

find.<sup>65</sup> A better solution is to give bonus credit to scholars who take on the harder task of studying the less observable.

6. "Counterfactual analysis can expand the number of observations available for theory-testing." James Fearon suggests this argument.<sup>66</sup> Counterfactual statements cannot provide a substitute for empirical observations, however. They can clarify an explanation: "I claim  $x$  caused  $y$ ; to clarify my claim, let me explain my image of a world absent  $x$ ." They can also help analysts surface hypotheses buried in their own minds (see the section "How Can Theories Be Made?" in this chapter). But counterfactual statements are not data and cannot replace empirical data in theory-testing.

65. Moreover, tests that are difficult for the time being may become feasible as new tests are devised or new data emerge. Thus theories of the Kremlin's conduct under Stalin were hard to test before the Soviet collapse but later became more testable. This is another reason to keep hard questions on the agenda.

66. James D. Fearon, "Counterfactuals and Hypothesis Testing in Political Science," *World Politics* 43 (January 1991): 171 and passim.

## C H A P T E R 2

### What Are Case Studies? How Should They Be Performed?

A large literature on the case-study method has appeared in recent years,<sup>1</sup> but that literature remains spotty. No complete catalog of research designs for case studies has emerged.<sup>2</sup> No textbook covers the gamut of study design considerations.<sup>3</sup> There

1. A good survey of the case-study literature is David Collier, "The Comparative Method," in Ada W. Finifter, ed., *Political Science: The State of the Discipline*, 2d ed. (Washington, D.C.: American Political Science Association, 1993), pp. 105–20. Landmark writings on the case method include Alexander L. George and Timothy J. McKeown, "Case Studies and Theories of Organizational Decision Making," in *Advances in Information Processing in Organizations* (Greenwich, Conn.: JAI Press, 1985), 2:21–58; Arend Lijphart, "Comparative Politics and the Comparative Method," *American Political Science Review* 65 (September 1971): 682–93; Harry Eckstein, "Case Study and Theory in Political Science," in Fred I. Greenstein and Nelson W. Polsby, ed., *Handbook of Political Science*, vol. 7, *Strategies of Inquiry* (Reading, Mass.: Addison-Wesley, 1975), pp. 79–137; and Robert K. Yin, *Case Study Research: Design and Methods*, 2d ed. (Thousand Oaks, Calif.: Sage, 1994). A more developed discussion by George is Alexander L. George, "Case Studies and Theory Development" (paper presented to the Second Annual Symposium on Information Processing in Organizations, Carnegie-Mellon University, Pittsburgh, Pa., October 15–16, 1982). An earlier statement is Alexander 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. Additional works on case methods are listed in the bibliography.

2. Yin, *Case Study Research*, pp. 18–19.

3. *Ibid.*, p. 18. Useful steps toward such a text are Yin's *Case Study Research* and

is no soup-to-nuts cookbook on the case method for beginning practitioners, and many texts on social science methodology slight or omit the case-study method.<sup>4</sup> Accordingly, the following chapter distills, elaborates, and qualifies the observations and suggestions of existing literature. I focus on assessing the case-study method and offering practical how-to-do-it advice for beginners doing case studies.

### Case Studies in Perspective

As I noted in Chapter 1, we have two basic ways to test theories: experimentation and observation.<sup>5</sup> Observational tests come in two varieties: large-*n* and case study. Thus, overall we have a universe of three basic testing methods: experimentation, observation using large-*n* analysis, and observation using case-study analysis.

Which testing method is best? Is case study inferior to other methods?

Social scientists have long considered case studies the weakest

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

4. Yin, *Case Study Research*, pp. 13, 18–19; Jennifer Platt, "'Case Study' in American Methodological Thought," *Current Sociology* 40 (Spring 1992): 42–43. Works slighting the case-study method include Earl Babbie, *The Practice of Social Research*, 7th ed. (Belmont, Calif.: Wadsworth, 1995); Julian L. Simon and Paul Burstein, *Basic Research Methods in Social Science*, 3d ed. (New York: Random House, 1985); Kenneth D. Bailey, *Methods of Social Research*, 4th ed. (New York: Free Press, 1994); David Dooley, *Social Research Methods*, 3d ed. (Upper Saddle River, N.J.: Prentice-Hall, 1995); and Norman K. Denzin, *The Research Act*, 3d ed. (Englewood Cliffs, N.J.: Prentice-Hall, 1989). Babbie mentions case studies once (p. 280); Simon and Burstein give case-study methods a two-page mention (pp. 37–38); Bailey has three pages that touch the subject (pp. 301–3); Dooley has a chapter on "qualitative research" (pp. 257–74) but no direct mention of case studies.

