

Chapter 2

Case Study Methods and Research on the Interdemocratic Peace

Political scientists have amassed growing evidence in the past three decades that democracies seldom if ever make war upon one another. This finding has sparked a rich literature on whether and how the international behavior of democracies is different from that of other kinds of regimes. Because the resulting “democratic peace” research program has developed so recently and rapidly, it has involved a broad range of sophisticated contemporary research methods and provides an excellent illustration of the methodological themes highlighted in Chapter 1.¹ Though disagreements among researchers over results and methods have often been sharp, it is clear that work on this subject by numerous scholars using several methods has achieved a progressively better understanding of when and how democracies use force, and the differences between their behavior and that of other types of regimes. Statistical methods, case studies, and formal models have all made important contributions to this cumulation of knowledge, and typological theories have been useful in synthesizing the literature on this topic, and in creating useful case study research designs.

This chapter analyzes the methodological lessons of the democratic peace research program, rather than directly engaging theoretical arguments about whether we should or should not expect democracies to be-

1. Of course, many different research programs in the social sciences illustrate the complementary nature of formal, statistical, and qualitative methods. For an analysis of how these methods have contributed to research in comparative politics, for example, see David Collier, “Comparative Politics: The Case for a New Synthesis,” in David Collier and James Mahoney (eds.), *Explaining Institutional Change: Ambiguity, Agency, and Power* (Cambridge: Cambridge University Press, 1993), pp. 1–25.

have differently from other kinds of regimes. One challenge in carrying out such an analysis is that the democratic peace research program has grown to encompass many different propositions. There is some evidence, for example, that democracies are more likely than other kinds of governments to ally with one another, trade with one another, form long-lasting intergovernmental organizations, accept mediation in disputes with one another, obey international law, avoid militarized disputes with one another short of war, and win the wars in which they choose to participate.² As the literature on these questions is vast and includes hypotheses with varying degrees of support, we focus on the hypothesis that democracies rarely if ever make war upon one another. We use the term "interdemocratic peace" to distinguish this hypothesis from the related argument, for which the evidence is more ambiguous, that democracies are generally less prone to war. The interdemocratic peace hypothesis is one of the earliest, most familiar, and best substantiated claims of the research program, and it has thus arguably generated the most methodologically diverse and sophisticated research.³ This chapter assesses three methodological strands of the literature on this question that roughly succeeded one another.

The first generation of empirical research on the democratic peace, from the early 1960s through the late 1980s, for the most part utilized statistical methods to assess correlations between regime types and war. This research sought to establish whether democracies have been more peaceful generally or toward one another, and it attempted to determine whether correlations to this effect were spurious. The result was a fairly

2. For a listing of these and related hypotheses and the authors who introduced them, see James Lee Ray, "A Lakatosian View of the Democratic Peace Research Program," in Colin Elman and Miriam Fendius Elman, eds., *Progress in International Relations Theory: Appraising the Field* (Cambridge, Mass.: MIT Press, 2003), p. 221. As Ray points out, one of the most convincing arguments on behalf of the progressivity of the broader research program is that so many of the diverse auxiliary hypotheses that it has engendered have proven to have some merit.

3. Several of the auxiliary hypotheses have also involved sophisticated research using a variety of statistical, formal and case study methods. On the hypothesis that democracies tend to win the wars in which they participate, for example, see Dan Reiter and Alan Stam, *Democracies at War*, (Princeton, N.J.: Princeton University Press, 2002). This study uses both statistical and case study methods. See also David Lake, "Powerful Pacifists: Democratic States and War," *American Political Science Review*, Vol. 86, No. 1 (March 1992), pp. 24-37. For a critique of these works that turns on methodological issues, see Michael Desch, "Democracy and Victory: Why Régime Type Hardly Matters," *International Security*, Vol. 27, No. 2 (Fall 2002), pp. 5-47. For rejoinders, see David Lake, "Fair Fights? Evaluating Theories of Democracy and Victory"; Dan Reiter and Alan Stam, "Understanding Victory: Why Political Institutions Matter"; and Mi-

robust, but not unanimous consensus that democracies have rarely if ever fought wars against one another, but that they have engaged in war in general with about the same frequency as have other types of regime.

Yet adequate causal explanations must include two things: correlational or probabilistic statements associating purported causes with observed effects, *and* logically coherent and consistent assertions on the underlying causal mechanisms through which purported causes affect outcomes. As the focus of the research program began to shift from the “whether” to the “why” of the democratic peace, a second generation of research began to use case studies to test purported causal mechanisms more directly, develop more finely differentiated variables and typological theories, and identify new variables. This research was more cognizant of the possibility that the democratic peace might manifest the phenomenon of equifinality. In other words, as in the title of a book edited by Miriam Fendius Elman, there might be *Paths to Peace*, rather than one single path to peace among democracies.⁴ Key claim

The third and most recent generation of literature on the interdemocratic peace has used formal models to refine theories on this phenomenon and has tested these revised theories with both statistical and case study research. Formal models have helped clarify the logic of how democratic institutions might both constrain democracies’ foreign policy behaviors and inform other states of the credibility of commitments democratic leaders make regarding the possible use of force.

This chapter looks at these three generations of the literature on the interdemocratic peace. Yet this tripartite categorization of research on this topic should not be taken as suggesting that any one method has or will supplant others in the democratic peace research program or that the evolution of social science research programs generally proceeds from one method to another. Research using all three methods usually proceeds simultaneously and iteratively, as each method confronts new research tasks where another method is superior. Much useful work on the democratic peace remains to be done using all three approaches. As case studies and formal models refine the concepts and logic of democratic peace theories, statistical tests can fruitfully be redone using these new concepts and their associated measurements. Such tests will in turn help identify new sets of anomalous cases for further case studies, which can provide fertile ground for both inductive and formal refinements to extant theories, which will need to be tested by new statistical studies, and so on.

The First Generation: Contributions of Statistical Methods

As James Lee Ray points out in his thorough review of the literature on the democratic peace, arguments for the existence of a democratic peace can be traced back to such liberal theorists as Immanuel Kant and Woodrow Wilson, and critiques of these arguments have an equally distinguished pedigree among realist thinkers like E. H. Carr and Hans Morgenthau and neorealists such as Kenneth Waltz.⁵ Much of the contemporary research on this subject, however, can be traced back to a 1964 article by Dean Babst, a research scientist at the New York State Narcotic Addiction Control Commission.⁶ In this four-page article, Babst concluded that "no wars have been fought between independent nations with elective governments between 1789 to 1941," and he calculated the difference between the proportions of democratic and mixed or nondemocratic dyads at war in World Wars I and II to be significant at the 1 percent level.⁷ Yet Babst's article was theoretically underdeveloped, positing a monadic explanation (the purported reluctance of democratic publics to vote to take on the costs of war) for this dyadic result, and it did not control for important variables.

J. David Singer and Melvin Small rescued Babst's argument from obscurity among political scientists by critiquing it in a 1976 article that contended that the war involvement of democratic states between 1816 and 1965, in terms of duration and battle deaths, was not significantly different from that of other types of regimes. Singer and Small suggested that the absence of wars between democracies was due to the fact that democratic states rarely bordered upon one another, but they did not test this assertion.⁸ In the late 1970s and 1980s, a rapidly expanding body of statistical research made three key contributions to the democratic peace research program. First statistical studies refined the research question from whether democratic states were more peaceful in general to whether they were more peaceful only or primarily toward one another (the interdemocratic peace).⁹ Some research continued on the monadic

5. James Lee Ray, *Democracies and International Conflict: An Evaluation of the Democratic Peace Process* (Columbia: University of South Carolina Press, 1995), pp. 1-9.

6. *Ibid.*, p. 11.

7. Dean V. Babst, "Elective Governments: A Force for Peace," *Wisconsin Sociologist*, Vol. 3, No. 1 (1964), pp. 9-14.

8. Melvin Small and J. David Singer, "The War-Prone-ness of Democratic Regimes," *Jerusalem Journal of International Relations*, Vol. 1, No. 1 (1976), pp. 50-69.

9. Rudolph I. Rummel. *War. Power. Peace* (Beverly Hills. Calif.: Sage Publications.

proposition that democracies might be more peaceful in general, but research increasingly focused on the stronger evidence for an interdemocratic peace.¹⁰ Researchers also used statistical methods to test whether democracies have been less likely than other states to engage in conflicts short of war—both generally and vis-à-vis one another.¹¹ Some researchers also began to examine whether subtypes of states, such as states in transition to democracy, were more or less prone to war.¹² Second, many statistical studies tested for whether findings of an interdemocratic peace were spurious by controlling for the effects of numerous variables—including contiguity, wealth, alliance membership, relative military capabilities, rates of economic growth, and the presence of a hegemon.¹³ Third, researchers using statistical methods theorized on and began to test the potential causal mechanisms behind an interdemocratic peace, often grouping them together under explanations relating to democratic norms or institutions or some interaction between the two.¹⁴

Erich Weede, "Democracy and War Involvement," *Journal of Conflict Resolution*, Vol. 28, No. 4 (December 1984), pp. 649–664; Michael Doyle, "Liberalism and World Politics," *American Political Science Review*, Vol. 80, No. 4 (December 1986), pp. 1151–1161; Nasrian Abdolali and Zeev Maoz, "Regime Types and International Conflict, 1817–1976," *Journal of Conflict Resolution*, Vol. 33, No. 1 (March 1989), pp. 3–35; and Zeev Maoz and Bruce Russett, "Normative and Structural Causes of Democratic Peace, 1946–1986," *American Political Science Review*, Vol. 87, No. 3 (September 1993), pp. 624–638.

10. David Rousseau et al., "Assessing the Dyadic Nature of the Democratic Peace," *American Political Science Review*, Vol. 90, No. 3 (September 1996), pp. 512–533.

11. T. Clifton Morgan and Valerie L. Schwebach, "Take Two Democracies and Call Me in the Morning: A Prescription for Peace?" *International Interactions*, Vol. 19, No. 4 (1992), p. 305; Bruce Bueno de Mesquita and David Lalman, *War and Reason* (New Haven, Conn.: Yale University Press, 1992); and William J. Dixon, "Democracy and the Peaceful Settlement of International Conflict," *American Political Science Review*, Vol. 88, No. 1 (March 1994), pp. 14–32.

12. Jack Snyder and Edward D. Mansfield make a monadic argument that states in transition to democracy are particularly prone to war. See Jack Snyder and Edward D. Mansfield, "Democratization and the Danger of War," *International Security*, Vol. 20, No. 1 (Summer 1995), pp. 5–38. In their subsequent book on this topic, these authors use both statistical tests and case studies to elaborate upon this argument. See Edward D. Mansfield and Jack Snyder, *Electing to Fight: Why Emerging Democracies Go to War* (Cambridge, Mass.: MIT Press, forthcoming).

13. Maoz and Russett, "Normative and Structural Causes of Democratic Peace"; and Stuart A. Bremer, "Democracy and Militarized Interstate Conflict, 1816–1965," *International Interactions*, Vol. 18, No. 3 (1993), pp. 231–249. For a more complete list of variables and studies, see Margaret Hermann and Charles Kegley, Jr., "Rethinking Democracy and International Peace: Perspectives From Political Psychology," *International*

On the first two tasks of refining the research question and testing for possible spuriousness, statistical methods greatly advanced the research program and achieved a growing consensus among researchers.¹⁵ The conflict behavior of democracies and other regime types gained attention as a research program worthy of intensive study, even among skeptics of a democratic peace. Also, a consensus emerged that democracies are not markedly more peaceful in general, although some studies continue to challenge and qualify this conclusion.¹⁶ The consensus view is also that democracies have fought wars substantially less frequently against one another than they have against other types of states, although opinions differ on the number and seriousness of exceptions to this generalization.¹⁷ A weaker consensus emerged around the idea that democracies are less likely to engage in militarized disputes with one another short of war.¹⁸

Statistical methods proved less successful at explaining why an interdemocratic peace might exist. Researchers using statistical methods had theorized and rigorously defined several potential causal mechanisms that might explain the democratic peace, focusing on democratic institutions and democratic norms. However, the posited causal mechanisms were often contradictory, and no consensus existed on which of these variables caused an interdemocratic peace. Statistical methods proved inadequate to test these mechanisms for three reasons. First, they

pp. 624–638; and Dixon, “Democracy and the Peaceful Settlement of International Conflict.”

15. For similar assessments of where the consensus lies on these issues, see Ray, *Democracies and International Conflict*; Bear F. Braumoeller, “Causal Complexity and the Study of Politics,” *Political Analysis*, Vol. 13, No. 3 (2003), pp. 209–233; Bueno de Mesquita and Lalman, *War and Reason*; and Elman, ed., *Paths to Peace*.

16. Rousseau et al., “Assessing the Dyadic Nature of the Democratic Peace.”

17. For an account that uses statistical methods to question the existence of an interdemocratic peace, see David E. Spiro, “The Insignificance of the Liberal Peace,” *International Security*, Vol. 19, No. 2 (Fall 1994), pp. 50–86. In subsequent published correspondence, Bruce Russett critiques Spiro’s argument, particularly Spiro’s assumption that dyadic data points lack independence and offers some of the most convincing statistical tests yet for the existence of an interdemocratic peace (Bruce Russett, “Correspondence: The Democratic Peace,” *International Security*, Vol. 19, No. 4 (Spring 1995), pp. 164–184). For an additional account that finds Russett’s statistical tests more convincing, see Braumoeller, “Causal Complexity and the Study of Politics.” For an additional account that questions the statistical validity of the finding of an interdemocratic peace, see Henry Farber and Joanne Gowa, “Politics and Peace,” *International Security*, Vol. 20, No. 2 (Fall 1995), pp. 123–146.

18. Dixon, “Democracy and the Peaceful Settlement of International Conflict.”

faced daunting measurement problems.¹⁹ One of the most methodologically sophisticated efforts to test for the normative versus institutional causes of the democratic peace, by Bruce Russett and Zeev Maoz, illustrates these problems. Maoz and Russett use well-established and straightforward measures to control for wealth, economic growth, and contiguity. They also employ careful measures of more complex variables such as alliance membership and ratios of military capabilities. Their measurement of democratic institutions is more complex, though there is at least some consensus on this issue, as many quantitative studies have joined Maoz and Russett in relying on the "Polity II" data set, or modified versions of this data set.²⁰ The most difficult measurement problem, however, is that there is no easy way to quantify the slippery variable of "democratic norms" and no widely accepted database for this variable. Consequently, Maoz and Russett used the longevity of political regimes as a proxy for the prevalence of their norms, and they used the average number of recent deaths from domestic political violence or executions within a dyad as a measure of the "democraticness" of that dyad's norms.²¹ Clearly, these proxy measures are problematic, as authoritarian and totalitarian states that persist for decades may minimize the use of domestic violence by monopolizing the instruments of force and creating powerful police and intelligence institutions that deter domestic violence and political opposition.

To some extent, even measurement problems on complex variables like democratic norms can be addressed, and statistical researchers have proven adept at devising creative ways of measuring complex variables. One study by Bear Braumoeller, for example, has developed a dedicated definition and data set for looking at democratic norms as they relate to the democratic peace. This study even measures the differences between the norms of elites and those of mass publics.²² This is a very labor-

19. On the difficulties of conceptualizing and measuring democracy, and on the strengths and weaknesses of various statistical databases in doing so, see Gerardo Munck and Jay Verkuilen, "Conceptualizing and Measuring Democracy: Evaluating Alternative Indices," *Comparative Political Studies*, Vol. 35, No. 1 (February 2002), pp. 5–34.

20. Ted Gurr et al., "The Transformation of the Western State: The Growth of Democracy, Autocracy, and State Power Since 1800," *Studies in Comparative International Development*, Vol. 25, No. 1 (1990), pp. 73–108. The "Polity" data set, begun in the 1970s, has been updated several times, and the current "Polity IV" version is available at <<http://weber.ucsd.edu/~kgledits/Polity.html>>.

21. Maoz and Russett, "Normative and Structural Causes of Democratic Peace," p. 630.

intensive task and it is all but impossible to implement for states for which extensive and dedicated polling data is not available. More generally, data sets that quantify or dichotomize variables can achieve reproducible results across many cases (external validity), but only at the cost of losing some of the ability to devise measures that faithfully represent the variables that they are designed to capture (internal validity).

A second problem is that statistical methods are not well suited to testing causal mechanisms in the context of particular cases. These methods are optimized for assessing correlations across cases or among data points within a case, rather than for testing whether every aspect of a case is consistent with a hypothesized causal process. In contrast to statistical methods, if process-tracing shows that a single step in a hypothesized causal chain in a case study is not as the theory predicts, then the variable in question cannot explain that case without modification, even if it does explain most or even all other cases. If, for example, we find a case in which a democratic public clamored for going to war, the hypothesized propensity of democratic citizens to avoid voting upon themselves the cost of war cannot explain this case, even if it might explain other cases. Conversely, if a complex hypothesis involved one hundred steps and ninety-nine of these were as predicted in a case, a statistical test would confirm the hypothesized process at a high level of significance, but a case study analysis would continue to probe the missing step.

