequently reach false conclusions. This danger does not manifest itself as a “degrees of freedom” problem, however, and it is not necessarily amplified by the intentional selection of cases based on a preliminary knowledge of their variables (indeed, intentional selection can address the issue of the lack of independence of cases).

The question of whether the independence of cases is a relevant consideration depends on the research objectives of a particular study, what theory or hypothesis is being developed or tested, and how the comparison of cases is structured.⁶⁶ Process-tracing can inductively uncover linkages between cases and may thereby reduce the danger of any unanticipated lack of independence of cases. When learning or diffusion processes are anticipated or uncovered and taken into account, they need not undercut the value of studying partially dependent cases. Indeed, only perfectly dependent cases are capable of providing additional information.⁶⁷ Moreover, process-tracing can be particularly effective at examining the kinds of detailed sequences in learning and diffusion processes that can create relationships between cases, allowing researchers to gauge

64. Ronald Rogowski makes this point, citing works by Arend Lijphart, William Sheridan Allen, and Peter Alexis Gourevitch. (Ronald Rogowski, “The Role of Theory and Anomaly in Social-Scientific Inference,” *American Political Science Review*, Vol. 89, No. 2 (June 1995), pp. 467–468.

65. George, “Case Studies and Theory Development,” pp. 19–23; and King, Keohane, and Verba, *Designing Social Inquiry*, p. 222.

66. George, “Case Studies and Theory Development,” p. 21.

67. King, Keohane, and Verba, *Designing Social Inquiry*, p. 222.

more accurately how much of the variance in outcomes is explained by learning or diffusion and how much is explained by other variables.⁶⁸

A lack of independence of cases is useful in research that aims to test whether the lessons of an earlier case played a causal role in a later one. Hugh Heclo made use of this in studying the process of “political learning.” Stephen Stedman’s study of four sequential efforts at international mediation in Rhodesia’s civil war also used the lack of case independence to identify possible learning from earlier cases. And, more generally, Jack Levy has suggested that intensive case studies that make use of process-tracing may be better suited than large-N quantitative studies for exploring the possibility of learning.⁶⁹

Opportunities for Multi-Method Collaborative Research

The increasingly evident complementarity of case studies, statistical methods, and formal models is likely to lead toward more collaborative work by scholars using these various methods. The recent interest among rational choice theorists in using historical case studies to test their theories, for example, is an important step in this direction.⁷⁰ More generally, there are a variety of ways in which the three methods can be used together, either in a single study or sequentially.⁷¹ Statistical analysis can help identify outliers or deviant cases, and case studies can then investigate why these cases are deviant, perhaps leading to the identification of omitted variables. Case studies can also explore the possible causal mechanisms behind the correlations or patterns observed in statistical studies, providing a check on whether correlations are spurious or potentially

68. George, “Case Studies and Theory Development,” p. 21; see also commentaries on the studies by Hugh Heclo and Stephen Stedman in the Appendix, “Studies That Illustrate Research Design.”

69. Hugh Heclo, *Modern Social Politics in Britain and Sweden* (New Haven, Conn.: Yale University Press, 1974); Stephen John Stedman, *Peacemaking in Civil War: International Mediation in Zimbabwe, 1974–1980* (Boulder, Colo.: Lynne Rienner, 1991); and Jack S. Levy, “Learning and Foreign Policy: Sweeping a Conceptual Minefield,” *International Organization*, Vol. 48, No. 2 (Spring 1994), pp. 279–312. For a fuller discussion of the Heclo and Stedman studies, see the Appendix, “Studies That Illustrate Research Design.” See also Jack Levy, “Explaining Events and Developing Theories: History, Political Science, and the Analysis of International Relations,” in Colin Elman and Miriam Fendius Elman, eds., *Bridges and Boundaries: Historians, Political Scientists, and the Study of International Relations* (Cambridge, Mass.: MIT Press, 2001).

70. Bates et al., *Analytic Narratives*.

71. These techniques of multi-method research are discussed more fully in Andrew Bennett, “Where the Model Frequently Meets the Road: Combining Statistical, Formal, and Case Study Methods,” presented at the American Political Science Association annual conference, Boston, Massachusetts, August 2002.

causal and adding details on how hypothesized causal mechanisms operate. Alternatively, when case studies lead to the specification of new variables or the refinement of concepts, statistical studies can explore whether these new variables and concepts are relevant to larger populations of cases. Formal models can be tested in case studies to see if their hypothesized causal mechanisms were in fact in operation, and the variables and concepts developed through case studies can be formalized in models.

Because case studies, statistical methods, and formal modeling are all increasingly sophisticated, however, it is becoming less likely that a single researcher can be adept at more than one set of methods while also attaining a cutting-edge theoretical and empirical knowledge of his field. Successful collaboration is therefore likely to take the form of several researchers working together using different methods, or of researchers more self-consciously building on the findings generated by scholars who have used different methods. In either form, effective collaboration requires that even as researchers become expert in one methodological approach, they must also become conversant with alternative approaches, aware of their strengths and limitations, and able to make informed reading of their substantive results. The next chapter shows how varied research methods have contributed to the progress of the democratic peace research program, from a broad hypothesis to a refined set of contingent generalizations.

Organization of the Book

This is a large book, and many readers may wish to focus on chapters that meet their current needs. Chapter 2 is about the research methods that political scientists have used to develop and study the democratic peace theory. It provides an extended illustration of what purposes are best served by different research methods; how knowledge cumulates within a research agenda; and how typological theories draw on the results of a large number of researchers. Chapter 2 reflects our strong belief that each research method is strong at answering particular kinds of questions, and that beyond the din of social scientists' sometimes heated disagreements, one can discern the cumulation of knowledge in the social sciences.

Part II is intended as a practical guide for graduate students. Chapter 3 introduces case study research design through a discussion of the method of structured, focused comparison. Chapter 4 covers the design of case studies; Chapter 5 discusses the work involved in actually carrying out the study; and Chapter 6 provides guidance for drawing implications for a theory from the findings of a case.

Part III addresses important methodological and epistemological issues of alternative case study methods and also discusses the use of typological theories. The section begins with Chapter 7, on the philosophical underpinnings of our methodological advice. Chapter 8, on comparative methods, focuses on the challenges of case methods that rely on the logic of controlled comparisons and highlights a need for methods that do not rely upon the covariance of variables. Chapter 9 discusses the congruence method, in which the researcher examines the correspondence between the values of the independent and dependent variables in a case. Chapter 10 discusses the method of process-tracing, and identifies its differences and similarities to historical explanation. Chapter 11, on the use of typological theories, provides guidance for the inductive and deductive construction of such theories, and the research designs supported by each. Chapter 12 offers additional advice on how to design research that will be relevant to policymakers; this chapter will also be useful to more senior academics who have not considered this issue.

We have also included an Appendix, “Studies that Illustrate Research Design,” which briefly reviews the research designs of numerous books. This may be useful to graduate students who want to explore research designs in well-regarded studies; it may also be helpful to professors as they design classes in case study methods.