5. See the section "How Can Theories Be Tested?" in Chapter 1.

of these three testing methods for two reasons.<sup>6</sup> First, some argue that case studies provide the least opportunity to control for the effect of perturbing third variables. In this view experiments are the best method (the investigator eliminates the possible effect of omitted variables by exposing the group to only one stimulus, while holding the others constant). Large-*n* analysis is next-best, because the investigator can run partial correlations to control the effect of specific omitted variables and can rely on the randomizing effect of examining many cases to reduce the effects of other omitted variables. Studies of one or a few cases are worst, because the data is unrandomized and partial correlations are infeasible, since data points are too few.<sup>7</sup>

This criticism of case studies is unfair, however. Case studies offer two fairly strong methods for controlling the impact of omitted variables. First, tests of predictions of within-case variance

6. See, e.g., Yin, *Case Study Research*, pp. 9–11, who notes the "traditional prejudice against the case study strategy," and the "disdain for the [case-study] strategy" held by many researchers (p. 9). As Yin further notes, texts on social science methodology reflect this disdain by neglecting or omitting the case method: "Most social science textbooks have failed to consider the case method a formal research strategy at all" (p. 13). Randy Stoecker likewise notes the disrepute of case studies among sociologists, who "see the case study as barely better than journalism." Randy Stoecker, "Evaluating and Rethinking the Case Study," *The Sociological Review* 39 (February, 1991): 88. See also Jacques Hamel with Stéphane Dufour and Dominic Fortin, *Case Study Methods* (Newbury Park, Calif.: Sage, 1993), pp. 18–28.

7. Lijphart and Smelser advance this view. See Lijphart, "Comparative Politics and the Comparative Method," pp. 683–84, and Neil J. Smelser, "The Methodology of Comparative Analysis," in Donald P. Warwick and Samuel Osherson, eds., *Comparative Research Methods* (Englewood Cliffs, N.J.: Prentice Hall, 1973), pp. 45–57. Collier, "Comparative Method," pp. 106–8, summarizes Lijphart. See also Donald T. Campbell and Julian C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Boston: Houghton Mifflin, 1963), p. 6, who claim that single case studies are "of almost no scientific value." But see further Campbell's later retraction: Donald T. Campbell, "'Degrees of Freedom' and the Case Study," in Donald T. Campbell, *Methodology and Epistemology for Social Science: Selected Papers* (Chicago: University of Chicago Press, 1988, first pub. 1974), pp. 377–88; and noting this retraction, Collier, "Comparative Method," p. 115.

(that is, tests using a multiple “congruence procedure”<sup>8</sup> or a “process-tracing” methodology<sup>9</sup>) gain strong controls from the uniform character of the background conditions of the case.<sup>10</sup> Most cases offer a backdrop of fairly uniform case conditions, and many cases allow a number of observations of values on the independent (IV) and dependent variables (DV). If case conditions are uniform, we can discount third-variable influence as a cause of observed within-case covariance between values on IV and DV. (The uniform background conditions of the case create a semi-controlled environment that limits the effects of third variables by holding them constant.)<sup>11</sup>

Second, we can control the effects of omitted variables by selecting for study cases with extreme (high or low) values on the study variable (SV). This lowers the number of third factors with the strength to produce the result that the test theory predicts,

8. In a multiple-congruence procedure the investigator explores the case looking for congruence or incongruence between observed and predicted values on several or more measures of the independent and dependent variables of the test hypothesis. See the discussion of congruence procedure in the next section of this chapter. To test a theory fully one would look for congruence or incongruence between values of independent and dependent variables, between independent and intervening variables, between intervening variables (if there are several), and between intervening and dependent variables.

9. On “process tracing” see George and McKeown, “Case Studies and Theories,” pp. 34–41; George, “Case Studies and Theory Development” (1979), pp. 18–19; and the discussion of process tracing in the next section of this chapter. George and McKeown use “process tracing” to refer to a tracing of “the decision process by which various initial conditions are translated into outcomes.” “Case Studies and Theories,” p. 35. I use the term more broadly, to refer to the tracing of *any* causal process by which initial conditions are translated into outcomes. Thus my definition includes the tracing of both decision processes and causal processes that do not involve decisions. We might reserve “decision-process tracing” to capture George and McKeown’s narrower meaning.

10. Noting the controls that congruence procedure and process tracing allow (and referring to them jointly as “pattern matching”) is Campbell, “Degrees of Freedom’ and the Case Study,” p. 380.

11. This logic applies to analysis of any hypothesized causal relationship—between IV and IntV, IntV and IntV, or IntV and DV, as well as IV and DV.

which lowers the possibility that omitted variables account for passed tests.<sup>12</sup>

A second criticism of case studies—that “case-study results cannot be generalized to other cases”—has more merit, but applies only to single-case studies. A single case is a poor laboratory for identifying a theory’s antecedent conditions (background conditions that activate or magnify its action),<sup>13</sup> because as noted above, most cases provide a backdrop of fairly uniform case conditions. This uniformity masks the action of antecedent conditions that the theory requires, since the antecedent condition does not vary, hence it causes no telltale variance on the DV. Thus a theory that passes a single-case-study test with flying colors may require rare antecedent conditions and hence have little explanatory range,<sup>14</sup> but this weakness can remain hidden from an investigator who studies only one or two cases. (Thus a strength of the case method is also a weakness. The uniformity of case background conditions controls the effects of third variables but also masks antecedent conditions.) The identity and importance of antecedent conditions emerges more clearly from large-*n* studies. In large-*n* studies, cases that lack these antecedent conditions emerge as outliers that exhibit the theory’s cause without its predicted outcome. The existence of outliers signals that the theory needs

12. On this technique see the discussion of congruence procedure type 1 in the next section of this chapter. A third means of omitted-variable control in case studies is found in the method of controlled comparison, using John Stuart Mill’s “method of difference,” but this is a fairly weak tool. See the discussion of controlled comparison in the next section of this chapter.