Third, the relative infrequency of both wars and contiguous democracies presents a sharp methodological limitation for statistical research. Given the small number of potential wars between democracies, the existence of even a few wars between democracies or the omission of a single relevant variable could call into question the statistical support for an interdemocratic peace.²³ Because there are at least twenty hotly debated potential exceptions or near-exceptions to the assertion that democracies have never fought wars with one another, the results of statistical studies remain provisional despite the emerging consensus that an interdemocratic peace exists.²⁴ For case study researchers, this is an opportunity rather than a problem: it is easily possible for the field as a whole to intensively study every one of the possible exceptions to the democratic peace and to also include a number of comparative cases of mixed dyads and nondemocratic dyads.

23. Ray, *Democracies and International Conflict*; and Spiro, "The Insignificance of the Liberal Peace."

Explain
this
claim

The Second Generation: Case Study Contributions

As one researcher argued in the 1990s, “generalizations about the democratic peace are fine—we have many of them—but now is the time to explore via comparative case studies the causal chains, if they exist.”²⁵ The limitations of statistical methods as applied to the democratic peace were greatest precisely where case study methods had the most to contribute.²⁶ Case studies on the democratic peace in the past decade illustrate the comparative advantages of qualitative methods and offer commendable examples of alternative research designs.

One of the main advantages of case studies is their ability to serve the heuristic purpose of inductively identifying additional variables and generating hypotheses.²⁷ Statistical methods lack accepted procedures for inductively generating new hypotheses. Moreover, case studies can analyze qualitatively complex events and take into account numerous variables precisely because they do not require numerous cases or a restricted number of variables. Case study researchers are also not limited to variables that are readily quantified or those for which well-defined data sets already exist. Case studies on the democratic peace have thus identified or tested several new variables, including issue-specific state structures, specific norms on reciprocity and the use of deadly force, leaders’ perceptions of the democraticness of other states, transparency, and the distinction between status quo and challenger states.²⁸

Second, process-tracing can test individual cases regarding the claims made about causal mechanisms that might account for a democratic peace. Miriam Elman, for example, asserts that

25. Correspondence of Kalevi Holsti, cited in Elman, ed., *Paths to Peace*, p. 44.

26. Elman, ed., *Paths to Peace*, p. 43.

27. 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; Alexander L. George, “Case Studies and Theory Development,” in Paul G. Lauren, ed., *Diplomacy: New Approaches in Theory, History, and Policy* (New York: Free Press, 1979), pp. 43–68.

28. See, respectively, Susan Peterson, “How Democracies Differ: Public Opinion, State Structure, and the Lessons of the Fashoda Crisis,” *Security Studies*, Vol. 5, No. 1 (Autumn 1995), pp. 3–37; William Hoelt, *Explaining Interdemocratic Peace: The Norm of Cooperatively Biased Reciprocity* (Ph.D. dissertation, Georgetown University, 1993); John M. Owen, “How Liberalism Produces Democratic Peace,” *International Security*, Vol. 19, No. 2 (Fall 1994), pp. 87–125; Bernard Finel and Kristin Lord, “The Surprising Logic of Transparency,” *International Studies Quarterly*, Vol. 42, No. 2 (June 1998), pp. 257–280.

The quantitative empirical analyses that find that democracy is associated with peace are correlational studies, and provide no evidence that leaders actually consider the opponent's regime type in deciding between war and peace. These studies focus primarily on foreign policy outcomes and ignore the decision-making process. If we want to move beyond correlation to causation, we need to reveal the decision-making processes of aggressive and pacific states.²⁹

Both proponents and critics of the existence of a democratic peace agree on the importance of process-tracing on causal mechanisms, and researchers who had once relied largely on statistical methods have turned to case study methods because of these methods' ability to test causal mechanisms.³⁰ Some have combined statistical and case study techniques.³¹ Since the 1990s, scholars have used case study methods to test many of the hypothesized causal mechanisms and independent variables listed above, but there is not yet a consensus on which causal mechanisms might account for a democratic peace. However, case studies have been able to rule out the presence of some causal mechanisms in important cases. For example, the assertion that democratic mass publics oppose wars with other democracies does not hold for the Fashoda Crisis between Britain and France in 1898.³²

Third, case studies can develop typological theories (theories on how different combinations of independent variables interact to produce different levels or types of dependent variables). Researchers have begun to identify the conditions under which specified types of democracies behave in various contexts to produce specific types of conflict behavior within democratic or mixed dyads.³³ The resulting theories usually focus on interactions among combinations of variables, rather than variables considered in isolation.

The development of typological theories thus involves differentiating configurations of independent and dependent variables into qualitatively different "types," such as types of war or types of democracy. The task of

29. Elman, ed., *Paths to Peace*, p. 33.

30. Owen, "How Liberalism Produces Democratic Peace," p. 91; Christopher Layne, "Kant or Cant: The Myth of Democratic Peace," *International Security*, Vol. 19, No. 2 (Fall 1994), pp. 12–13; and Ray, *Democracies and International Conflict*, pp. 151; 158–159.

31. See Mansfield and Snyder, *Electing to Fight*; and Edgar Kiser, Kriss A. Drass, and William Brustein, "Ruler Autonomy and War in Early Modern Europe," *International Studies Quarterly*, Vol. 39, No. 1 (March 1995), pp. 109–138.

32. Ray, *Democracies and International Conflict*; and Peterson, "How Democracies Differ."

defining “war” and “democracy” is challenging for both statistical and case study researchers, and they respond to it differently. Statistical researchers attempt to develop rigorous but general definitions, with a few attributes that apply across a wide number of cases. Case study researchers usually include a larger number of attributes to develop more numerous types and subtypes, each of which may apply to a relatively small number of cases.³⁴ In the context of the democratic peace, for example, case study researchers have suggested differentiating between centralized and decentralized democracies, and among democracies where leaders and mass publics either share or have different norms regarding the use of force vis-à-vis other democracies.³⁵ It is also useful to distinguish among different kinds of peace. Alexander George has suggested, for example, that it is important to distinguish among three types of peace: “precarious peace,” which is the temporary cessation of hostilities when one side remains dissatisfied with the status quo and continues to see force as a legitimate means of changing it; “conditional peace,” such as the situation that existed during the Cold War, when the threat of mutual destruction by nuclear weapons helped deter war; and “stable peace,” when two states no longer even consider or plan for the possibility of using force against one another.³⁶

Two examples illustrate particularly well the kind of typological theories that case studies can develop to model complex interactions of variables. The first is Susan Peterson’s model of how war was averted in the Fashoda Crisis in ways that are not entirely consistent with either liberal or realist views of the democratic peace. Peterson argues that systemic variables (such as military balance) interacted with state institutions and the preferences of leaders and public opinion in France and Britain to avert war. In Britain, she argues, the dovish Prime Minister Lord Robert Cecil (the Earl of Salisbury) was constrained by a strong and hawkish cabinet, parliament, and public, and was pushed into more confrontational policies than he would have liked. In France, the hawkish Foreign Minister Theophile Delcasse was constrained by a more dovish parliament and public, as well as by France’s military inferiority, but not by his own cabinet. As a result, France pushed harder for concessions and Salisbury was more willing to make them than realists might expect, while

34. David Collier and Steven Levitsky, “Democracy with Adjectives: Conceptual Innovation in Comparative Research,” *World Politics*, Vol. 49, No. 3 (April 1997), pp. 430–451; and Elman, ed., *Paths to Peace*, pp. 35–39.

35. Elman, ed., *Paths to Peace*, pp. 36–37; 41.

36. Alexander L. George. “Foreword” in Arie Kacowicz, ed., *A Stable Peace Among*

traditional liberal theories on the democratic peace have difficulty explaining the British public's willingness to go to war against France.³⁷

A second example is Randall Schweller's study of how democracies behave with regard to preventive war. Like Peterson, Schweller incorporates both systemic and domestic variables, looking at how domestic structures affect state decisions on preventive wars during ongoing power shifts. Schweller concludes that only nondemocratic states wage preventive wars against rising opponents, and that democracies seek accommodation with rising democracies and form counterbalancing alliances against rising nondemocratic challengers.³⁸

Both these studies define useful subtypes of democracies, but not every subtype is useful or progressive. Researchers might allow their subjective biases to intrude, leading them to define away anomalies through the creation of subtypes. As Miriam Elman argues, for example, "defining democracy as a regime in an independent state that ensures full civil and economic liberties; voting rights for virtually all the adult population; and peaceful transfers of power between competing political groups makes it fairly easy to exclude numerous cases of warring democracies."³⁹ The creation of a new subtype is warranted if it helps explain not only the aspects of a case that led to the creation of this subtype, but also other unexplained dimensions of the case or of other cases. The assertion that "new" or "transitional" democracies are more war-prone and should be treated differently from other cases that might fit the democratic peace, for example, may warrant the creation of a new subtype. It posits testable correlations and causal mechanisms and suggests dynamics that should make states in transitions from as well as into democracy more war-prone.⁴⁰ More questionable is the exclusion from assertions on the "democratic peace" of civil wars, like the U.S. Civil War.⁴¹ Also debatable is the exclusion from some data sets of conflicts that fall somewhat below the arbitrary figure of 1,000 battle deaths, such as the conflict between Finland and Britain during World War II.⁴²

While case study methods are particularly amenable to creating subtypes and differentiating variables, they have no monopoly on such inno-

37. Peterson, "How Democracies Differ."

38. Randall L. Schweller, "Domestic Structure and Preventive War: Are Democracies More Pacific?" *World Politics*, Vol. 44, No. 2 (January 1992), pp. 235-269. The U.S. invasion of Iraq in 2003 may constitute an important exception to Schweller's argument.

39. Elman, *Paths to Peace*, p. 21.

40. Snyder and Mansfield, "Democratization and the Danger of War."

uations. Studies using statistical methods have addressed the behavior of “democratizing” states and have examined the behavior of states that have democratic institutions but not democratic norms.⁴³ Also, once case studies identify potentially useful subtypes, if a sufficient number of cases in the subtype exists statistical tests can assess whether these subtypes are indeed correlated with the specified outcome. In this way, case studies can often help develop sharper concepts, subtypes, or measurement procedures that can then be incorporated into statistical studies, though this can require considerable effort in recoding the cases in existing statistical datasets.

Examples of Case Study Research Design in the Interdemocratic Peace Literature

The democratic peace literature provides some of the best examples of how to implement case studies. These examples illustrate the important point that there is no single “case study research design.” Rather, different case study research designs use varying combinations of within-case analysis, cross-case comparisons, induction, and deduction for different theory-building purposes.⁴⁴ An excellent example of a case study research design using both within-case analysis and cross-case comparisons is *Paths to Peace*, edited by Miriam Elman. Elman carefully defines the class of cases to be studied—international crises between democratic, mixed, and nondemocratic dyads—while acknowledging that this class of cases cannot adequately test the assertion that democracies frequently resolve their conflicts with one another without resorting to war.⁴⁵ The alternative of trying to select “non-crisis” as well as crisis cases is obviously problematic, and crisis cases have the advantage of posing tough tests for hypotheses supporting the democratic peace. Moreover, Elman selects cases that provide substantial variance on the independent and dependent variables, and in contrast to many studies, includes cases of both wars and successful crisis management from all three types of dyads: demo-

43. Snyder and Mansfield, “Democratization and the Danger of War”; and Braumoeller, “Causal Complexity and the Study of Politics.”

44. There are, unfortunately, no good examples of a “crucial” or critical case in the democratic peace literature. As Eckstein argues, history seldom provides clear examples of cases that satisfy the demanding criteria of a crucial case. The second best alternative, Eckstein argues, are “most-likely” cases that a theory fails or “least-likely” cases that a theory passes (Harry Eckstein, “Case Studies and Theory in Political Science,” in Fred Greenstein and Nelson Polsby, eds., *Handbook of Political Science*, Vol. 7,

cratic dyads, mixed dyads, and nondemocratic dyads.⁴⁶ Elman also evaluates alternative hypotheses via process-tracing and cross-case comparison tests.

The democratic peace literature has also produced commendable examples of research designs that incorporate "least similar" and "most similar" cases. In the most similar case design, the researcher attempts to select cases that are similar in all of their independent variables except one and differ in their dependent variable. James Lee Ray uses this design to compare the cases of the Fashoda Crisis and the Spanish-American War.⁴⁷ Ray carefully addresses each of the standard categories of confounding variables identified by Donald Campbell and Julian Stanley, including "regression" (selection of cases with extreme scores on the variables of interest) and other kinds of selection bias, as well as "mortality" (the differential loss of respondents from the study: in this instance, the possibility that states may become more authoritarian as they see a war coming).⁴⁸ Ray obviates other standard problems, such as the effects of history, maturation, and changes in instrumentation, by selecting cases from the same year. Ray also addresses six other variables that might account for the different outcomes of the two cases: proximity, power ratios, alliances, levels of economic development, militarization, and political stability. Ray's systematic attention to these aspects of cross-case comparison, as well as his use of process-tracing evidence, bolsters his conclusions that Spain's autocracy contributed to the Spanish-American War and that democracy in France and Britain helped peacefully resolve the Fashoda Crisis.⁴⁹

As for the research design of least similar cases, Carol and Melvin Ember and Bruce Russett use the logic of this design, together with the instruments of statistical research, to test assertions about the democratic peace. In a least similar cases design, the researcher selects cases that are dissimilar in all but one independent variable, but that share the same dependent variable. This can provide evidence that the single common independent variable, in this instance the democratic decision-making processes, may account for the common dependent variable. In their study, Ember, Ember, and Russett test the findings of studies on the democratic

46. Ibid., pp. 47–52.

47. Ray, *Democracies and International Conflict*, pp. 159–200.

48. Donald Campbell and Julian Stanley, *Experimental and Quasi-Experimental Design for Research* (Chicago: Rand McNally, 1963).

49. Ray, *Democracies and International Conflict*, pp. 159–200; Carol Ember, Melvin Ember and Bruce Russett "Peace Between Participatory Politics: A Cross-Cultural Test of

peace among modern states and test them against pre-industrial societies; they find support for the proposition that participatory decision-making processes are conducive to peace in otherwise very different industrial and pre-industrial societies.

Both the most similar and least similar designs for case study comparisons, which rely on the logic of John Stuart Mill's "method of difference" and "method of agreement," respectively, are subject to methodological limitations that Mill and others have identified.⁵⁰ In particular, the omission of relevant variables can entirely invalidate the results of cross-case comparisons in either design. Yet there are safeguards against this, as exemplified by Ray's careful attention to a wide range of alternative hypotheses to ensure that no relevant variables are omitted from the comparison. In addition, process-tracing (undertaken by Ray, by Ember, and by Russett) provides an additional check on the results of cross-case comparisons.

Critiques and Challenges of Case Study Methods as Applied to the Democratic Peace

The case study literature on the democratic peace reveals two problems in case study methods: the problem of case selection and that of reconciling conflicting interpretations of the same cases. On the issue of case selection, there is always the danger that case study researchers' subjective biases and commitments to certain theoretical propositions will lead them to select cases that over-confirm their favorite hypotheses (a different and potentially more serious problem than that addressed in standard discussions of selection biases in statistical studies, which result in truncated samples and under-confirmation of hypotheses).⁵¹ Biased case selection can also arise from the fact that evidence on certain cases is more readily accessible than that on others and from the tendency for historically important cases to be overrepresented relative to studies of obscure—but theoretically illuminating—events. Miriam Elman argues, for example, that democratic peace case studies overemphasize cases involving the United States and that they have focused excessively on the study of the Fashoda Crisis and the Spanish-American War compared to possible exceptions of the democratic peace. She also maintains that demo-

50. Alexander L. George and Timothy J. 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.

cratic dyads have been over-studied relative to mixed and nondemocratic dyads.⁵² On the other hand, for some theory-building purposes mixed dyads are less interesting, and existing studies of wars in mixed and non-democratic dyads may help fill this gap. Studies that show that states have initiated wars despite inferior military capabilities, for example, call into question assertions that military imbalances alone help explain cases of successful crisis management by democracies. Still, Elman is justified in arguing that more dedicated case studies of mixed and even nondemocratic dyads are needed for comparative research designs like Ray's study of the Spanish-American War and the Fashoda Crisis.