13. The process of defining and measuring the prevalence of the antecedent conditions is often referred to as testing a theory’s “external validity,” meaning tests “establishing the domain to which a theory can be generalized.” Yin, *Case Study Research*, p. 33. Tests for external validity contrast with tests of “internal validity,” which address the capacity of the theory to explain a given case. See, for example, Yin, *Case Study Research*, pp. 33, 35–36, and Collier, “Comparative Method,” p. 113. I avoid these binary categories because they omit an important third type of validity—the ability of the theory to pass tests in a given case.

14. On explanatory range see the discussion of a theory’s explanatory power in the section “What Is a Good Theory?” in Chapter 1.

special conditions to operate; study of these outliers can identify these conditions. The study of a single case offers no parallel method for uncovering antecedent conditions; however, these antecedent conditions can be uncovered by doing more case studies, so this weakness in the case method is reparable.<sup>15</sup>

The case method has two strengths that offset this weakness. First, tests performed with case studies are often strong, because the predictions tested are quite unique (these predictions are not made by other known theories).<sup>16</sup> Specifically, case studies allow the test of predictions about the private speech and writings of policy actors. Often these predictions are singular to the theory that makes them; no other theory predicts the same thoughts or statements. The confirmation of such predictions strongly corroborates the test theory. Case studies are the best format for capturing such evidence. Hence case studies can supply quite decisive evidence for or against political theories. Often this evidence is more decisive than large-*n* evidence.

Second, inferring and testing explanations that define *how* the independent causes the dependent variable are often easier with case-study than large-*n* methods. If case-study evidence supports a hypothesis, the investigator can then explore the case further to deduce and test explanations detailing the operation of the hypothesis. Most important, one can “process trace,” that is, examine the process whereby initial case conditions are translated into case outcomes. How does the theory work? Tracing process can tell us. Congruence procedures can also illuminate explanations. (More on process tracing and congruence procedures below.) Both procedures are fairly easy to perform after a case has been initially

15. Methods of inferring and testing antecedent conditions with case studies are discussed later in this chapter.

16. As noted in Chapter 1, a test is strong if it evaluates a unique prediction (a forecast not made by other known theories), because the prediction's fulfillment cannot be explained except by the theory's action. Tests are also strong if they evaluate certain predictions (forecasts that are unequivocal and must occur if the theory is valid). On strong and weak tests see the section “Strong vs. Weak Tests” in Chapter 1 and the section “Strong vs. Weak Tests” in this chapter.

studied because the background spadework on the case—establishing the case background and chronology, and so on—has already been done. In contrast, a large-*n* test of a hypothesis provides little or no new insight into the causal process that comprises the hypothesis' explanation, nor does it generate data that could be used to infer or test explanations of that process. Overall, large-*n* methods tell us more about whether hypotheses hold than why they hold. Case studies say more about why they hold.

Thus the case method is a strong method of testing theories. Is a theory valid? How does it operate? Even single-case studies can give clear answers. They are less able to identify a theory's antecedent conditions. How broad is the range of cases that the theory governs? Case studies say little unless several are performed.

Which method of inquiry—experiment, large-*n*, or case study—is superior? The answer turns on the nature of our question and structure of the data in the domain we study. Experiments can be best if experiments are feasible (but they seldom are in social science). Large-*n* can be best if we want to test a prime hypothesis, and if we have many well-recorded cases to study. Case studies can be best if we want to infer or test explanatory hypotheses, or if cases have been unevenly recorded—a few are recorded in great detail, many in scant detail. There is no uniform answer to the question “which method is best?”

### Testing Theories with Case Studies

Case studies can serve five main purposes: testing theories, creating theories, identifying antecedent conditions, testing the importance of these antecedent conditions, and explaining cases of intrinsic importance.<sup>17</sup> The first four purposes are similar in

17. These purposes overlap and several (such as, for example, explaining cases, creating theories and testing theories; and identifying and testing antecedent conditions) often can and should be pursued simultaneously. For another typology of case-study formats see Lijphart, “Comparative Politics and the Compara-

their logic and are realized using the same basic methods. Although each purpose merits its own discussion, readers familiar with this material may wish to skip to the section "Explaining Cases" after reading this one.

Case studies offer three formats for testing theories:<sup>18</sup> controlled comparison, congruence procedures, and process tracing. Controlled comparison uses comparative observations across cases to test theories. Congruence procedures are of two types, with one type using comparative observations across cases to test theories, the other using observations within cases. Process tracing tests theories using observations within cases. Congruence procedure and process tracing are stronger test methods than controlled comparison. (All three are also used to create theories and to infer and test antecedent conditions.)

In each testing format the same three steps should be followed: (1) state the theory; (2) state expectations about what we should observe in the case if the theory is valid, and what we should observe if it is false; and (3) explore the case (or cases) looking for congruence or incongruity between expectation and observation.

#### *Controlled Comparison*

In controlled comparison<sup>19</sup> the investigator explores paired observations in two or more cases, asking if values on the pairs are

tive Method." He distinguishes six types of case study (p. 691): (1) atheoretical, (2) interpretive, (3) hypothesis-generating, (4) theory confirming, (5) theory infirming, and (6) deviant. Five of Lijphart's categories overlap with three of mine: his study types 4 and 5 are types of theory-testing study, his type 2 is a case-explaining study, and his types 3 and 6 are types of theory-making study. Lijphart's first study-type is descriptive history, a study-type I do not consider here. He omits antecedent-condition-identifying and condition-testing studies, which I consider later in this chapter.

18. Theory-testing case studies are also known as "theory confirming" and "theory infirming" studies. Lijphart, "Comparative Politics and the Comparative Method," p. 692.

19. See George and McKeown, "Case Studies and Theories," pp. 24–29, and works discussed in Collier, "Comparative Method," pp. 111–12 (his section on "Focus on Comparable Cases").

congruent or incongruent with the test theory's predictions. For example, if values on the independent variable (IV) are higher in case *A* than case *B*, values on the dependent variable (DV) should also be higher in case *A* than *B*. If values on the DV are in fact higher in case *A* than *B*, the theory passes the test. If they are much higher, this indicates that the theory has large importance—that variance in the value on the IV will cause large variance in the value on the DV. If they are only a little higher, the test is again passed but the result suggests that the theory has little importance.

Case selection follows John Stuart Mill's "method of difference" or "method of agreement."<sup>20</sup> In the method of difference the investigator chooses cases with similar general characteristics and different values on the study variable (the variable whose causes or effects we seek to establish). If we seek to establish the causes of the study variable, the investigator then asks if values on the study variable correspond across cases with values on variables that define its possible causes. If we seek to establish the effects of the study variable, the investigator asks if its values correspond across cases with values on variables that define its possible effects. In each instance similar cases are picked to control for the effect of third variables: the more similar the cases, the less likely that the action of third variables explains passed tests.

In the method of agreement the investigator chooses cases with different general characteristics and similar values on the study variable. The investigator then asks if values on the study variable correspond across cases with values on variables that define its possible effects (or its causes, if we seek to establish these).

Controlled comparison is the most familiar case-study method but also the weakest. The method of difference is weak because in social science the characteristics of paired cases are never nearly

20. John Stuart Mill, *A System of Logic*, ed. J. M. Robson (Toronto: University of Toronto Press, 1973), chap. 8, "Of the Four Methods of Experimental Inquiry," pp. 388–406. George and McKeown discuss Mill in "Case Studies and Theories," pp. 26–28.

identical (as the method of difference requires). The method of agreement is even weaker because paired cases usually deviate even further from having wholly different characteristics (as the method of agreement requires).<sup>21</sup>

### *Congruence Procedures*

When using congruence procedures<sup>22</sup> the investigator explores the case looking for congruence or incongruence between values observed on the independent and dependent variable and values predicted by the test hypothesis. Two types of congruence procedure are used.

1. *Congruence procedure type 1: Comparison to typical values.* The investigator observes values on the IV and DV within a particular case and observes the world (without doing further case studies) to ascertain values on the IV and DV that are typical in most other cases. The investigator then deduces from these observations and from the test theory expected relative values for the IV and DV in the study case and measures the congruence or incongruence between expectation and observation. For example, in a given case, if the IV's value is above the typical norm, the value on the DV should also be above normal if the theory holds water.<sup>23</sup> If values on the DV are in fact above normal, the theory passes the test. If they are far above normal, this indicates that the theory has large

21. Noting these and other difficulties with controlled comparison are George and McKeown, "Case Studies and Theories," p. 27, and Stanley Lieberman, "Small N's and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases," *Social Forces* 70 (December 1991): 307-20.

22. On congruence procedure see Alexander L. George, "The Causal Nexus between Cognitive Beliefs and Decision-Making Behavior: The 'Operational Code' Belief System," in Lawrence S. Falkowski, ed., *Psychological Models in International Politics* (Boulder: Westview, 1979), pp. 105-13, and George and McKeown, "Case Studies and Theories," pp. 29-34.

23. If the value on the IV is below normal, the DV's value should also be below normal.

importance, that variance in the value on the IV will cause large variance on the DV. If they are only a little above normal, the test is again passed but the result suggests that the theory has little importance.