Yet the substantial convergence among supporters and critics of the democratic peace on which cases deserve study demonstrates that case selection is not an arbitrary process. Several cases have been mentioned by numerous scholars as possible deviant cases, or exceptions to the democratic peace, including the War of 1812, the U.S. Civil War, conflicts between Ecuador and Peru, the Fashoda Crisis, the Spanish-American War, and Finland's conflict with Britain in World War II. Many of the fourteen other possible exceptions to the democratic peace listed by Ray have also been cited by more than one author or subjected to more than one case study.⁵³ The initial focus on "near wars" between democracies and "near democracies" that went to war was appropriate for the first wave of case studies of the interdemocratic peace, as it offered tough tests of such a theory. As researchers accumulate adequate studies of these cases, they can branch out into more comparisons to mixed and nondemocratic dyads, as Elman has begun to do.

As researchers conduct multiple studies of particular cases, how can they reconcile or judge conflicting interpretations of the same cases? Olav Njølstad emphasizes this problem in case study research, noting that differing interpretations may arise from several sources. First, competing explanations or interpretations could be equally consistent with the process-tracing evidence, making it hard to determine whether both are at play and the outcome is overdetermined, whether the variables in competing explanations have a cumulative effect, or whether one variable is causal and the other spurious. Second, competing explanations may address different aspects of a case, and they may not be commensurate. Third, studies may simply disagree on the "facts" of the case.

Njølstad offers several useful suggestions on these problems.⁵⁴ These suggestions include: identifying and addressing factual errors, disagree-

52. Elman, ed., *Paths to Peace*, pp. 47-52.

ments, and misunderstandings; identifying all potentially relevant theoretical variables and hypotheses; comparing various case studies of the same events that employ different theoretical perspectives (analogous to paying careful attention to all the alternative hypotheses in a single case study); identifying additional testable and observable implications of competing interpretations of a single case; and identifying the scope conditions for explanations of a case or category of cases.

The democratic peace literature illustrates how these suggestions work in practice. There is some factual disagreement on whether both British and French public opinion was bellicose in the Fashoda Crisis or whether British public opinion was substantially more supportive of going to war, and some argue that foreign policymaking was so dominated by elites in both cases that public opinion made little difference.⁵⁵ Similarly, there is some disagreement on the nature and salience of public opinion in Spain at the time of the Spanish-American War.⁵⁶

On the Fashoda Crisis, there is disagreement on whether democracy in both states and a wide power imbalance overdetermined the peaceful outcome, whether they had cumulative effects, or whether one factor was more causal and the other more spurious.⁵⁷ This may be resolvable through more systematic analysis of process-tracing data, or careful counterfactual analysis, but likely will not be entirely determined to the satisfaction of a scholarly consensus.⁵⁸ The same is true of discussions on whether a large power disparity and a (perceived) absence of democracy

ory-Building," in Olav Njølstad, ed., *Arms Races: Technological and Political Dynamics* (Newbury Park, Calif.: Sage Publications, 1990), pp. 240–244. We disagree with Njølstad's suggestion that these approaches are substantially different from the standard methodological advice offered by those who have outlined the method of structured, focused comparisons between cases. See George, "Case Studies and Theory Development," and George and McKeown, "Case Studies and Theories of Organizational Decision Making," pp. 21–58.

55. See, respectively, Kevin Wang and James Lee Ray, "Beginners and Winners: The Fate of Initiators of Interstate Wars Involving Great Powers Since 1495," *International Studies Quarterly*, Vol. 38, No. 1 (March 1994), pp. 139–154; Robert Bates et al., *Analytic Narratives* (Princeton, N.J.: Princeton University Press, 1998); Peterson, "How Democracies Differ"; and Ray, *Democracies and International Conflict*.

56. Owen, "How Liberalism Produces Democratic Peace," pp. 87–125; and José Varela Ortega, "Aftermath of Splendid Disaster: Spanish Politics Before and After the Spanish-American War of 1898," *Journal of Contemporary History*, Vol. 15, No. 2 (April 1980), pp. 317–344.

57. Ray, *Democracies and International Conflict*; Peterson, "How Democracies Differ;" and Layne, "Kant or Cant."

58. Philin Tetlock and Aaron Belkin, eds., *Counterfactual Thought Experiments: Logical*

in Spain were both necessary conditions for the Spanish-American War.⁵⁹ In case study methods, as in statistical methods, scholars may at times have to live with some degree of indeterminacy when competing variables push in the same direction.

One disagreement that has been narrowed by additional research concerns the question of how to interpret Finland's decision to side with Germany against several democracies in World War II. Democratic peace proponents note that Finland did not undertake any offensive operations against democratic states, and the only attack against Finland by a democracy consisted of a single day of British bombing.⁶⁰ Critics argue that the Finnish case should be considered an important exception to the democratic peace because Finland became a co-belligerent with Germany and several democracies declared war on Finland.⁶¹ Miriam Elman's careful case study of the Finnish case suggests that more centralized or semi-presidential democracies like Finland are more likely than decentralized democracies to engage in war with other democracies. She indicates that the Finnish parliament resisted aligning with Germany, but was overruled by the Finnish president. Thus, while the case does not fit neorealist theories arguing that systemic pressures are paramount, neither does it strongly vindicate interdemocratic peace theories.⁶²

The Third Generation: Formal Modeling Contributions

Researchers have more recently begun to use formal models to help unravel the causal mechanisms that might explain the correlational and case study findings on the interdemocratic peace. We concentrate here on Kenneth Schultz's work on this subject, which provides an excellent exemplar both of formal work, and of multi-method research that tests a formal model with statistical and case study evidence.⁶³ Schultz frames his

59. Ray, *Democracies and International Conflict*.

60. Russett, "Correspondence: The Democratic Peace."

61. Spiro, "The Insignificance of the Liberal Peace."

62. Elman, ed., *Paths to Peace*, pp. 192–197.

63. Kenneth A. Schultz, *Democracy and Coercive Diplomacy* (Cambridge: Cambridge University Press, 2001). Additional formal modeling work relevant to the interdemocratic peace includes Bueno de Mesquita and Lalman, *War and Reason*; Bruce Bueno de Mesquita, James D. Morrow, Randolph M. Siverson, and Alastair Smith, "An Institutional Explanation of the Democratic Peace," *American Political Science Review*, Vol. 92, No. 3 (September 1998), pp. 623–638; and George W. Downs and David Rocke, "Conflict, Agency, and Gambling for Resurrection: The Principal-Agent Problem Goes

research around the question whether democratic institutions primarily constrain or inform decisions on the use of force. The constraint theory argues that democratic publics are reluctant to vote upon themselves the costs of war, and will vote against any democratic leaders who use force unsuccessfully or unjustifiably. An alternative theory, which Schultz favors, emphasizes that the transparency inherent in democracy makes it hard for democratic leaders to bluff; a threat to use force will lack credibility when a democratic leader's opposition party or the public does not wish to use force. At the same time, transparency makes a democratic leader's threat of force highly credible when the opposition party or the public support the use of force. In this view, democratic leaders are more selective than authoritarian leaders in their threats to use force, and when they do threaten to use force, these threats carry high credibility when the opposition party supports them and low credibility when the opposition party vocally objects to the threat of force.⁶⁴

Schultz provides a tight logic for his theory by developing it through a formal model of crisis bargaining that incorporates democratic leaders' preferences, opposition leaders' preferences, the information leaders have, and the signals they send to and receive from domestic audiences and the opposing actor in a crisis. This model highlights the bargaining problem, or the challenge that actors in a crisis face when they attempt to negotiate a peaceful outcome without complete information. In this view, the use of force, costly to all parties, is always to some degree suboptimal, as the side that ultimately loses on the battlefield would almost always have been better off conceding on the dispute prior to the costly resort to force (though if the public strongly favors war or the leader wants to have an international reputation for being "tough," there may be incentives to fight losing battles). Even when a negotiated outcome would be preferable for both parties, however, they may resort to force because they are unable to accurately assess each other's intentions and capabilities. Actors in a crisis have private information about their willingness and ability to fight, and they have incentives to misrepresent this information by bluffing to achieve a favorable outcome at the bargaining table. This is where the transparency of democratic politics enters in, helping to resolve the bargaining problem by making it hard for leaders to bluff but easy to issue credible threats when they have the support of the opposition party.

ies, see Andrew Kydd, "The Art of Shaker Modeling: Game Theory and Security Studies," in Detlef F. Sprinz and Yael Wolinsky-Nahmias, eds., *Models, Numbers, and Cases: Methods for Studying International Relations* (Ann Arbor: University of Michigan

Schultz tests the implications of his formal model through a statistical analysis of 1,785 cases of militarized interstate disputes. In this test, he finds strong evidence that democratic institutions decrease the probability of a crisis being initiated by a threat of force, decrease the likelihood of resistance to a threat if one is issued, and decrease the probability of war.⁶⁵ Schultz further tests his model against fifty-six cases in which states attempted to deter threats made against their allies, finding a tendency for democratic governments to be more successful in their deterrent threats when their opposition parties support them (though this finding falls short of standard levels of statistical significance).⁶⁶ Schultz then turns to case studies so that his analysis can provide "both a statistical correlation that is consistent with the argument . . . and historical evidence that the hypothesized causal mechanisms underlie this correlation."⁶⁷ Here, Schultz studies one case where the credibility of a democratic government's threat to use force is confirmed by the support of its opposition party (the British side of the Fashoda Crisis) and several cases where a democratic government decided against threatening force or issued a threat that was less credible because of objections from the opposition party (the French side of the Fashoda Crisis, the British threat and use of force in the 1899 Boer War, French and British behavior in the 1936 Rhineland Crisis, and British behavior in the 1956 Suez crisis and the 1965 Rhodesian crisis).

Schultz's work in each methodological approach is generally rigorous and well done. He devotes ample attention to alternative explanations, including explanations that emphasize democratic norms (nonviolence and respect for democratic regimes) and neorealist variables (particularly alliance portfolios), as well as the constraining and informing aspects of democratic institutions. Schultz is careful not to overstate his findings, and his case studies are convincing in showing that opposition parties played an important role in forestalling, bolstering, or undercutting democratic leaders' threats to use force. He is not as systematic in treating the outliers in his statistical and case study work, however. Out of the thirty-two cases of extended-immediate deterrence in which the defending state had a competitive political system, for example, ten had outcomes that do not fit Schultz's argument. Yet he only discusses one of these cases (the British-Greek crisis over Crete in 1897) for the purpose of arguing that it might deserve recoding in a way that would make it fit his thesis. Similarly, the cases chosen for individual study all fit the argu-

65. *Ibid.*, p. 158.

66. *Ibid.* pp. 169-170.

ment; this is defensible in the early stages of an innovative research program such as Schultz's where the goal is to illustrate as much as test the mechanisms that might explain a correlational finding, but even so he might have paid more attention to anomalous cases that might have helped to delimit the scope conditions of his theory. Schultz justifiably notes that the cases that do not fit his theory tend to be the more spectacular and memorable ones, resulting in wars rather than negotiated settlements. Yet after listing World War I, World War II, and the Vietnam War as "some of the most prominent international conflicts of the last century" and indicating that they do not fit his theory, he does not discuss how these anomalies might be explained or how they might limit his findings.⁶⁸ Despite this shortcoming, Schultz's successful effort to integrate different methods is one worthy of emulation, as it demonstrates that the value of carrying out statistical and case study tests of a formal model is worth the considerable difficulties involved in doing so.

One final example illustrates how the latest work on the interdemocratic peace has been able to build on prior statistical, case study, and formal research toward a more complete and integrated theory of the interdemocratic piece. Charles Lipson's *Reliable Partners* uses the insights developed in Schultz's work, as well as other findings from formal theories on bargaining, contracting, audience costs, self-binding, and transparency, to construct a model of the superior ability of democracies to create credible and enforceable commitments or contracts with one another that make it unnecessary to use costly military force to resolve disputes.⁶⁹ Lipson's model aspires to explain not only the interdemocratic peace, but many of the other findings that have emerged from the broader democratic peace research program. Lipson tests his model against numerous brief case studies and the results of existing statistical studies. His goal is largely to integrate existing studies rather than to carry out exhaustive and detailed primary research or develop and test a single statistical model. Because Lipson conscientiously considers alternative explanations throughout, and because he has so many excellent prior studies to draw upon, what emerges is the most convincing and complete treatment of the interdemocratic peace thus far.

68. Ibid., pp. 241–242. Schultz points out in a footnote, without elaboration, that his model predicts that such deterrence failures by united democracies should not happen, but that a different model he used in earlier work allows for the possibility of resistance against united democracies (note 1, p. 242); the referenced work is Kenneth A. Schultz, "Domestic Opposition and Signaling in International Crises," *American Politi-*

Methodological Suggestions for Future Research on the Interdemocratic Peace

We end this chapter by offering several suggestions for future research on the interdemocratic peace that will further enrich the development of typological theory on this subject. First, researchers can intensify efforts, like that undertaken by Braumoeller, to study states that have democratic institutions but lack democratic norms, as well as those that have democratic norms but lack democratic institutions. Researchers can then compare such cases to those that have both or neither of these attributes of democracy as a test of institutional and normative causal mechanisms.

Second, researchers can follow up Peterson's research on the interaction between leaders and publics by examining how leaders have tried to reconcile their own preferences with public opinion.

Third, researchers can look for other testable process-tracing implications of democratic peace assertions. For example, if norms and institutions affect the international use of force, they should also affect the conditions under which domestic police forces are allowed to use deadly force. William Hoelt, for example, has argued that the domestic police forces of democratic states are more likely to be allowed to use deadly force only to prevent the use of such force against themselves or others, whereas nondemocratic states allow the use of deadly force and of state-sanctioned executions for property crimes.⁷⁰ Also, researchers can look at civil-military relationships in democracies and in other regime types and at variations in civil-military relations among democracies and within democracies over time.⁷¹

Fourth, researchers can look at the origins of democratic norms and institutions and assess whether differing origins lead to different foreign policy behaviors. For example, do the foreign policies of democracies established through domestic revolutions against monarchs (France) differ from those created through anti-colonial uprisings (the United States), or those established through defeat in war and occupation by other democracies (Germany and Japan)? Does one democracy treat another differently depending on the origins of their respective norms and institutions?

Fifth, researchers might move beyond statistical, case study, and formal research to use surveys and other techniques to study the democratic peace. In particular, researchers might undertake surveys of the attitudes that elites and mass publics in democracies hold toward the use of force vis-à-vis other democracies and other types of regimes. Although recent

research like that by Schultz and Lipson has focused on institutional and informational dynamics, one possible data bias regarding cases prior to the 1940s is that no systematic survey data exists on public and elite opinion. Moreover, although standard surveys indicate that citizens of contemporary democracies generally feel more warmly toward other democracies than toward other kinds of states, little dedicated survey work has been done on attitudes toward the possible use of force in ongoing disputes between democracies. There is thus a danger that the role of democratic values in promoting peace among democracies has been understated in works that emphasize institutions and information, although surveys might also help validate the role of institutions and information as well.

Sixth, researchers can look more assiduously for closely matched pairs of democratic and mixed dyads that might be amenable to most similar research designs like Ray's study of the Spanish-American War and the Fashoda Crisis. One possibility here is to undertake longitudinal studies of particular dyads, as John Owen has done in the case of U.S.-British relations. This allows a before-after comparison of dyadic relations after domestic developments that make one partner in the dyad more democratic.⁷² Statistical methods can also carry out or augment such longitudinal comparisons.

Seventh, researchers might focus on the cases that pose anomalies to Schultz's theory that democracies find it hard to make convincing bluffs but easy to issue credible threats. This can help set the theory's scope conditions and perhaps uncover additional causal mechanisms that explain Schultz's anomalies.

Finally, researchers should look for relationships between democracies that have varying levels of power imbalances. This can test whether democratic norms function to the point of altruism or whether democracies are willing to exploit materially weaker democracies through the use or threat of force short of war.

In sum, the interdemocratic peace literature amply demonstrates the complementary nature of alternative methods and the value of combining them or using them sequentially for the research tasks to which they are best suited. We turn now to a detailed examination of the methods that have allowed case study researchers to contribute to the cumulation of knowledge in this and many other research programs.

A Pedagogical Note to Parts II and III

Readers of this book who are or will be teaching Ph.D.-level courses on qualitative methods may be interested in how the materials presented in Part II were developed. The origins of the method of structured, focused comparison were already described in some detail in the Preface. This note indicates, first, how the method was developed and tested in the Ph.D.-level research seminar Alex George taught over a period of years at Stanford. Then, a brief commentary is provided on Parts II and III, which follow, to indicate that they provide a manual for case study methods.

In the seminar, students first read the current description of the method. Then, each student selected a book of interest that consisted of a study of a single case or comparative cases. For this assignment each student employed the requirements of structured, focused comparison as a basis for critiquing the chosen book's methodology. Students prepared written evaluations of the relevance and utility of the structured, focused method's requirements for developing an incisive critique of their chosen book. Was the method useful for this purpose, and how might it be made more useful? After critiquing their chosen study in this way, students then consulted published reviews of that book to judge what their use of the structured, focused method had added. Generally, they found that use of structured, focused comparison added substantially to the published reviews. This assignment gave students useful hands-on experience with the method. It also contributed, together with classroom discussion, to the clarification and further development of the method.

For their second assignment, students prepared a research design on a problem they were considering as a possible topic for a Ph.D. disserta-

signs for a possible dissertation, what problems they encountered, and what they had learned from the experience. Each student's research design paper was discussed in the class and the writer of the paper then produced an addendum to his or her paper indicating what had been learned as a result of the discussion.

The *modus operandi* of the research seminar has been described in this note in order to indicate that the chapters that follow in Part II and Part III are the result of sustained efforts over a period of years to develop and refine the method of structured, focused comparison and related material in Parts Two and Three. Many students who took the seminar later drew upon that experience in their Ph.D. dissertations. The seminar became a required course at Stanford for all Ph.D. students in comparative politics and was taken by most Ph.D. students in international relations.

We emphasize in this book the critical importance of *research designs*. After a brief discussion of the essential components of the structured, focused method in Chapter 3, we discuss in Chapter 4, "Phase One: Designing Case Study Research," five interrelated requirements for developing effective research design. This chapter should be used as a reference guide to be read not just once, but as often as necessary; first, in initial efforts to develop a research design and, then, as needed to redesign one's research strategy to better approximate the desiderata set forth in the chapter. Readers planning to undertake case study research would be well advised to use the criteria for research design identified in that chapter to see how well they enable one to critique and build on existing publications of interest to them. In teaching these research seminars, we found it a quite useful first step to have students familiarize themselves with the challenge of good research design by applying these criteria as guidelines for reviewing existing studies.

Research design is an integral part of the method of structured, focused comparison. Readers should keep in mind, as emphasized in Chapter 4, that the guidelines for research design are intimately interrelated and must be integrated to produce an appropriate set of general questions to ask of each case in order to obtain the data needed to meet the study's research objectives. "Appropriate" general questions are those highly likely to provide the data from the case studies that will be needed when one turns to drawing conclusions from the cases that contribute to meeting the research objectives of the study. The reader's attention is called also to the Appendix, which describes the variety, flexibility, and ingenuity of research designs in some thirty studies within the field of American politics, comparative politics, and international relations. Reading these accounts will be helpful in designing one's own study.

vide good data. And Chapter 6, "Phase Three: Drawing the Implications of Case Study Findings for Theory," discusses various methods for using case results to meet the research objectives of a study.

Part III closely examines additional research methods available to case researchers, and presents chapters on process-tracing and typological theory, which we see as two of a researcher's most important tools for empirically identifying causal mechanisms and for modeling phenomena that reflect complex causation. Graduate students may wish to consult these chapters as they select their methods, and then later as a check that they are using their chosen methods in a disciplined way.

Chapter 7 in Part III surveys recent developments in philosophy of science that are relevant for theory-oriented case study research. We call attention in particular to the emergence of the scientific realism school, which supports the emphasis we give to the role of causal mechanisms in explanation and to within-case analysis and process-tracing.

Chapter 8 provides a detailed discussion of the limitations of "controlled comparison," which is still the standard comparative method. This chapter also discusses various ways to cope with these limitations. We offer an alternative to controlled comparison, the within-case method, which makes use of process-tracing in analyzing individual cases. Chapter 9 calls attention to another within-case method, the congruence method, which does not make use of process-tracing. Illustrations of both types of within-case analysis are provided.

Chapter 10 provides a detailed discussion of process-tracing, its different types and uses. Accompanying it is a discussion of similarities and differences between theory-oriented process-tracing and historical explanation.

Chapter 11 presents one of the most important contributions of our book: a discussion of how to develop typological theories of problems characterized by equifinality and complex causation. "Equifinality," a term used in general systems theory, is referred to by some scholars as multiple causality. It identifies a pervasive characteristic of social phenomena, namely the fact that *different* causal processes can lead to *similar* outcomes of a given dependent variable. Equifinality complicates the task of theory development and testing and must be taken into account in the design and implementation of all research, not just case study investigation. We emphasize also that many real-world problems are characterized by considerable causal complexity, which also complicates the task of theory development. Both equifinality and causal complexity are discussed in detail at various points in the book. Both can be dealt with effectively in theory-oriented case study research that develops more lim-

Finally, we note in Chapter 12 the kind of theory for which case study research is particularly applicable. This is “middle-range” theory, to distinguish it from efforts to develop and apply broad-spanning paradigmatic theories such as realism, liberalism, and constructivism. In contrast, middle-range theory focuses on specific subtypes of a general phenomenon—for example, not all but each specific type of military intervention and not all but each type of effort to employ a particular variant of coercive diplomacy. This approach contributes greatly by filling in the theoretical vacuum left by these general paradigmatic models.

Middle-range theories carefully delimit the scope of their findings to each particular subclass of a general phenomenon. Individual middle-range theories of each specific subclass constitute building blocks for constructing broader but also internally differentiated theories of a general phenomenon. Middle-range theories, as noted in Chapter 12, are particularly relevant for the development of policy-relevant theoretical findings—or “generic knowledge,” as they are sometimes called—of strategies and problems repeatedly encountered in different contexts in the conduct of foreign policy.

In sum, Parts II and III provide a manual for developing theory through a variety of case study methods. We have attempted to make this manual as “user-friendly” as possible. We hope that it provides an important, usable approach for efforts to raise the standards for case study research and to explicate the procedures for doing so, the two objectives for our study we have pursued for several decades.

Part II
How to Do Case Studies

Chapter 3

The Method of Structured, Focused Comparison

The method and logic of structured, focused comparison is simple and straightforward. The method is "structured" in that the researcher writes general questions that reflect the research objective and that these questions are asked of each case under study to guide and standardize data collection, thereby making systematic comparison and cumulation of the findings of the cases possible. The method is "focused" in that it deals only with certain aspects of the historical cases examined. The requirements for structure and focus apply equally to individual cases since they may later be joined by additional cases.

The method was devised to study historical experience in ways that would yield useful generic knowledge of important foreign policy problems. The particular challenge was to analyze phenomena such as deterrence in ways that would draw the explanations of each case of a particular phenomenon into a broader, more complex theory. The aim was to discourage decision-makers from relying on a single historical analogy in dealing with a new case.¹

1. This discussion draws upon earlier publications: Alexander L. George, "Case Studies and Theory Development: The Method of Structured, Focused Comparison," in Paul Gordon Lauren, ed., *Diplomacy: New Approaches in History, Theory, and Policy* (New York: Free Press, 1979), pp. 43-68; Alexander L. George, "The Causal Nexus Between Cognitive Beliefs and Decision-Making Behavior," in Lawrence S. Falkowski, ed., *Psychological Models in International Politics* (Boulder, Colo.: Westview Press, 1979), pp. 95-124; and Alexander L. George and Timothy J. McKeown, "Case Studies and Theories of Organizational Decision Making," in Robert F. Coulam and Richard A. ...

Before we discuss each of these two characteristics of structured, focused comparison, it will be instructive to show how they improve upon previous case study approaches. Following the end of World War II, many political scientists were quite favorably disposed toward or even enthusiastic about the prospect of undertaking individual case studies for the development of knowledge and theory. Many case studies were conducted, not only in the field of international relations but also in public administration, comparative politics, and American politics. Although individual case studies were often instructive, they did not lend themselves readily to strict comparison or to orderly cumulation. As a result, the initial enthusiasm for case studies gradually faded, and the case study as a strategy for theory development fell into disrepute.² In 1968 James Rosenau critiqued case studies of foreign policy and called attention to their nonscientific, noncumulative character. These studies of foreign policy by political scientists and historians, Rosenau observed, were not conducted in ways appropriate for scientific inquiry. In his view, most of them lacked “scientific consciousness” and did not accumulate. Individual studies may have made interesting contributions to knowledge, but a basis for systematic comparison was lacking.³

his *Organizing for Foreign Policy Crisis* (Ann Arbor, Mich.: University of Michigan Press, 1997). Haney develops ways of surveying cases that are capable of combining the advantages of structured, focused comparison with large-N analysis. He suggests that the findings of a number of studies that address the same problem can be combined and the results averaged—i.e., a form of what statisticians refer to as “meta-analysis.” This particular case survey method was proposed earlier by Robert Yin and Karen A. Heald, “Using the Case Survey Method to Analyze Policy Studies,” *Administrative Science Quarterly*, Vol. 20, No. 3 (September 1975), pp. 371–381. The rather obvious limitations of the case survey approach are noted by Yin and Heald.

A cogent statement of key research steps in small-n research is provided by Ronald Mitchell and Thomas Bernauer in “Empirical Research in International Environmental Policy: Designing Qualitative Case Studies,” *Journal of Environment and Development*, Vol. 7, No. 1 (March 1998), pp. 4–31.

Dwayne Medford outlines a way of extending and generalizing structured, focused comparisons that focus on the actor’s cognitive processes in Charles F. Hermann, Charles W. Kegley, Jr., and James N. Rosenau, eds., *New Directions in the Study of Foreign Policy* (Boston, Mass.: Allen & Unwin, 1987).

See also our commentary on the important work by Thomas Homer-Dixon in the Appendix, “Studies That Illustrate Research Design.”

2. Of course, as noted in Chapter 10, well-researched case studies that are largely descriptive and atheoretical are useful in providing a form of vicarious experience for students and others interested in a particular phenomenon, and sometimes they provide data that can be of some use in case studies devoted to theory development.

3. James N. Rosenau, “Moral Emancipation,” *International Studies Quarterly*, Vol. 32, No. 1, 1988, pp. 1–20.

Writers in other fields of political science offered similar critiques of extant case studies. In 1955, Roy Macridis and Bernard Brown criticized the old "comparative politics" for being, among other things, not genuinely comparative. These earlier studies consisted mainly of single case studies which were often essentially descriptive and monographic rather than theory-oriented. In the field of public administration, similar concerns were expressed, and, in the field of American politics, an important critique of the atheoretical case study was presented by Theodore Lowi.⁴

What, then, are some of the requirements that case study research must meet to overcome these difficulties?

First, the investigator should clearly identify the universe—that is, the "class" or "subclass" of events—of which a single case or a group of cases to be studied are instances. Thus, the cases in a given study must all be instances, for example, of only one phenomenon: either deterrence, coercive diplomacy, crisis management, alliance formation, war termination, the impact of domestic politics on policymaking, the importance of personality on decision-making, or whatever else the investigator wishes to study and theorize about. The identification of the class or subclass of events for any given study depends upon the problem chosen for study.

Second, a well-defined research objective and an appropriate research strategy to achieve that objective should guide the selection and analysis of a single case or several cases within the class or subclass of the phenomenon under investigation. Cases should not be chosen simply because they are "interesting" or because ample data exist for studying them.

Third, case studies should employ variables of theoretical interest for purposes of explanation. These should include variables that provide some leverage for policymakers to enable them to influence outcomes.

We turn now to a discussion of the two characteristics of the method of structured, focused comparison. From the statistical (and survey) research model, the method of structured, focused comparison borrows the device of asking a set of standardized, general questions of each case, even in single case studies. These questions must be carefully developed to reflect the research objective and theoretical focus of the inquiry. The use of a set of general questions is necessary to ensure the acquisition of comparable data in comparative studies. This procedure allows researchers to avoid the all too familiar and disappointing pitfall of traditional, in-

4. Roy C. Macridis and Bernard E. Brown, eds., *Comparative Politics: Notes and Readings* (Homewood, Ill.: Dorsey Press, 1955); Herbert Kaufmann, "The Next Step in Case Studies," *Public Administration Review*, Vol. 18 (Winter 1958), pp. 52–59; and Theodore

tensive single case studies. Even when such cases were instances of a class of events, they were not performed in a comparable manner and hence did not contribute to an orderly, cumulative development of knowledge and theory about the phenomenon in question. Instead, each case study tended to go its own way, reflecting the special interests of each investigator and often being unduly shaped by whatever historical data was readily available. As a result, idiosyncratic features of each case or the specific interests of each investigator tended to shape the research questions. Not surprisingly, single case studies—lacking “scientific consciousness”—did not accumulate.

The method also requires that the study of cases be “focused”: that is, they should be undertaken with a specific research objective in mind and a theoretical focus appropriate for that objective. A single study cannot address all the interesting aspects of a historical event. It is important to recognize that a single event can be relevant for research on a variety of theoretical topics. For example, the Cuban Missile Crisis offers useful material for developing many different theories. This case may be (indeed, has been) regarded and used as an instance of deterrence, coercive diplomacy, crisis management, negotiation, domestic influence on foreign policy, personality involvement in decision-making, etc. Each of these diverse theoretical interests requires the researcher to adopt a different focus, to develop and use a different theoretical framework, and to identify a different set of data requirements. A researcher’s treatment of a historical episode must be selectively focused in accordance with the type of theory that the investigator is attempting to develop.

One reason so many case studies of a particular phenomenon in the past did not contribute much to theory development is that they lacked a clearly defined and common focus. Different investigators engaged in research on a particular phenomenon tended to bring diverse theoretical (and nontheoretical) interests to bear on their case studies. Each case study tended to investigate somewhat different dependent and independent variables. Moreover, many case studies were not guided by a well-defined theoretical objective. Not surprisingly, later researchers who had a well-defined theoretical interest in the phenomenon often found that earlier studies were of little value for their purposes.

It is important for researchers to build self-consciously upon previous studies and variable definitions as much as possible—including studies using formal, statistical, and qualitative methods. “Situating” one’s research in the context of the literature is key to identifying the contribution the new research makes. Of course, researchers will sometimes find it necessary to modify existing definitions of variables or add

edge that this reduces the comparability to or cumulativeness with previous studies.

It should be noted that a merely formalistic adherence to the format of structured, focused comparison will not yield good results. The important device of formulating a set of standardized, general questions to ask of each case will be of value only if those questions are grounded in—and adequately reflect—the theoretical perspective and research objectives of the study. Similarly, a selective theoretical focus for the study will be inadequate by itself unless coupled with a relevant set of standardized general questions.

In comparative case studies, structure and focus are easier to achieve if a single investigator not only plans the study, but also conducts all of the case studies. Structured, focused comparison is more difficult to carry out in collaborative research when each case study is undertaken by a different scholar. Collaborative studies must be carefully planned to impress upon all participants the requirements of structure and focus. The chief investigator must monitor the conduct of case studies to ensure that the guidelines are observed by the case writers and to undertake corrective actions if necessary. Properly coordinating the work of case writers in a collaborative study can be a challenging task for the chief investigator, particularly when the contributors are well-established scholars with views of their own regarding the significance of the case they are preparing.

This can be seen in comparing two collaborative studies. One study of Western democratic political opposition brought together a distinguished group of scholars, each studying the democratic opposition in a Western democracy. The study was not tightly organized to meet the requirement of a structured comparison, so the organizer of the study was left with the difficult task of drawing together the disparate findings of the individual case studies for comparative analysis in the concluding chapter.⁵ In contrast, Michael Krepon and Dan Caldwell developed a tight version of structured, focused comparison for their collaborative study of cases of U.S. Senate ratification of arms control treaties. They

5. Robert A. Dahl, *Political Oppositions in Western Democracies* (New Haven, Conn.: Yale University Press, 1966). As Sidney Verba notes in his detailed commentary on this book, it "highlights a problem that arises in the multiauthored book. There are great advantages in having a large number of country specialists, but specialists are hard to discipline. In *Political Oppositions*, the major theoretical chapters that attempt to tie together the individual country chapters are found at the end of the book. . . . If we want to have as collaborators men of the stature of the authors of this book, we must let them go their own way." Sidney Verba, "Some Dilemmas in Comparative Research,"

closely monitored the individual authors' adherence to the guidelines and intervened as necessary to ensure that they adhered to the original or revised guidelines.⁶

The next chapter provides a more specific discussion of procedures for the design and implementation of case studies—either single case analyses or comparative investigations that are undertaken within the framework of the structured, focused method.

6. Michael Krepon and Dan Caldwell, eds., *The Politics of Arms Control Treaty Ratification* (New York: St. Martin's Press, 1991). We are indebted to Michael Krepon for providing us with a detailed account of how he and Caldwell accomplished this difficult task.

Chapter 4

Phase One: Designing Case Study Research

There are three phases in the design and implementation of theory-oriented case studies. In phase one, the objectives, design, and structure of the research are formulated. In phase two, each case study is carried out in accordance with the design. In phase three, the researcher draws upon the findings of the case studies and assesses their contribution to achieve the research objective of the study. These three phases are interdependent, and some iteration is often necessary to ensure that each phase is consistent and integrated with the other phases.¹ The first phase is discussed in this chapter, and phases two and three in the chapters that follow.