Thus to test the hypothesis that "economic downturns cause scapegoating of ethnic minorities," we would explore cases of downturns (for example, the United States in the Great Depression of the 1930s), asking if ethnic scapegoating was above normal in these cases; or we would explore cases of prosperity (the United States in the 1960s) and ask if ethnic scapegoating was below normal.

How do we ascertain normal IV and DV values? Often the normal background levels of phenomena are a matter of common knowledge. Thus we know that the U.S. economy of the 1930s was more depressed than typical modern industrial economies without doing new studies to prove it. We know that Nazi Germany and Stalin's Soviet Union were more murderous than typical modern industrial states, and we can safely compare their conduct to this typical conduct in a case study. We know that elite belief that conquest was feasible was above the historical average in Europe in 1914. If common knowledge is thin or unreliable, however, research to establish typical values is necessary.

Congruence procedure type 1 works best if we select cases with extreme (very high or very low) values on the SV. Thus to test the hypothesis that "economic downturns cause scapegoating of ethnic minorities," we would explore the United States in the 1930s (an extreme downturn) instead of a lesser U.S. or European recession. To test the hypothesis that "belief that conquest is easy causes war," we would explore Europe during 1910-14 (when such beliefs were unusually widespread) instead of more normal times.<sup>24</sup> Cases that exhibit extreme values on study variables are

24. An example from physics is found in the test of Einstein's general theory of relativity conducted with photographs of the May 29, 1919, solar eclipse. Eins-

good test laboratories because theories make more unique and certain predictions about them. This allows stronger tests. In a case where a theory's causal phenomenon is extremely abundant its effects (including both its intervening and dependent variables) should also be abundant. Likewise, if the cause is unusually scarce, its effects should also be scarce. If we observe these extreme results, it is unlikely that they arise from measurement error, since only a large error would cause the observed result. The action of a third variable is also an unlikely cause of the observation, since it is unlikely that another cause operates strongly enough to produce the striking effect that the theory predicts. And any third variable that was responsible would also be abundant, hence it would stand out against the case background, making it easy to spot. Hence we can more easily rule out measurement error and omitted-variable explanations for passed tests. (In other words, the tested prediction is quite unique, hence the test is strong.)

If we fail to observe the predicted result, it is less likely that measurement error or the countervailing effect of other variables caused the failure. Since a large result was predicted, that result should have overpowered any measurement errors or countervailing variables, appearing despite them. Moreover, a countervailing variable would probably need to be abundant and hence be easy to spot. Hence a test failure in the absence of a visible powerful countervailing variable casts large doubt on the theory. (In other words, the tested prediction is quite certain, hence the test is strong.)

---

tein's theory predicted that gravity would bend the path of light. Accordingly, scientists looked to the strongest gravity source they could find—the sun—and asked if it could bend the path of light. (Photographs of an eclipse were used because its darkness made stars near the sun visible, letting scientists observe if these stars appeared displaced by its gravity.) In other words, scientists selected a case where the value on the IV (gravity) was as high as possible. On this test see the discussion on process tracing in this section.

Hence both the passage and the flunking of tests provides decisive evidence in cases with extreme values on the study variable. Passage strongly corroborates the hypothesis, a flunk strongly infirms it.

Congruence procedure type 1 is a close cousin of controlled comparison. Both rest on comparisons across cases, not within them.<sup>25</sup> Both offer means to reduce the possibility that passed tests result from the action of third variables. They differ in the method of reducing this possibility. Controlled comparison holds the case background constant, thus preventing the variance of potentially perturbing third variables. Thus it narrows the range of variables that vary across cases, which lowers the number of potential perturbing variables. In contrast, if an "extreme value on study variable" case-selection method is used, congruence type 1 reduces omitted-variable problems by expanding the impact that omitted variables must generate to produce the result predicted by the theory. This lowers the likelihood that any third variables have enough impact to produce this result and also ensures that these variables' necessarily extreme values will call attention to themselves if they do produce this result. Thus the number of potential perturbing variables is again reduced, this time by a different method.

2. *Congruence procedure type 2: Multiple within-case comparisons.* The investigator makes a number of paired observations of values

25. George describes congruence procedure (by which he means Type 1 congruence procedure) as a within-case case study because he argues that congruence or incongruence is established by deduction, not by comparison to typical values in other cases. Specifically, he argues that once we know the value on the IV we can deduce the expected value on the DV from the test theory. Then we assess the congruence or incongruence of this expectation with observed values. He omits the idea that expected within-case DV values are instead established by comparison to typical IV and DV values. George, "Case Studies and Theory Development" (1982), p. 14. However, it seems to me that any such deductive exercise must rest on comparison to typical values in other cases and on expectations about the study case that are calibrated to these typical values. Hence it rests on cross-case comparison.

on the IV and DV across a range of circumstances within a case. Then the investigator assesses whether these values covary in accordance with the predictions of the test hypothesis.<sup>26</sup> If they covary, the test is passed. The greater the amplitude of the DV's covariance with the IV, the greater the theory's importance.