Phase one—the research design—consists of five tasks. These tasks are relevant not only for case study methodology but for all types of systematic, theory-oriented research. They must be adapted, of course, to different types of investigation and to whether theory testing or theory development is the focus of the study. The design phase of theory-oriented case study research is of critical importance. If a research design

1. The procedure of organizing such studies on the basis of these three phases was introduced by Alexander L. George and Richard Smoke in their book *Deterrence in American Foreign Policy: Theory and Practice* (New York: Columbia University Press, 1974). It has proven to be a useful organizing device in subsequent studies and has also provided a framework for reviewing and evaluating existing studies. We are omitting here a fourth phase, presentation of the results of the study, that was mentioned in Alexander L. George and Timothy J. McKeown, "Case Studies and Theories of Organizational Decision Making," in Robert F. Coulam and Richard A. Smith, eds., *Advances in Information Processing in Organizations*. Vol. 2 (Greenwich, Conn.: JAI Press,

proves inadequate, it will be difficult to achieve the research objectives of the study. (Of course, the quality of the study depends also on how well phases two and three are conducted.)

Task One: Specification of the Problem and Research Objective

The formulation of the research objective is the most important decision in designing research. It constrains and guides decisions that will be made regarding the other four tasks.

The selection of one or more objectives for research is closely coupled with identification of an important research problem or "puzzle." A clear, well-reasoned statement of the research problem will generate and focus the investigation. A statement that merely asserts that "the problem is important" is inadequate. The problem should be embedded in a well-informed assessment that identifies gaps in the current state of knowledge, acknowledges contradictory theories, and notes inadequacies in the evidence for existing theories. In brief, the investigator needs to make the case that the proposed research will make a significant contribution to the field.

The research objective must be adapted to the needs of the research program at its current stage of development. Is there a need for testing a well-established theory or competing theories? Is it important to identify the limits of a theory's scope? Does the state of research on the phenomenon require incorporation of new variables, new subtypes, or work on different levels of analysis? Is it considered desirable at the present stage of theory development to move up or down the ladder of generality?² For example, as noted in Chapter 2, in the 1990s the democratic peace research program moved largely from the question of whether such a peace existed to that of identifying the basis on which democratic peace rests. It now needs to go further to explain how a particular peace between two democratic states developed over time. Similarly, in the 1960s deterrence theory needed to bring in additional variables to add to excessively parsimonious and abstract deductive models.

In general, there are six different kinds of theory-building research objectives. Arend Lijphart and Harry Eckstein identified five types. We outline these below and add a sixth type of our own.³

2. Giovanni Sartori, "Concept Misformation in Comparative Politics," *American Political Science Review*, Vol. 64, No. 4 (December 1970), pp. 1033–1053.

3. Arend Lijphart, "Comparative Politics and the Comparative Method," *American*

- *Atheoretical/configurative idiographic* case studies provide good descriptions that might be used in subsequent studies for theory building, but by themselves, such cases do not cumulate or contribute directly to theory.
- *Disciplined configurative* case studies use established theories to explain a case. The emphasis may be on explaining a historically important case, or a study may use a case to exemplify a theory for pedagogical purposes. A disciplined configurative case can contribute to theory testing because it can “impugn established theories if the theories ought to fit it but do not,” and it can serve heuristic purposes by highlighting the “need for new theory in neglected areas.”⁴ However, a number of important methodological questions arise in using disciplined configurative case studies and these are discussed in Chapter 9 on the congruence method.
- *Heuristic* case studies inductively identify new variables, hypotheses, causal mechanisms, and causal paths. “Deviant” or “outlier” cases may be particularly useful for heuristic purposes, as by definition their outcomes are not what traditional theories would anticipate. Also, cases where variables co-vary as expected but are at extremely high or low values may help uncover causal mechanisms.⁵ Such cases may not allow inferences to wider populations if relationships are nonlinear or involve threshold effects, but limited inferences might be possible if causal mechanisms are identified (just as cancer researchers use high dosages of potential carcinogens to study their effects).
- *Theory testing* case studies assess the validity and scope conditions of single or competing theories. As discussed in Chapter 6, it is important in tests of theories to identify whether the test cases are most-likely, least-likely, or crucial for one or more theories. Testing may also be devised to identify the scope conditions of theories (the conditions under which they are most- and least-likely to apply).
- *Plausibility probes* are preliminary studies on relatively untested theories and hypotheses to determine whether more intensive and laborious testing is warranted. The term “plausibility probe” should not be used too loosely, as it is not intended to lower the standards of evidence and inference and allow for easy tests on most-likely cases.

son Polsby, eds., *Handbook of Political Science*, Vol. 7 (Reading, Mass.: Addison-Wesley, 1975), pp. 79–138.

4. Eckstein, “Case Studies and Theory,” p. 99.

- “*Building Block*” studies of particular types or subtypes of a phenomenon identify common patterns or serve a particular kind of heuristic purpose. These studies can be component parts of larger contingent generalizations and typological theories. Some methodologists have criticized single-case studies and studies of cases that do not vary in their dependent variable.⁶ However, we argue that single-case studies and “no variance” studies of multiple cases can be useful if they pose “tough tests” for theories or identify alternative causal paths to similar outcomes when equifinality is present.⁷ (See also the more detailed discussion of “building blocks” theory below.)

Researchers should clearly identify which of these six types of theory-building is being undertaken in a given study; readers should not be left to find an answer to this question on their own. The researcher may fail to make it clear, for example, whether the study is an effort at theory testing or merely a plausibility probe. Or the researcher may fail to indicate whether and what kind of “tough test” of the theory is supposedly being conducted.⁸

These six research objectives vary in their uses of induction and deduction. Also, a single research design may be able to accomplish more than one purpose—such as heuristic and theory testing goals—as long as it is careful in using evidence and making inferences in ways appropriate to each research objective. For example, while it is not legitimate to derive a theory from a set of data and then claim to test it on the same data, it is sometimes possible to test a theory on different data, or new or previously unobserved facts, from the same case.⁹

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

7. David Collier, “Translating Quantitative Methods for Qualitative Researchers: The Case of Selection Bias,” *American Political Science Review*, Vol. 89, No. 2 (June 1995), pp. 461–466; and Ronald Rogowski, “The Role of Theory and Anomaly in Social-Science Inference,” *American Political Science Review*, Vol. 89, No. 2 (June 1995), pp. 467–470. Theory development via building blocks is useful also in the absence of equifinality. Contingent generalizations are possible, and indeed easier to formulate, when equifinality is not present. For an example of this approach see George and Smoke, *Deterrence in American Foreign Policy*.

8. Joseph Grieco criticizes Robert O. Keohane’s *After Hegemony: Cooperation and Discord in the World Political Economy* (Princeton, N.J.: Princeton University Press, 1984) on both counts in his detailed criticism of the research design in this important study, to which Keohane replies in David A. Baldwin, ed., *Neorealism and Neoliberalism: The Con-*

Specific questions that need to be addressed in designating the research objectives include:

- What is the phenomenon or type of behavior that is being singled out for examination; that is, what is the class or subclass of events of which the cases will be instances?
- Is the phenomenon to be explained thought to be an empirical universal (i.e., no variation in the dependent variable), so that the research problem is to account for the lack of variation in the outcomes of the cases? Or is the goal to explain an observable variation in the dependent variable?
- What theoretical framework will be employed? Is there an existing theory or rival candidate theories that bear on those aspects of the phenomenon or behavior that are to be explained? If not, what provisional theory or theories will the researcher formulate for the purpose of the study? If provisional theories are lacking, what theory-relevant variables will be considered?
- Which aspects of the existing theory or theories will be singled out for testing, refinement, or elaboration?
- If the research objective is to assess the causal effects or the predictions of a particular theory (or independent variable), is that theory sufficiently specified and operationalized to enable it to make specific predictions, or is it only capable of making probabilistic or indeterminate predictions? What other variables and/or conditions need to be taken into account in assessing its causal effects?

Researchers' initial efforts to formulate research objectives for a study often lack sufficient clarity or are too ambitious. Unless these defects are corrected, the study will lack a clear focus, and it will probably not be possible to design a study to achieve the objectives.

Better results are achieved if the "class" of the phenomenon to be investigated is not defined too broadly. Most successful studies, in fact, have worked with a well-defined, smaller-scope *subclass* of the general phenomenon.¹⁰ Case study researchers often move down the "ladder of generality" to contingent generalizations and the identification of more circumscribed scope conditions of a theory, rather than up toward broader but less precise generalizations.¹¹

10. For illustrative examples, see the Appendix, "Studies That Illustrate Research Design."

11. A similar point is made by Robert Kachera in his critique of structural equation

Working with a specified subclass of a general phenomenon is also an effective strategy for theory development. Instead of trying in one study to develop a general theory for an entire phenomenon (e.g., all “military interventions”), the investigator should think instead of formulating a typology of different kinds of interventions and proceed to choose one type or subclass of interventions for study, such as “protracted interventions.” Or the study may focus on interventions by various policy instruments, interventions on behalf of different goals, or interventions in the context of different alliance structures or balances of power. The result of any single circumscribed study will be one part of an overall theory of intervention. Other studies, focusing on different types or subclasses of intervention, will be needed to contribute to the formulation of a general theory of interventions, if that is the broader, more ambitious research program. If the typology of interventions identifies six major kinds of intervention that are deemed to be of theoretical and practical interest, each subtype can be regarded as a candidate for separate study and each study will investigate instances of that subtype.

This approach to theory development is a “building block” procedure. Each block—a study of each subtype—fills a “space” in the overall theory or in a typological theory. In addition, the component provided by each building block is itself a contribution to theory; though its scope is limited, it addresses the important problem or puzzle associated with the type of intervention that led to the selection and formulation of the research objective. Its generalizations are more narrow and contingent than those of the general “covering laws” variety that some hold up as the ideal, but they are also more precise and may involve relations with higher probabilities.¹² In other words, the building block developed for a subtype is self sufficient; its validity and usefulness do not depend upon the existence of other studies of different subclasses of that general phenomenon.

If an investigator wishes to compare and contrast two or more different types of intervention, the study must be guided by clearly defined puzzles, questions, or problems that may be different from or similar to those of a study of a single subclass. For example, the objective may be to discover under what conditions (and through what paths) Outcome X occurs, and under what conditions (and through what paths) Outcome Y

ory.” Robert O. Keohane, ed., *Neorealism and Its Critics* (New York: Columbia University Press, 1986), pp. 187–188.

12. For example, see the discussion in the Appendix of Ariel Levite, Bruce Jentleson, and Larry Berman, eds., *Foreign Military Intervention: The Dynamics of Protracted Conflict*

occurs. Alternatively, the objective may be to examine under what conditions Policy A leads to Outcome Y and under what other conditions Policy A leads to Outcome X. Similarly, the focus may be on explaining the outcome of a case or a subclass or type of cases, or it may be on explaining the causal role of a particular independent variable across cases.

Task Two: Developing a Research Strategy: Specification of Variables

In the course of formulating a research objective for the study—which may change during the study—the investigator also develops a *research strategy* for achieving that objective. This requires early formulation of hypotheses and consideration of the elements (conditions, parameters, and variables) to be employed in the analysis of historical cases. Several basic decisions (also subject to change during the study) must be made concerning questions such as the following:

- What exactly and precisely is the dependent (or outcome) variable to be explained or predicted?
- What independent (and intervening) variables comprise the theoretical framework of the study?
- Which of these variables will be held constant (serve as parameters) and which will vary across cases included in the comparison?

The specification of the problem in Task One is closely related to the statement of what exactly the dependent variable will be. If a researcher defines the problem too broadly, he or she risks losing important differences among cases being compared. If a researcher defines the problem too narrowly, this may severely limit the scope and relevance of the study and the comparability of the case findings.¹³ As will be noted, the definition of variance in the dependent variable is critical in research design.

In analyzing the phenomenon of “war termination,” for instance, a researcher would specify numerous variables. The investigator would decide whether the dependent (outcome) variable to be explained (or predicted) was merely a cease-fire or a settlement of outstanding issues over which the war had been fought. Variables to be considered in explaining the success or failure of war termination might include the fighting capabilities and morale of the armed forces, the availability of

13. This research dilemma is discussed by Sidney Verba in his detailed commentary on Robert A. Dahl, ed., *Political Opposition in Western Democracies* (New Haven, Conn.:

economic resources for continuing the war, the type and magnitude of pressures from more powerful allies, policymakers' expectation that the original war aim was no longer attainable at all or only at excessive cost, the pressures of pro-war and anti-war opinion at home, and so on. The researcher might choose to focus on the outcome of the dependent variable (e.g., on cases in which efforts to achieve a cease-fire or settlement failed, but adding cases of successful cease-fires or settlements for contrast) to better identify the independent and intervening variables associated with such failures. Alternatively, one might vary the outcome, choosing cases of both successes and failures in order to identify the conditions and variables that seem to account for differences in outcomes.

Alternatively, the research objective may focus not on outcomes of the dependent variable, but on the importance of an independent variable—e.g., war weariness—in shaping outcomes in a number of cases.

We conclude this discussion of Task Two with a brief review of the strengths and weaknesses of the common types of case study research designs in relation to the kinds of research objectives noted above.

First, single case research designs can fall prey to selection bias or over-generalization of results, but all of the six theory-building purposes identified above have been served by studies of single well-selected cases that have avoided or minimized such pitfalls. Obviously, single-case studies rely almost exclusively on within-case methods, process-tracing, and congruence, but they may also make use of counterfactual analysis to posit a control case.¹⁴

For theory testing in single cases, it is imperative that the process-tracing procedure and congruence tests be applied to a wide range of alternative hypotheses that theorists and even participants in the events have proposed, not only to the main hypotheses of greatest interest to the researcher. Otherwise, left-out variables may threaten the validity of the research design. Single cases serve the purpose of theory testing particularly well if they are "most-likely," "least-likely," or "crucial" cases. Prominent case studies by Arend Lijphart, William Allen, and Peter Kourevitch, for example, have changed entire research programs by impugning theories that failed to explain their most-likely cases.¹⁵

14. David Laitin, "Disciplining Political Science," *American Political Science Review*, Vol. 89, No. 2 (June 1995), pp. 454-456. We say "almost" since single case studies take place within the context of ongoing research programs, so that studies of single cases may draw comparisons to existing studies; thus, "the community of scientists," rather than the "individual researcher" is the relevant context in which to judge case selection.

15. Bergewski, "The Role of Theory and Anomaly in Social Scientific Inference".

Similarly, studies of single “deviant” cases and of single cases where a variable is at an extreme value can be very useful for heuristic purposes of identifying new theoretical variables or postulating new causal mechanisms. Single-case studies can also serve to reject variables as being necessary or sufficient conditions.¹⁶

Second, the research objective chosen in a study may require comparison of several cases. There are several comparative research designs. The best known is the method of “controlled comparison”—i.e., the comparison of “most similar” cases which, ideally, are cases that are comparable in all respects except for the independent variable, whose variance may account for the cases having different outcomes on the dependent variable. In other words, such cases occupy neighboring cells in a typology, but only if the typological space is laid out one change in the independent variable at a time. (See Chapter 11 on typological theories.)

As we discuss in Chapter 8 on the comparative method, controlled comparison can be achieved by dividing a single longitudinal case into two—the “before” case and an “after” case that follows a discontinuous change in an important variable. This may provide a control for many factors and is often the most readily available or strongest version of a most-similar case design. This design aims to isolate the difference in the observed outcomes as due to the influence of variance in the single independent variable. Such an inference is weak, however, if the posited causal mechanisms are probabilistic, if significant variables are left out of the comparison, or if other important variables change in value from the “before” to the “after” cases.

However, even when two cases or before-after cases are not perfectly matched, process-tracing can strengthen the comparison by helping to assess whether differences other than those in the main variable of interest might account for the differences in outcomes. Such process-tracing can focus on the standard list of potentially “confounding” variables identified by Donald Campbell and Julian Stanley, including the effects of history, maturation, testing, instrumentation, regression, selection, and mortality.¹⁷ It can also address any idiosyncratic differences between the two

Seizure of Power: The Experience of a Single German Town, 1930–1935 (New York: Watts, 1965); and Peter Alexis Gourevitch, “The International System and Regime Formation: A Critical Review of Anderson and Wallerstein,” *Comparative Politics*, Vol. 10, No. 3 (April 1978), pp. 419–438.

16. For an example, see Lijphart’s study summarized in the Appendix, “Studies That Illustrate Research Design”; Douglas Dion, “Evidence and Inference in Comparative Case Study,” *Comparative Politics*, Vol. 3, No. 2 (January 1998); and Collier, “Trans-

cases that scholars or participants have argued might account for their differences.

Another comparative design involves "least similar" cases and parallels John Stuart Mill's method of agreement.¹⁸ Here, two cases are similar in outcome but differ in all but one independent variable, and the inference might be made that this variable contributes to the invariant outcome. For example, if teenagers are "difficult" in both postindustrial societies and tribal societies, we might infer that their developmental stage, and not their societies or their parents' child-rearing techniques, account for their difficult natures. Here again, left-out variables can weaken such an inference, as Mill recognized, but process-tracing provides an additional source of evidence for affirming or infirming such inferences.

Another type of comparative study may focus on cases in the same cell of a typology. If these have the same outcome, process-tracing may still reveal different causal paths to that outcome. Conversely, multiple studies of cases with the same level of a manipulable independent variable can establish under what conditions that level of the variable is associated with different outcomes. In either approach, if outcomes differ within the same type or cell, it is necessary to look for left-out variables and perhaps create a new subtype.

Often, it is useful for a community of researchers to study or try to identify cases in all quadrants of a typology. For example, Sherlock Holmes once inferred that a dog that did not bark must have known the person who entered the dog's house and committed a murder, an inference based on a comparison to dogs that do bark in such circumstances. To fully test such an assertion, we might also want to consider the behavior of non-barking non-dogs on the premises (was there a frightened cat?) and barking non-dogs (such as a parrot). The process of looking at all the types in a typology corresponds with notions of Boolean algebra and those of logical truth tables.¹⁹ However, it is not necessary for each researcher to address all the cells in a typology, although it is often useful

Designs for Research (Chicago: Rand McNally College Publishing, 1963); for a good example, see James Lee Ray, *Democracies and International Conflict: An Evaluation of the Democratic Peace Proposition* (Columbia: University of South Carolina Press, 1995), pp. 158-200.

18. For a detailed discussion of Mill's methods, see Chapter 8.

19. Charles C. Ragin, *The Comparative Method* (Berkeley: University of California Press, 1987); Daniel Little, *Varieties of Social Explanation: An Introduction to the Philosophy of Science* (Boulder, Colo.: Westview Press, 1991); and Daniel Little, *Microfoundations, Method, and Causation: On the Philosophy of the Social Sciences* (New Brunswick, N.J.:

for researchers to offer suggestions for future research on unexamined types or to make comparisons to previously examined types.

Finally, a study that includes many cases may allow for several different types of comparisons. One case may be most similar to another and both may be least similar to a third case. As noted below, case selection is an opportunistic as well as a structured process—researchers should look for whether the addition of one or a few cases to a study might provide useful comparisons or allow inferences on additional types of cases.

Task Three: Case Selection

Many students in the early stages of designing a study indicate that they find it difficult to decide which cases to select. This difficulty usually arises from a failure to specify a research objective that is clearly formulated and not overly ambitious. One should select cases not simply because they are interesting, important, or easily researched using readily available data. Rather, case selection should be an integral part of a good research strategy to achieve well-defined objectives of the study. Hence, the primary criterion for case selection should be relevance to the research objective of the study, whether it includes theory development, theory testing, or heuristic purposes.

Cases should also be selected to provide the kind of control and variation required by the research problem. This requires that the universe or subclass of events be clearly defined so that appropriate cases can be selected. In one type of comparative study, for example, all the cases must be instances of the same subclass. In another type of comparative study that has a different research objective, cases from different subclasses are needed.

Selection of a historical case or cases may be guided by a typology developed from the work in Tasks One and Two. Researchers can be somewhat opportunistic here—they may come across a pair of well-matched before-after cases or a pair of cases that closely fit “most similar” or “least similar” case research designs. They may also come upon cases that have many features of a most- or least-likely case, a crucial case, or a deviant case.

Often researchers begin their inquiry with a theory in search of a test case or a case in search of a theory for which it is a good test.²⁰ Either approach is viable, provided that care is taken to prevent case selection bias and, if necessary, to study several cases that pose appropriate tests for a

candidate theory once one is identified. Often, the researcher might start with a case that interests her, be drawn to a candidate theory, and then decide that she is more interested in the theory than in the case and conclude that the best way to study the theory is to select several cases that may not include the case with which the inquiry began. Some such iteration is usually necessary—history may not provide the ideal kind of cases to carry out the tests or heuristic studies that a research program most needs at its current stage of development.

Important criticisms have been made of potential flaws in case selection in studies with one or a few cases; such concerns are influenced by the rich experience of statistical methods for analyzing a large-N. David Collier and James Mahoney have taken issue with some widespread concerns about selection bias in small studies; we note four of their observations.²¹ They question the assertion that selection bias in case studies is potentially an even greater problem than is often assumed (that it may not just understate relationships—the standard statistical problem—but may overstate them). They argue that case study designs with no variance in the dependent variable do not inherently represent a selection bias problem. They emphasize that case study researchers sometimes have good reasons to narrow the range of cases studied, particularly to capture heterogeneous causal relations, even if this increases the risk of selection bias. They point out (as have we) that case study researchers rarely “overgeneralize” from their cases; instead, they are frequently careful in providing circumscribed “contingent generalizations” that subsequent researchers should not mistakenly overgeneralize.

Task Four: Describing the Variance in Variables

The way in which variance is described is critical to the usefulness of case analyses in furthering the development of new theories or the assessment or refinement of existing theories. This point needs emphasis because it is often overlooked in designing studies—particularly statistical studies of a large-N. The researcher’s decision about how to describe variance is important for achieving research objectives because the discovery of potential causal relationships may depend on how the variance in these variables is postulated. Basing this decision on *a priori* judgments may be risky and unproductive; the investigator is more likely to develop sensitive ways of describing variance in the variables after he or she has become familiar with how they vary in the historical cases examined. An it-

21. David Collier and James Mahoney, “Insights and Pitfalls: Selection Bias in Quali-

erative procedure for determining how best to describe variance is therefore recommended.²²

The variance may in some instances be best described in terms of qualitative types of outcomes. In others, it may be best described in terms of quantitative measures. In either case, one important question is how many categories to establish for the variables. Fewer categories—such as dichotomous variables—are good for parsimony but may lack richness and nuance, while greater numbers of categories gain richness but sacrifice parsimony. The trade-off between parsimony and extreme richness should be determined by considering the purposes of each individual study.

In a study of deterrence, for example, Alexander George and Richard Smoke found it to be inadequate and unproductive to define deterrence outcomes simply as “successes” or “failures.”²³ Instead, their explanations of individual cases of failure enabled them to identify different types of failures. This led to a typology of failures, with each type of failure having a different explanation. This typology allowed George and Smoke to see that deterrence failures exemplified the phenomenon of equifinality. The result was a more discriminating and policy-relevant explanatory theory for deterrence failures.²⁴

The differentiation of types can apply to the characterization of independent as well as dependent variables. In attempting to identify conditions associated with the success or failure of efforts to employ a strategy of coercive diplomacy, one set of investigators identified important variants of that strategy.²⁵ In their study, coercive diplomacy was treated as an independent variable. From an analysis of different cases, four types of the coercive diplomacy strategy were identified: the explicit ultimatum, the tacit ultimatum, the “gradual turning of the screw,” and the “try and see” variant. By differentiating the independent variable in this way, it was possible to develop a more discriminating analysis of the effectiveness of coercive diplomacy and to identify some of the factors that favored or handicapped the success of each variant. A very general or undifferentiated depiction of the independent variable would have

22. See also the discussion of this point in Chapter 9 on “The Congruence Method.”

23. George and Smoke, *Deterrence in American Foreign Policy*.

24. See the Appendix, “Studies That Illustrate Research Design,” for a fuller discussion of their study.

25. Alexander L. George, David K. Hall, and William E. Simons, *The Limits of Coercive Diplomacy* (Boston: Little, Brown, 1971); an extended second edition under the same title that examines additional cases was published in 1994, edited by Alexander

"washed out" the fact that variants of coercive diplomacy may have different impacts on outcomes, or it might have resulted in ambiguous or invalid results. In addition, the identification of different variants of coercive diplomacy strategy has important implications for the selection of cases.

Task Five: Formulation of Data Requirements and General Questions

The case study method will be more effective if the research design includes a specification of the data to be obtained from the case or cases under study. Data requirements should be determined by the theoretical framework and the research strategy to be used for achieving the study's research objectives. The specification of data requirements should be integrated with the other four design tasks. Specification of data requirements structures the study. It is an essential component of the method of structured, focused comparison.

Whether a single-case study or a case comparison is undertaken, specification of the data requirements should take the form of general questions to be asked of each case. This is a way of standardizing data requirements so that comparable data will be obtained from each case and so that a single-case study can be compared later with others. Case study methodology is no different in this respect from large-N statistical studies and public opinion surveys. Unless one asks the same questions of each case, the results cannot be compared, cumulated, and systematically analyzed.

This is only to say—and to insist—that case researchers should follow a procedure of systematic data compilation. The questions asked of each case must be of a general nature; they should not be couched in overly specific terms that are relevant to only one case but should be applicable to all cases within the class or subclass of events with which the study is concerned. Asking the same questions of each case does *not* prevent the case writer from addressing more specific aspects of the case or bringing out idiosyncratic features of each case that may also be of interest for theory development or future research.

A problem sometimes encountered in case study research is that data requirements are missing altogether or inadequately formulated. The general questions must reflect the theoretical framework employed, the data that will be needed to satisfy the research objective of the study, and the kind of contribution to theory that the researcher intends to make. In other words, a mechanical use of the method of structured, focused comparison will not yield good results.

integrated with the four other elements of the research design. For example, in a comparative study of policymakers' approaches to strategy and tactics toward political opponents in the international arena, one might start by asking questions designed to illuminate the orientations of a leader toward the fundamental issues of history and politics that presumably influence his or her processing of information, policy preference, and final choice of action.²⁶ In this type of study, the investigator examines an appropriate body of material in order to infer the "answers" a political leader might have given to the following questions:

PHILOSOPHICAL QUESTIONS

- What is the essential nature of political life? Is the political universe essentially one of harmony or conflict? What is the fundamental character of one's political opponents?
- What are the prospects for eventual realization of one's fundamental political values and ideological goals? Can one be optimistic or pessimistic?
- In what sense and to what extent is the political future predictable?
- How much control or mastery can one have over historical developments? What is the political leader's (or elite's) role in moving and shaping history?
- What is the role of chance in human affairs and in historical development?

INSTRUMENTAL QUESTIONS

- What is the best approach for selecting goals or objectives for political action?
- How are the goals of action pursued most effectively?
- How are the risks of political action best calculated, controlled, and accepted?
- What is the best timing of action to advance one's interests?

26. See 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. The problem of judging the causal role of such beliefs in a policymaker's choice of action was discussed in 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 Relations* (Boulder, Colo.: Westview Press, 1979), pp. 95–124. Since then, numerous studies have been made of the "operational codes" of a variety of leaders using this standardized approach. See 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.

- What is the utility and role of different means for advancing one's interests?

Integration of the Five Design Tasks

The five design tasks should be viewed as constituting an integrated whole. The researcher should keep in mind that these tasks are interrelated and interdependent. For example, the way in which Task Two is performed should be consistent with the specification of Task One. Similarly, both the selection of cases in Task Three and the theoretical framework developed in Task Four must be appropriate and serviceable from the standpoint of the determinations made for Tasks One and Two. And finally, the identification of data requirements in Task Five must be guided by the decisions made for Tasks One, Two, and Three.

Yet a satisfactory integration of the five tasks usually cannot be accomplished on the first try. A good design does not come easily. Considerable iteration and respecification of the various tasks may be necessary before a satisfactory research design is achieved. The researcher may need to gain familiarity with the phenomenon in question by undertaking a preliminary examination of a variety of cases before finalizing aspects of the design.

Despite the researcher's best efforts, the formulation of the design is likely to remain imperfect—and this may not be apparent until the investigator is well into phase two or even phase three of the study. If these defects are sufficiently serious, the researcher should consider halting further work and redesigning the study, even if this means that some of the case studies will have to be redone. In drawing conclusions from the study, the researcher (or others who evaluate it) may be able to gain some useful lessons for a better design of a new study of the problem.²⁷

27. For additional discussion of the critical importance of research design, see the "Pedagogical Note to Parts Two and Three."

Chapter 5

Phase Two: Carrying Out the Case Studies

The fifth task in a research design—the formulation of general questions to ask of each of the cases to be studied in phase two—allows the researcher to analyze each case in a way that will provide “answers” to the general questions.¹ These answers—the product of phase two—then constitute the data for the third phase of research, in which the investigator will use case findings to illuminate the research objectives of the study.

Usually one’s first step in studying a case with which one is not already intimately familiar is to gather the most easily accessible academic literature and interview data on the case and its context. This preliminary step of immersing oneself in the case, known as “soaking and poking,” often leads to the construction of a chronological narrative that helps both the researcher and subsequent readers understand the basic outlines of the case.²

1. This chapter draws on earlier publications by Alexander L. George, “Case Studies and Theory Development: The Method of Structured, Focused Comparison,” in Paul Gordon Lauren, ed., *Diplomacy: New Approaches in Theory, History, and Policy* (New York: Free Press, 1979), pp. 3–68; Alexander L. George, “The Causal Nexus Between Cognitive Beliefs and Decision-Making Behavior,” in Lawrence S. Falkowski, ed., *Psychological Models in International Politics* (Boulder, Colo.: Westview Press, 1979), pp. 95–124; and Alexander L. George and Timothy J. McKeown, “Case Studies and Theories of Organizational Decision Making,” in Robert F. Coulam and Richard A. Smith, eds., *Advances in Information Processing in Organizations*, Vol. 2 (Greenwich, Conn.: JAI Press, 1985), pp. 21–58.

2. An interesting example of “soaking and poking” and a description of how it mixes inductive and deductive reasoning is found in Richard F. Fenno’s *Homestyle*

After a period of “soaking and poking,” the researcher turns to the task of case study analysis, establishing the values of independent and dependent variables in a case through standard procedures of historical inquiry. (If appropriate, the researcher may be able to quantify and scale variables in some fashion.) The researcher should always articulate the criteria employed for “scoring” the variables so as to provide a basis for inter-coder reliability.

Next, the researcher develops explanations for the outcome of each case. This is a matter of detective work and historical analysis rather than a matter of applying an orthodox quasi-experimental design.³ Social scientists performing case studies will need to familiarize themselves with the craft of the historian’s trade—learning, for the context in which the case is embedded, the special difficulties presented by various kinds of evidence that may be available; using multiple weak inferences rather than single strong inferences to buttress conclusions; developing procedures for searching through large masses of data when the objectives of the search are not easily summarized by a few simple search rules.⁴

This chapter provides advice on these topics. The first three sections focus on the provisional nature of case explanations, and the challenges involved in weighing explanations offered by other researchers who have analyzed a given case, and the task of transforming a descriptive explanation for a case into an explanation that adequately reflects the researcher’s theoretical framework. We then turn to issues that researchers encounter when working with a variety of primary and secondary materials. Notable issues with secondary sources include the biases of their authors, and a tendency to overestimate the rationality of the policy-making process while underestimating the complexity and the multitude of interests that may be at play. Scholars face numerous issues in assessing the evidentiary value of primary sources. Finally, we describe some of the tasks faced by those who critically read others’ case studies, and urge that researchers make their methods as transparent as possible to the reader.

The Provisional Character of Case Explanations

Case explanations must always be considered to be of a provisional character. Therefore, the theoretical conclusions drawn from case study findings (in phase three) will also be provisional. The explanations pro-

3. For discussion of this point, see George and McKeown, “Case Studies and Theories of Organizational Decision Making,” pp. 38–39.

vided by the case writer may be challenged by other scholars on one or another ground—for example, the original research may have overlooked relevant data or misunderstood its significance, failed to consider an important rival hypothesis, and so forth. If case explanations are later successfully challenged, the researcher will have to reassess the implications for any theory that has been developed or tested. Such a reassessment would also be necessary if new historical data bearing on the cases become available at a later date and lead to a successful challenge of earlier explanations.

In seeking to formulate an explanation for the outcome in each case, the investigator employs the historian's method of causal imputation, which differs from the mode of causal inference in statistical-correlational studies. These causal interpretations gain plausibility if they are consistent with the available data and if they can be supported by relevant generalizations for which a measure of validity can be claimed on the basis of existing studies. The plausibility of an explanation is enhanced to the extent that alternative explanations are considered and found to be less consistent with the data, or less supportable by available generalizations.