Thus to test the theory that "economic downturns cause scapegoating of ethnic minorities," we might ask whether periodic measures of scapegoating are possible at timed intervals during the Great Depression in the United States during 1929–41; and if they are, we might then ask whether scapegoating increased as the depression deepened during 1929–33, and whether scapegoating eased as the depression later eased.

Congruence procedure type 2 works best if we select cases with two characteristics: (1) many observations of values on the IV and DV are possible; and/or (2) values on the IV or DV vary sharply over time or across space (across region, institution, group, and so on) within the case.

Cases allowing many observations are better test laboratories because they allow more measures of congruence, and tests that rest on more measures are stronger.

Cases with large variation in values on the IV or DV are also good test laboratories because theories make more unique and certain predictions about these cases. For example, if values on a theory's IV vary sharply, values on its DV should also vary sharply. This sharp variance on the DV is unlikely to arise from measurement error, since the error would need to be large and to gyrate in step with the IV—an unlikely combination. The action of a third variable is also an unlikely cause, since this would require a

26. Alexander George, who coined the concept of congruence procedure, does not mention multiple within-case comparisons as a type of congruence procedure in his various writings on case studies, but his discussions of congruence procedure are consistent with the possibility of multiple observations and comparisons. See, for example, George, "Case Studies and Theory Development" (1982), pp. 13–15, and George and McKeown, "Case Studies and Theories," pp. 29–34.

third variable that gyrates in step with *A* and as markedly as *A*—an unlikely possibility, and one that is easily assessed, since such a variable will leap out from the case. Hence we can more easily rule out measurement error and omitted-variable explanations for passed tests. For parallel reasons we can also rule out these explanations for failed tests. As a result both the passage and the flunking of tests provide decisive evidence in cases with sharp variance on the IV. Passage strongly corroborates the hypothesis, failure strongly infirms it.

Congruence procedure type 2 is a within-case case study, but it shades into large-*n* analysis at some point as the number of within-case observations grows larger and values on variables are assigned numeric values. For example, the 1994 U.S. election is a bounded example of something more general—a parliamentary election in a democracy—hence it has aspects of a case. It also allows hundreds of observations expressed in numeric form—for example, the 435 U.S. house races. These can be studied and statistically compared. Such a study has aspects of a congruence procedure type 2 and a large-*n* analysis.

Case studies could, in principal, be hybrids of congruence procedure types 1 and 2. An analyst could make many observations of a case and compare these observations both to each other and to a typical value. An analyst could also select cases that offer many observations of IV and DV, extreme values on IV or DV, and large variance in these values, all at the same time.

Both types of congruence procedure can be used to test a theory's explanatory hypotheses as well as its prime hypothesis.<sup>27</sup> To test explanatory hypotheses, we prefer cases that permit multiple measures of, and/or display extreme or sharply varying values on, either or both variables in the tested explanatory hypothesis.

27. If a theory holds that " $A \rightarrow q \rightarrow B$ ," then " $A \rightarrow B$ " is its prime hypothesis, " $A \rightarrow q$ " and " $q \rightarrow B$ " are its explanatory hypotheses. See "What Is a Theory?" in Chapter 1.



*Process Tracing*

In process tracing<sup>28</sup> the investigator explores the chain of events or the decision-making process by which initial case conditions are translated into case outcomes. The cause-effect link that connects independent variable and outcome is unwrapped and divided into smaller steps; then the investigator looks for observable evidence of each step.

For example, if "asteroid impacts cause mass extinctions," we should find evidence of an asteroid-caused mass killing mechanism in the sedimentary record of mass extinctions that coincide with asteroid impacts. Perhaps an impact would kill by spraying the world with molten rock, igniting global forest fires that blacken the skies in smoke, shutting out sunlight and freezing the earth. If so, the sedimentary record of mass extinctions should contain the remains of a vast continental or global molten rock shower, a layer of soot, and evidence of an abrupt mass dying of species—evidence of each step of the killing process.<sup>29</sup>

Likewise, if "bipolar distributions of international power cause peace," we should find, in cases of peaceful bipolarity, evidence of intervening phenomena that form the causal chains by which bipolarity causes peace. Kenneth Waltz, the prime exponent of the peacefulness of bipolarity, suggests that bipolarity causes the following pacifying phenomena: less false optimism by governments about the relative power of opponents; easier cooperation and faster learning by each side about the other, leading to thicker

28. See George and McKeown, "Case Studies and Theories," pp. 34–41; also see King, Keohane, and Verba, *Designing Social Inquiry*, pp. 226–28.

29. Evidence from the sedimentary record laid down at the time of the dinosaurs' demise 65 million years ago—which coincided with an asteroid impact—confirms these predictions. Walter Alvarez and Frank Asaro, "An Extraterrestrial Impact," *Scientific American*, October 1990, pp. 79–82. In the section "How Can Specific Events Be Explained?" in Chapter 1 I noted that this same evidence could be used to test an explanation for a specific event (the impact theory of the dinosaur's demise). This illustrates that the same evidence can test both general theories and specific explanations.

rules of the game; faster and more efficient internal and external moves by each side to balance growth in the other's power or to check the other's aggressive moves, causing deterrence; and the selection of fewer inept national political leaders.<sup>30</sup> A process-tracing test would look for evidence of these phenomena in cases of bipolarity (for example, the cold war, 1947–89) and, if they are found, for evidence that they stemmed from bipolarity (for example, testimony by policymakers that reveals motives and perceptions that fit this interpretation).