An investigator must demonstrate that he or she has seriously considered alternative explanations for the case outcome in order to avoid providing the basis for a suspicion, justified or not, that he or she has "imposed" a favored theory or hypothesis as the explanation. Such a challenge is likely if the reader believes that case selection was biased by the investigator's commitment to a particular theory or hypothesis.⁵

The Problem of Competing Explanations

A familiar challenge that case study methods encounter is to reconcile, if possible, conflicting interpretations of a case or to choose between them. This problem can arise when the investigator provides an explanation that differs from an earlier scholar's but does not adequately demonstrate the superiority of the new interpretation. As Olav Njølstad notes, competing explanations may arise from several sources.⁶ There are different types of explanation stemming, for example, from historiographical is-

5. The need to avoid selecting cases that favor a particular theory and that constitute easy rather than tough tests of a theory was emphasized in Chapter 4.

6. This brief discussion draws from the fuller discussion of these problems in Chapter 2, "Case Study Methods and Research on the Interdemocratic Peace," which also provides illustrative materials. See also Olav Njølstad's chapter, "Learning from History? Case Studies and the Limits to Theory-Building," in Nils Petter Gleditsch and

sues such as the relative importance of ideology or historical context. Sometimes competing explanations can be equally consistent with the available historical evidence; this makes it difficult to decide which is the correct explanation or, alternatively, whether both interpretations may be part of the overall explanation—i.e., whether the outcome may be overdetermined. Another possibility is that each of the ostensibly competing explanations in fact addresses different parts of a complex longitudinal development. In such cases, the task of the investigator is to identify different turning points in the causal chain and to sort out which independent variables explain each step in the causal chain—for example, those explaining why a war occurred, those that explain the form of the attack, those that explain its timing, and so on. Still another possibility is that the key variable in one explanation is causal and the proposed causal variable in the other explanation is spurious.

The problem of apparently competing explanations may also arise when the rival interpretations address and attempt to explain different aspects of a case and therefore cannot be reconciled. When this happens, the investigator and readers of the case account should not regard the two interpretations as competing with each other. Another possibility is that the rival explanations emerge because the scholars advancing them have simply disagreed on the “facts” of the case.

In any case, if the data and generalizations available to the investigator do not permit him or her to choose from competing explanations, then both explanations for the case should be retained as equally plausible, and the implications of both for theory development should be considered in phase three of the study.

Transforming Descriptive Explanations Into Analytical Explanations

In addition to developing a specific explanation for each case, the researcher should consider transforming the specific explanation into the concepts and variables of the general theoretical framework specified in Task Two.⁷ (In Harry Eckstein’s terminology, such research is “disciplined-configurative” rather than “configurative-idiographic.”) To transform specific explanations into general theoretical terms, the researcher’s theoretical framework must be broad enough to capture the major elements of the historical context. That is, the set of independent and inter-

vening variables must be adequate to capture and record the essentials of a causal account of the outcome in the case. The dividing line between what is essential and what is not is whether aspects of a causal process in a given case are expected or found to operate across the entire class of cases under consideration. For example, if some instance of organizational decision-making was decisively affected by the fact that one of the key participants in the decision process caught a cold and was unable to attend an important meeting, this would *not* constitute a basis for revising our theory of organizational decision-making to endogenize the susceptibility of actors to disease. It *would*, however, constitute a basis for a general argument about how outcomes are affected by the presence or absence of important potential participants.

Some historians will object to this procedure for transforming a rich and detailed historical explanation into a more abstract and selective one couched in theoretical concepts, arguing that unique qualities of the explanation inevitably will be lost in the process. This is undoubtedly true: some loss of information and some simplification is inherent in any effort at theory formulation or in theoretically formulated explanations. The critical question, however, is whether the loss of information and the simplification jeopardize the validity of the conclusions drawn from the cases for the theory and the utility of that theory. This question cannot be answered abstractly. The transition from a specific to a more general explanation may indeed lead a researcher to dismiss some of the causal processes at work in the case simply because they are not already captured by the general theory or because the researcher fails to recognize a variable's general significance. To say that avoiding these errors depends on the sensitivity and judgment of the researcher, while true, is not very helpful. One slightly more specific guideline is that researchers seem more susceptible to this error when trying to discern new causal patterns than when attempting to evaluate claims about some causal patterns already hypothesized to be operating in a particular case; and second, that the more fine-tuned and concrete the description of variance, the more readily the analysis will accommodate a more differentiated description of the causal processes at work.⁸

To the extent that the case study method has arisen from the practice of historians, it has tended to follow certain procedures that are not really appropriate for social scientists. One feature of most historians' work is a relative lack of concern with or discussion of methodological issues en-

⁸ See Chapter 4 for a discussion of Task Four and the critical importance of how

countered in the performance of research. We believe that case researchers should explicitly discuss the major research dilemmas the case study researcher faced in the analysis of a case and the justifications for solving those dilemmas in a particular way. Therefore, we recommend that the investigator give some indication of how his or her initial expectations about behavior and initial data-collection rules were revised in the course of the study. This would permit readers to make a more informed analysis of the process by which a case and the conclusions based on the case were reached.

Most historians also rely heavily on chronological narrative as an organizing device for presenting the case study materials. Preserving some elements of the chronology of the case may be indispensable for supporting the theory-oriented analysis, and it may be highly desirable to do so in order to enable readers not already familiar with the history of the case to comprehend the analysis. Striking the right balance between a detailed historical description of the case and development of a theoretically-focused explanation of it is a familiar challenge. Analysts frequently feel it necessary to reduce the length of a case study to avoid overly long accounts that exceed the usual limits for journal articles or even books! The more cases, the more difficult this problem becomes.

There is no easy answer to this dilemma. Still, it has been dealt with in a reasonably effective way by a number of writers. A brief résumé of the case at the beginning of the analysis gives readers the essential facts about the development and outcome of the case. The ensuing write-up can blend additional historical detail with analysis.⁹ Presentation of a case need not always include a highly detailed or exclusively chronological narrative. As a theory becomes better developed and as research focuses on more tightly defined targets, there will be less need to present overly long narratives. Moreover, narrative accounts of a case can be supplemented by such devices as decision trees, sketches of the internal analytical structure of the explanation, or even computer programs to display the logic of the actors' decisions or the sequence of internal developments within the case.

Some Challenges in Attempting to Reconstruct Decisions

Scholars who attempt to reconstruct the policymaking process in order to explain important decisions face challenging problems. An important limitation of the analysis presented here is that it is drawn solely from the

9. See, for example, how this task was dealt with in studies such as Alexander I.

study of U.S. foreign policy.¹⁰ We discuss first the task of acquiring reliable data on factors that entered into the policy process and evaluating their impact on the decision. Political scientists must often rely upon, or at least make use of, historians' research on the policy in question. Such historical studies can be extremely useful to political scientists, but several cautions should be observed in making use of these studies.

First, researchers should forgo the temptation to rely on a single, seemingly authoritative study of the case at hand by a historian. Such a shortcut overlooks the fact that competent historians who have studied that case often disagree on how best to explain it. As Ian Lustick has argued, "the work of historians is not . . . an unproblematic background narrative from which theoretically neutral data can be elicited for the framing of problems and the testing of theories."¹¹ Lustick approvingly notes Norman Cantor's argument that a historian's work represents "a picture of 'what happened' that is just as much a function of his or her personal commitments, the contemporary political issues with which s/he was engaged, and the methodological choices governing his or her work."¹² The danger here, Lustick argues, is that a researcher who draws upon too narrow a set of historical accounts that emphasizes the variables of interest may overstate the performance of favored hypotheses.

It is thus necessary to identify and summarize important debates among historians about competing explanations of a case, and wherever possible to indicate the possible political and historical biases of the contending authors. The researcher should translate these debates into the competing hypotheses and their variables as outlined in phase one. If there are important historical interpretations of the case that do not easily translate into the hypotheses already specified, the researcher should consider whether these interpretations should be cast as additional hypotheses and specified in terms of theoretical variables. The same procedures apply to the primary political debates among participants in the case and their critics. Even such overtly political debates may draw upon

10. Similar problems arise in efforts by scholars to make use of archival materials and interviews from Soviet sources. See, for example, the correspondence between Mark Kramer, who expressed concern about the use of oral histories by Bruce J. Allyn, James G. Blight, and David A. Welch, and their responses in "Remembering the Cuban Missile Crisis: Should We Swallow Oral History?" *International Security*, Vol. 15, No. 1 (Summer 1990), pp. 212–218. See also "Commentaries on 'An Interview With Sergo Mikoyan'" by Raymond L. Garthoff, Barton J. Bernstein, Marc Trachtenberg, and Thomas G. Paterson in *Diplomatic History*, Vol. 14, No. 2 (Spring 1990), pp. 223–256.

11. Ian S. Lustick, "History, Historiography, and Political Science," *American Political*

generalizable variables that historians and researchers may have overlooked.

One way to avoid the risk of relying on a single historical analysis would be to follow the practice of Richard Smoke, who at the outset of his research, asked several historians to help him identify the best available accounts of each of the cases he planned to study. Later, Smoke obtained reviews of the first drafts of his cases from eight historians and made appropriate changes.¹³

Second, social scientists making use of even the best available historical studies of a case should not assume that they will provide answers to the questions they are asking. As emphasized in Chapter 3 on "The Method of Structured, Focused Comparison," the political scientist's research objectives determine the general questions to be asked of each case. The historian's research objectives and the questions addressed in his or her study may not adequately reflect those of subsequent researchers.¹⁴ We may recall that historians have often stated that if history is approached from a utilitarian perspective, then it has to be rewritten for each generation. History does not speak for itself to all successive generations. When new problems and interests are brought to a study of history by later generations, the meaning and significance of earlier historical events to the present may have to be studied anew and reevaluated. Hence, the study of relevant historical experience very much depends on the specific questions one asks of historical cases.

One of the key tasks during the "soaking and poking" process is to identify the gaps in existing historical accounts. These gaps may include archival or interview evidence that has not been examined or that had previously been unavailable. They may also include the measurement of variables the researcher identified in phase one that historians have not measured or have not measured as systematically as the explanatory goals of subsequent researchers require. It is also possible that researchers can make use of technologies, such as computer-assisted content analysis, that were not available to scholars writing earlier historical accounts.

Third, having identified possible gaps in existing accounts, the re-

13. See the preface to Richard Smoke, *War: Controlling Escalation* (Cambridge, Mass.: Harvard University Press, 1977).

14. The different ways historians and political scientists tend to define the task of explanation and the different questions they often ask of available data is discussed in helpful detail in Deborah Larson, "Sources and Methods in Cold War History: The Need for a Theory-Based Archival Approach," 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), pp. 327-350. The dangers of in-

searcher must reckon with the possibility that good answers to his or her questions about each case can be obtained only by going to original sources—archival materials, memoirs, oral histories, newspapers, and new interviews. In fact, political scientists studying international politics are increasingly undertaking this task. In doing so, however, they face the challenging task of weighing the evidentiary value of such primary sources.

Fourth, the researcher should not assume that going to primary sources and declassified government documents alone will be sufficient to find the answers to his or her research questions. The task of assessing the significance and evidentiary worth of such sources often requires a careful examination of contemporary public sources, such as daily media accounts of the developments of a case unfolding over time. Contemporary public accounts are certainly not a substitute for analysis of archival sources, but they often are an important part of contextual developments to which policymakers are sensitive, to which they are responding, or which they are attempting to influence. Classified accounts of the process of policymaking cannot be properly evaluated by scholars unless the public context in which policymakers operate is taken into account.¹⁵ We have at times found students who have become intimately familiar with hard-to-get primary source materials of a case but who have only a vague sense of the wider context because they have not taken the relatively easy (but often time-consuming) step of reading the newspapers or journals from the period.¹⁶

15. The importance of studying contemporary journalistic sources in order to understand part of the context in which policymakers were operating became a central methodological procedure in Deborah Larson's research. In conjunction with thorough research into archival sources, Larson spent a great deal of time going through contemporary journalists' accounts of developments, a procedure which helped her to appreciate the impact of events that came to the attention of policymakers on their perceptions and responses. Careful study of the public context of private deliberations was useful in evaluating the evidentiary significance of archival sources. See Deborah Welch Larson, *The Origins of Containment* (Princeton, N.J.: Princeton University Press, 1985). Larson amplifies and illustrates different ways in which contemporary newspaper accounts help the investigator to discern important elements of the context in which policymakers operate. See Larson, "Sources and Methods in Cold War History."

16. One example comes from the work of one of the present authors, Andrew Bennett. In an unpublished study of the 1929 stock market crash for the Federal Reserve Board, he found by reading the newspapers of the period that there are strong reasons to question the often-cited argument that the crash was caused by excessive speculation on margin credit rates "as low as 10 percent," or the supposedly common practice of buying stocks by putting up only 10 percent of their value as equity. In fact,

Finally, research on recent and contemporary U.S. foreign policy must be sensitive to the likelihood that important data may not be available and cannot be easily retrieved for research purposes, e.g., important discussions among policymakers that take place over the telephone or within internal e-mail and fax facilities—the results of which are not easily acquired by researchers.

The Risk of Over-Intellectualizing the Policy Process

When academic scholars attempt to reconstruct how and why important decisions were made, they tend to assume an orderly and more rational policymaking process than is justified. For example, overly complex and precise formal models may posit decision-making heuristics that are “too clever by half,” or that no individual would actually utilize. Also, scholars sometimes succumb to the common cognitive bias toward univariate explanations—explanations in which there appears to be a single clear and dominating reason for the decision in question. Instead, analysts should be sensitive to the possibility that several considerations motivated the decision.

In fact, presidents and top-level executives often seek multiple payoffs from any decision they take. Leaders known for their sophistication and skill, such as Lyndon B. Johnson, use this strategy to optimize political gains from a particular decision. Disagreements among scholars as to the particular reason for why a certain action was taken often fail to take this factor into account.

Several considerations can enter into a decision in other ways as well. Particularly in a pluralistic political system in which a number of actors participate in policymaking, agreement on what should be done can emerge for different reasons. It is sufficient that members of the policymaking group agree only on *what* to do without having to agree on *why* to do it. In some situations, in fact, there may be a tacit agreement among members of the group that not all those who support the decision have to share the same reason or a single reason for doing so. To obtain sufficient

on the assumption that the crash was caused by a liquidity crisis as plunging stock values led to margin calls on stocks and forced sales of those stocks. The fact that this measure failed to stem the crash, and that bond purchases were strong during the crash, suggest that perhaps the crash was caused not so much by loose margin credit as by the classic bursting of a speculative bubble, and a revaluation of the relative value of stocks versus bonds. This explanation is more in line with modern theories of stock market behavior. In any event, a simple reading of the newspapers reveals that explanations of the crash cannot unproblematically accept that margins were typically 10 percent

consensus on a decision may be difficult for various reasons, and sufficient time and resources may not be available for achieving a completely shared judgment in support of the decision. In any action-oriented group, particularly one that operates under time pressure, it is often enough to agree on what needs to be done. It may not be feasible or wise to debate until everyone agrees not merely on what decision to take but also the precise reasons for doing so.

Assessing the Evidentiary Value of Archival Materials

Scholars doing historical case studies must find ways of assessing the evidentiary value of archival materials that were generated during the policymaking process under examination. Similarly, case analysts making use of historical studies produced by other scholars cannot automatically assume that these investigators properly weighed the evidentiary significance of documents and interviews.

Scholars are not immune from the general tendency to attach particular significance to an item that supports their pre-existing or favored interpretation and, conversely, to downplay the significance of an item that challenges it. As cognitive dissonance theory reminds us, most people operate with a double standard in weighing evidence. They more readily accept new information that is consistent with an existing mind-set and employ a much higher threshold for giving serious consideration to discrepant information that challenges existing policies or preferences.

All good historians, it has been said, are revisionist historians. That is, historians must be prepared to revise existing interpretations when new evidence and compelling new interpretations emerge. Even seemingly definitive explanations are subject to revision. But new information about a case must be properly evaluated, and this task is jeopardized when a scholar is overly impressed with and overinterprets the significance of a new item—e.g., a recently declassified document—that emerges on a controversial or highly politicized subject.

Analytical or political bias on the scholar's part can lead to distorted interpretation of archival materials. But questionable interpretations can also arise when the analyst fails to grasp the context of specific archival materials. The importance of context in making such interpretations deserves more detailed analysis than can be provided here, so a few observations will have to suffice.

It is useful to regard archival documents as a type of purposeful communication. A useful framework exists for assessing the meaning and evidentiary worth of *what* is communicated in a document, speech, or inter-

analyst should consider *who* is speaking to *whom*, for *what purpose* and *under what circumstances*.¹⁷ The evidentiary worth of what is contained in a document often cannot be reliably determined without addressing these questions. As this framework emphasizes, it is useful to ask what purpose(s) the document was designed to serve. How did it fit into the policymaking process? What was its relation to the stream of other communications and activities—past, present, and future?