Evidence that a given stimulus caused a given response can be sought in the sequence and structure of events and/or in the testimony of actors explaining why they acted as they did. For example, if "commercial competition causes war," case studies of outbreaks of war should reveal pressure for war by commercially interested elites, government decisions for war should follow (not precede) this pressure, and statements in the diaries, private correspondence, memoirs, and so on of government officials should indicate that government decisions reflected this pressure. If "public relations campaigns shape public opinion," shifts in public opinion should quickly follow public relations campaigns, and interviews with citizens should reveal that they absorbed the campaign's message and changed their views in response to it.

Process predictions are often unique—no other theories predict the same pattern of events or the same actor testimony on their motives—hence process tracing often offers strong tests of a theory.<sup>31</sup> Hence a thorough process-trace of a single case can provide a strong test of a theory. As noted above, the investigator will still be unsure what antecedent conditions the theory may require to operate, and discovering these conditions remains an important

30. Kenneth N. Waltz, *Theory of International Politics* (Reading, Mass.: Addison-Wesley, 1979), pp. 161–76. My list of Waltz's hypotheses is incomplete; for the rest see *ibid.*

31. In other words, process tracing often provides "smoking-gun" tests—see the section "Strong/Weak Tests" in Chapter 1.

task. They can be found only by exploring other cases. Still, the validity of the theory and its ability to explain at least one case are strongly corroborated.

Theories assume many causal patterns, and we can adapt process-traces to fit these patterns. Some theories (such as, for example, the asteroid impact theory of extinctions) posit a single causal chain:

$$A \rightarrow p \rightarrow q \rightarrow r \rightarrow B$$

Some (Waltz's polarity theory of war) posit several chains:

$$\begin{array}{c} \rightarrow p \rightarrow \\ A \rightarrow \rightarrow q \rightarrow \rightarrow B \\ \rightarrow r \rightarrow \end{array}$$

A complete process-trace looks for evidence of all links in all the chains.

Case studies can assume several or all of these four formats (controlled comparison, congruence type 1, congruence type 2, and process-trace) at once. In the same study we can compare a single case to another case chosen according to Mill's method of difference or to typical conditions, we can examine it to see if within-case measures of values on the IV and DV covary over time and space, and we can study it for evidence that corroborates or infirms the theory's explanatory hypotheses.

How strong are the theory-tests that case studies pose? Scientists tested Albert Einstein's general theory of relativity with a single real-time congruence procedure type 1 case study: the observation of the May 29, 1919, solar eclipse. Einstein's theory predicted that gravity would bend the path of light toward a gravity source by a specific amount. Hence it predicted that during a solar eclipse stars near the sun would appear displaced—stars actually

behind the sun would appear next to it, and stars lying next to the sun would appear farther from it—and it predicted the amount of apparent displacement. No other theory made these predictions. The passage of this one single-case-study test brought the theory wide acceptance because the tested predictions were unique—there was no plausible competing explanation for the predicted result—hence the passed test was very strong.<sup>32</sup> Any case study that reliably tests equally unique predictions can offer equally decisive results. Social science case studies will seldom be so decisive, but this problem stems from the messy nature of social science data and the complexity of social phenomena, not the inherent weakness of the case method.

### Creating Theories with Case Studies

Case studies can serve five main purposes: testing theories, creating theories, identifying antecedent conditions, testing the importance of these antecedant conditions, and explaining cases

32. A synopsis of these events is found in Albert Einstein, *Relativity: The Special and the General Theory: A Popular Exposition*, trans. Robert W. Lawson (New York: Crown Publishers, 1961), pp. 123–32. Scientists conducted this case study in real time, studying the eclipse as it occurred, but they could just as well have studied a past eclipse if one had been photographed showing the positions of nearby stars during the eclipse.

Another decisive real-time case study from physics (also a congruence procedure type 1 study) is found in the observation of the return of Halley's comet in 1758–59, which powerfully supported Isaac Newton's theory of gravity. In 1705 astronomer Edmond Halley applied Newton's theory to predict the comet's return in 1758–59; it appeared almost exactly on schedule. Donald K. Yeomans, *Comets: A Chronological History of Observation, Science, Myth, and Folklore* (New York: John Wiley, 1991), pp. 118–19, 136. Opponents of Newton's theory "were silenced" when this and another real-time single-case-study test, the 1737 measurement of the flattening of the earth, "confirmed Newton's theory beyond a shadow of a doubt." J. Lévy, "The Solar System," in René Taton, ed., *The Beginnings of Modern Science: From 1450 to 1800* (New York: Basic Books, 1964), p. 438.

of intrinsic importance.<sup>33</sup> The previous section discussed theory testing. This section covers theory-making.<sup>34</sup>

To infer new theories from cases we start by searching cases for associations between phenomena and for testimony by people who directly experienced the case (actors in the case, for instance) on their motives and beliefs about the case. These associations and participant accounts offer clues on cause and effect. Then we ask: "Of what more general phenomena are these specific causes and effects examples?" Once candidate causes and effects are framed in general terms the investigator has theories that can be tested against other evidence and applied to other cases.