It is also important to note the circumstances surrounding the document's release to the public, and to be sensitive to the possibility that documents will be selectively released to fit the political and personal goals of those officials who control their release. Much of the internal documentation on Soviet decision-making on the invasion and occupation of Afghanistan beginning in 1979, for example, was released by the government of Russian President Boris Yeltsin in the mid-1990s to embarrass the Soviet Communist Party, which was then on trial for its role in the 1991 Soviet coup attempt. Needless to say, the Yeltsin government did not release any comparable documents on its own ill-fated intervention in Chechnya in the mid-1990s.

In studying the outputs of a complex policymaking system, the investigator is well advised to work with a sophisticated model or set of assumptions regarding ways in which different policies are made in that system. For example, which actors and agencies are the most influential in a particular issue area? To whom does the leader turn for critical information and advice on a given type of policy problem? How do status differences and power variables affect the behavior of different advisers and participants in high-level policymaking?

Thus, it is advisable to observe a number of cautions in following the "paper trail" leading to a policy decision. Has a country's leader tipped his or her hand—at least in the judgment of participants in the pro-

17. This framework was initially developed and employed in a study that examined methods for inferring the intentions, beliefs, and other characteristics of a political elite from its propaganda by means of qualitative content analysis. See Alexander L. George, *Propaganda Analysis: A Study of Inferences Made from Nazi Propaganda in World War II* (Evanston, Ill.: Row, Peterson, 1959; and Westport, Conn.: Greenwood Press, 1973), pp. 107–121.

In a personal communication (March 26, 2000), Jeremi Suri drew on his own research experience to emphasize the need to distinguish between various types of archival materials. Personal correspondence and diaries of historical actors can be very helpful in developing understanding of their general beliefs about political life, particularly since such materials are often not designed to persuade others; such sources can reflect the emotions experienced at different junctures. Also, the "incoming files" of

cess—regarding what he or she will eventually decide? What effect does such a perception—or misperception—have on the views expressed or written by advisers? Are some of the influential policymakers bargaining with each other behind the leader's back regarding what advice and options to recommend in the hope and expectation that they can resolve their differences and protect their own interests?¹⁸ What role did policymakers play in writing their own public speeches and reports, and to what extent do specific rhetorical formulations represent these top officials' own words rather than those of speech writers and other advisers?

It is well known that those who produce classified policy papers and accounts of decisions often wish to leave behind a self-serving historical record. One scholar who recently spent a year stationed in an office dealing with national security affairs witnessed occasions on which the written, classified record of important decisions taken was deliberately distorted for this and other reasons.¹⁹ Diplomatic historian Stephen Pelz reminds us that "many international leaders take pains to disguise their reasoning and purposes, and therefore much of the best work on such figures as Franklin D. Roosevelt consists of reconstructing their assumptions, goals, and images of the world from a variety of sources."²⁰

In assessing the significance of "evidence" that a leader has engaged in "consultation" with advisers, one needs to keep in mind that he or she may do so for several different reasons.²¹ We tend to assume that he or she consults in order to obtain information and advice before making a final decision—i.e., to satisfy his or her "cognitive needs." But he or she may consult for any one or several other reasons. The leader may want to obtain emotional support for a difficult, stressful decision; or the leader may wish to give important advisers the feeling they have had an opportunity to contribute to the decision-making process so that they will be more likely to support whatever decision the president makes—i.e., to build consensus; or the leader may need to satisfy the expectation (generated by the nature of the political system and its political culture and

18. Some of these possibilities are among the various "malfunctions" of the policymaking system discussed in Alexander L. George, *Presidential Decisionmaking and Foreign Policy: The Effective Use of Information and Advice* (Boulder, Colo.: Westview Press, 1980), chap. 6.

19. This observation was provided by a scholar who must remain anonymous.

20. Stephen Pelz, "Toward A New Diplomatic History: Two and a Half Cheers for International Relations Methods," in Elman and Elman, eds., *Bridges and Boundaries*, p. 100.

norms) that important decisions will not be made without the participation of all key actors who have some relevant knowledge, expertise, or responsibility with regard to the matter being decided; that is, the president hopes to achieve “legitimacy” for a decision by giving evidence that assures Congress and the public that it was well-considered and properly made. (Of course, a leader’s consultation in any particular instance may combine several of these purposes.)

This last purpose—consultation—is of particular interest in the United States. The public wants to be assured that an orderly, rational process was followed in making important decisions. Consider the development in recent decades of “instant histories” of many important decisions by leading journalists on the basis of their interviews with policymakers shortly after the event. Knowing that the interested public demands to know how an important decision was made, top-level policymakers are motivated to conduct the decision process in ways that will enable them to assure the public later that the decision was made after careful multisided deliberation. Information to this effect is given to journalists soon after the decision is made. Since “instant histories” may be slanted to portray a careful, multidimensioned process of policymaking, the case analyst must consider to what extent such an impression is justified and how it bears on the evidentiary worth of the information conveyed in the instant history and in subsequent “insider” accounts of how and why a particular decision was made.

To weigh archival type material effectively, scholars need to be aware of these complexities. An excellent example of a study that captures the dynamics of decision-making is Larry Berman’s interpretation of President Johnson’s decision in July 1965 to put large-scale ground combat troops into Vietnam. Some archival sources suggest that Johnson employed a careful, conscientious version of “multiple advocacy” in which he thoughtfully solicited all views. But according to Berman’s analysis, Johnson had already decided what he had to do and went through the motions of consultation for purposes of consensus-building and legitimization of his decision.²²

In another example, many scholars assumed that President Dwight D. Eisenhower’s policymaking system was highly formalistic and bureaucratic, a perception shared by important congressional and other critics at the time. Working with this image of Eisenhower’s decision-making style, scholars could easily misinterpret the significance of archival sources generated by the *formal* track of his policymaking. Easily

22. Larry Berman, *Planning a Tragedy: The Americanization of the War in Vietnam* (New

overlooked was the *informal* track, which preceded and accompanied the formal procedures, awareness of which led Fred Greenstein to write about the “hidden hand style” by which Eisenhower operated.²³ Now, a more sophisticated way of studying Eisenhower’s policymaking has developed that pays attention to both the formal and informal policy tracks and to the interaction between them.

The relevance and usefulness of working with an analytical framework that considers both tracks is, of course, not confined to studying the Eisenhower presidency. The workings of the informal track are not likely to become the subject of a written archival document. It is important to use interviews, memoirs, the media, etc., to obtain this valuable material.

Another aspect of the importance of a contextual framework for assessing the evidentiary worth of archival sources has to do with the hierarchical nature of the policymaking system in most governments. We find useful the analogy of a pyramid of several layers. Each layer, beginning with the bottom one, sends communications upwards (as well as sideways), analyzing available data on a problem and offering interpretations of its significance for policy. As one moves up the pyramid, the number of actors and participants grows smaller but their importance (potential, if not actual) increases. As one reaches the layer next to the top—the top being the president—one encounters a handful of key officials and top advisers. At the same time, we find that researchers at times interview officials who are too high in the hierarchy to have had close involvement in or detailed recall of the events under study. Often, lower-level officials who worked on an issue every day have stronger recollections of how it was decided than the top officials who actually made the decision but who focused on the issues in question only intermittently. However, a researcher must take into account that even well-informed lower-level officials often do not have a complete or fully reliable picture of how and why a decision was made—i.e., the “Rashomon” problem, when different participants in the process have different views as to what took place.

This layered pyramid produces an enormous number of communications and documents that the scholar must assess. The possibility of erroneous interpretation of the significance of archival material is enormous. How do sophisticated historians and other scholars cope with this problem? What cautions are necessary when examining archival sources on top-level policymaking? How does a researcher deal with the fact that much of the material coming to the top-level group of policymakers from

23. Fred I. Greenstein, *The Hidden-Hand Presidency: Eisenhower As Leader* (New York:

below is inconsequential? How does one decide which material coming from below to the top-level officials made a difference in the decision? How can one tell why he or she really decided as he or she did as against the justifications given for his or her decisions?

The analyst's search for documentary evidence on reasons behind top-level decisions can also run into the problem that the paper trail may end before final decisions are made. Among the reasons for the absence of reliable documentary sources on such decisions is the role that secrecy can play. Dean Rusk, Secretary of State during the Kennedy administration, later stated that secrecy "made it very difficult for many to reconstruct the Bay of Pigs operation, particularly its planning, because very little was put on paper. [Allen] Dulles, [Richard] Bissell, and others proposing the operation briefed us orally."²⁴

No doubt there are important examples of scholarly disputes that illustrate these problems and indicate how individual analysts handled them. What general lessons can be drawn that would help train students and analysts? We have not yet found any book or major article that provides an adequate discussion of the problems of weighing the evidentiary worth of archival materials.²⁵ The most we can do, therefore, is to warn writers of historical case studies about some of these problems and to call attention to some of the methods historians and political scientists have employed in dealing with archival materials. Deborah Larson, for example, suggests that "to judge the influence of a memo written by a

24. As told to Richard Rusk in Daniel S. Papp, ed., *As I Saw It* (New York: Norton, 1990), cited by Richard Ned Lebow, "Social Science and History: Ranchers versus Farmers," in Elman and Elman, eds., *Bridges and Boundaries*, p. 132.

25. The most useful account we have found is the article by John D. Mulligan, "The Treatment of A Historical Source," *History and Theory*, Vol. 18, No. 2 (May 1979), pp. 177–196. Mulligan identifies various criteria historians employ for evaluating the authenticity, meaning, and significance of historical sources. He cites the observations on these issues made by a large number of distinguished historians and illustrates how each criterion applies to his own research, which focused on the importance of a correct evaluation of a primary source which sharply challenges accepted historical research on an aspect of the Civil War. This source was a personal letter, not a governmental document. Nonetheless, Mulligan's article illustrates the relevance of the framework we suggest, namely asking, "who says what to whom for what purpose in what circumstances?"

Also useful is the recent article by Cameron G. Thies, "A Pragmatic Guide to Qualitative Historical Analysis in the Study of International Relations," *International Studies Perspective*, Vol. 3, No. 4 (November 2002), pp. 351–372. This article includes a comprehensive list of sources that contributed to his essay. Readers may also want to consult the website "History Matters" <www.historymatters.gmu.edu> which is designed for high school and college teachers of history. This website includes sections on

lower-level official, one can look to see who initialed it. Of course, that a secretary of state initialed a memo does not prove that he read it, but it is a first step in analysis. Sometimes higher officials will make marginal comments—these can be quite important. Finally, paragraphs from memos written by lower officials sometimes appear in National Security Council policy memoranda."²⁶

Problems in Evaluating Case Studies

Case writers should become familiar with the variety of critiques their work may face. The importance of understanding the history and context of a case makes the difficulties of critiquing qualitative research different from those of assessing quantitative work. Readers cannot easily judge the validity of the explanation of a case unless they possess a degree of independent knowledge of that case. This requires that reader-critics themselves possess some familiarity with the complexity of the case and the range of data available for studying it; knowledge of the existence of different interpretations offered by other scholars and of the status of the generalizations and theories employed by the case writer; and an ability to evaluate the case writer's use of counterfactual analysis or to provide plausible counterfactual analysis of their own. These are tough requirements for readers who must evaluate case studies, and simply to state these desiderata suffices to indicate that they are not easily met. Our own commentaries of case study research designs in the Appendix, "Studies That Illustrate Research Design," should be read with the caveat that we are not theoretical or historical experts on all the subjects of these studies. This is a problem also for those who review these books in academic journals.

Let us discuss some of the problems likely to be encountered by readers who attempt to evaluate case studies. Much of the preceding discussion is relevant to the task of evaluating case studies, and a few additional observations can be made.

The task of evaluating case studies differs depending on the research objective of the case. When the investigator's research objective is to explain a case outcome, the reader-critic must consider whether the case analyst has "imposed" a favored theory as the explanation. Have alternative theories that might provide an explanation been overlooked or inadequately considered? When the case writer pursues the different research objective of attempting to use case findings to "test" an existing

theory, there are several questions the reader-critic has to consider in deciding whether such a claim is justified. Does the case (or cases) constitute an easy or tough test of the theory? Do case findings really support the theory in question? Do they perhaps also support other theories the investigator has overlooked or inadequately considered?

Reader-critics must consider the possibility that the case-writer has overlooked or unduly minimized potentially important causal variables, or has not considered the possibility or likelihood that the phenomenon is subject to multiple conjunctural causation or is affected by equifinality.

These and other problems in using case studies to develop or test theories are also discussed in Chapter 6. They are referred to here in order to emphasize that case writers should be familiar with the variety of criticisms that can be and often are made of their work.

In addition, we urge that case writers accept the obligation to assist readers in evaluating whether their case analyses have met relevant methodological standards. To meet this requirement, case writers should go as far as reasonably possible to make the analyses they offer transparent enough to enable readers to evaluate them. Transparency of case studies must be closely linked with standards for case studies. These standards include (but are not limited to) providing enough detail to satisfy as much as possible the criteria of replicability and of the validity and reliability of the way in which variables are scored. Certainly these standards are often difficult to meet in case study research, but case writers can often do more to at least approximate them. We strongly concur with the admonition of Gary King, Robert Keohane, and Sidney Verba that *"the most important rule for all data collection is to report how the data were created and how we came to possess them."*²⁷

In sum, case analysts should strive to develop and make use of appropriate rules for qualitative analysis. As argued in earlier chapters, however, the development of such guidelines should not be regarded as a matter of simply extending to qualitative analysis all of the standard conventions for quantitative analysis. Some of these conventions apply also to qualitative analysis, but guidelines for case studies must take into account the special characteristics of qualitative methodology.²⁸

27. Gary King, Robert O. Keohane, and Sidney Verba, *Designing Social Inquiry: Scientific Inference in Qualitative Research* (Princeton, N.J.: Princeton University Press, 1994), p. 51. Emphasis in original.

28. For a detailed analysis of this position, see Gerardo L. Munck, "Canons of Re-

Conclusion

The present book was in process of publication when we became aware of a new guidebook on how to make use of primary historical sources. The author, Marc Trachtenberg, has produced a superb manuscript which is in draft form for the time being. Its title is *Historical Method in the Study of International Relations*.

Himself a leading diplomatic historian, Trachtenberg joined the political science department at UCLA several years ago. He has succeeded in bringing together historical and political science approaches to the study of international relations. This book will be an invaluable source for students and professors who want to integrate the perspectives of history and political science for insightful research on foreign policies.

We will not attempt to summarize the rich materials he presents. The titles of several chapters may be noted: Chapter 3, "The Critical Analysis of Historical Texts"; and Chapter 5, "Working with Documents." A chapter is also provided on "Diplomatic History and International Relations Theory"; another chapter provides a detailed analysis of America's road to war in 1941.

Trachtenberg's treatment of these issues is unusually user-friendly. It is written in an engaging style. It will become standard text for research on foreign policy. Trachtenberg provides many incisive examples to illustrate his points.

We may also recall the statement that Trachtenberg made some time ago: "The basic methodological advice one can give is quite simple: documents are not necessarily to be taken at face value, and one has to see things in context to understand what they mean. One has to get into the habit of asking why a particular document was written—that is, what purpose it was meant to serve."²⁹

We have stressed in the preceding pages the necessity to regard archival sources as being instances of purposive communication. This advice is strongly reinforced by Deborah Larson on the basis of her experience in conducting in-depth research in archival sources in preparing her book *Origins of Containment*.³⁰ A recent article by Larson helps to fill the gap regarding the proper use of archival sources, at least for research on U.S.

ment of the canons for qualitative research imbedded in King, Keohane, and Verba, *Designing Social Inquiry*.

29. In a letter to Alexander L. George (January 29, 1998), Marc Trachtenberg indicated that he is currently studying methods for assessing archival and other sources in research on international politics.

foreign policy. In it she emphasizes that it is important to understand the purpose of a document and the events leading up to it in order to correctly interpret its meaning. . . . The author of a memorandum or speaker at a meeting may be trying to ingratiate himself with superiors, create a favorable impression of himself, put himself on the record in case of leaks, or persuade others to adopt his preferred policy. Whatever his goals, we cannot directly infer the communicator's state of mind from his arguments without considering his immediate aims.³¹

Larson also notes that study of contemporary accounts in leading newspapers sometimes can be essential for ascertaining the context of documents. "News accounts can help to establish the atmosphere of the times, the purpose of speeches or statements, or the public reaction to a statement. Newspapers help to show what information policymakers had and provide clues as to what events they regarded as important. . . . In this way, newspapers help us to recapture the perspective of officials at the time."³²

31. Deborah Welch Larson, "Sources and Methods in Cold War History," pp. 327-350.

32. Ibid. See also the project "Oral History Roundtables: The National Security Project," established in 1998 by Ivo H. Daalder and I.M. Destler, sponsored by the Brookings Institution and the Center for International and Security Studies at the University of Maryland. This series of roundtables, published periodically, brings together former officials specializing in foreign and security affairs to discuss specific historical problems in which they were involved. Daalder and Destler plan a final summary report.