Investigators can use four basic methods to infer theories from case studies: controlled comparison, congruence procedures, and process tracing (all covered in the previous section) and the Delphi method. Controlled comparison compares observations across cases to infer theories. Congruence procedure and process tracing deduce theories from observations within cases. The Delphi method consults the views of case participants.

#### *Controlled Comparison*

In a controlled comparison the investigator infers hypotheses from contrasts or similarities in aspects of several cases, following Mill's methods of difference and agreement. In the method of difference the investigator explores several cases with similar characteristics and different values on the study variable (the variable whose causes or effects we seek to discover), looking for

33. These purposes overlap and several (for example, explaining cases, creating theories and testing theories; identifying and testing antecedent conditions) often can and should be pursued simultaneously. For another typology of case-study formats see note 17.

34. Theory-making case studies are also called "heuristic," "hypothesis-generating," and "exploratory" case studies. Eckstein, "Case Study and Theory," pp. 104-8; Lijphart, "Comparative Politics and the Comparative Method," p. 692; Yin, *Case Study Research*, pp. 1, 3-4.

other differences between the cases. These other cross-case differences are nominated as possible causes of the study variable (if we seek to discover its causes) or possible effects (if we seek its effects). The investigator picks similar cases to reduce the number of candidate causes or effects that emerge: the more similar the cases, the fewer the candidate causes. This makes the real cause easier to spot.<sup>35</sup>

In the method of agreement the analyst explores cases with different characteristics and similar values on the study variable, looking for other similarities between the cases. These similarities are nominated as candidate causes or effects of the variable.

The method of difference is preferred when the characteristics of available cases are homogeneous (most things about most cases are quite similar). The method of agreement is preferred when the characteristics of available cases are heterogeneous (most things about most cases are different).

#### *Congruence Procedures*

When using congruence procedures, the investigator explores a case looking for within-case correlation between the study variable and other phenomena. These phenomena are nominated as possible independent variables in new hypotheses (if we seek to establish the study variable's causes) or as possible dependent variables (if we seek to establish its effects). Three specific formats are used.

1. The investigator "examines the outliers," exploring cases that are poorly explained by known causes, on the assumption that unknown causes explain their outcomes. Specifically, the investigator looks for cases where the study phenomenon is present but its known causes are absent. Still-undiscovered causes must explain the phenomenon. These causal phenomena should be ob-

35. For examples of method-of-difference theory making, see note 25 to Chapter 1.

served in above-normal amounts in the case and should be observed covarying with the study variable.<sup>36</sup>

2. The investigator selects cases with extreme high or low values on the study variable, and explores them looking for other phenomena that are present in above-normal or below-normal amounts. When the study phenomenon is present in abundance, its causes and effects should also be present in unusual abundance and so should stand out against the case background. When the study phenomenon is absent, its causes and effects should also be prominent by their absence.

3. The investigator selects cases with extreme within-case variance on the dependent variable, and explores them looking for phenomena that covary with the study variable. If values on the study variable vary sharply, its causes and effects should also vary sharply, standing out against the more static case background.

#### *Process Tracing*

The investigator traces backward the causal process that produces the case outcome, at each stage inferring from the context what caused each cause. If this backward process-trace succeeds, it leads the investigator back to a prime cause.

#### *The Delphi Method*

In the Delphi method the investigator mines the views of case participants or others who experienced the case for hypotheses. Those who experience a case often observe important unrecorded data that is lost to later investigators. The investigator uses their

36. Studies of outlier cases are also known as "deviant" case studies. Lijphart, "Comparative Politics and the Comparative Method," pp. 692–93. The logic of studying outliers follows the logic of John Stuart Mill's "method of residues"; on this method see Mill, *A System of Logic*, pp. 397–98.

memories and judgments to infer hypotheses that could not be made from direct observation alone.<sup>37</sup>

#### **Inferring Antecedent Conditions from Case Studies**

As noted above, a weakness of the single-case study is its concealment of theories' antecedent conditions—the background conditions required for theories to operate or that magnify their action. However, one can uncover these background conditions by examining selected new cases.

Four methods of inferring antecedent conditions are most useful. (These methods parallel the four methods of inferring theories, outlined above.)

1. *Controlled comparison.* The investigator uses Mill's method of difference to infer antecedent conditions from contrasts or similarities in the characteristics of several cases.<sup>38</sup> Specifically, the investigator selects and examines new cases that resemble previously studied cases in all ways—except their value on the dependent variable. For example, if previously examined cases had high values on the independent and dependent variables, we would now examine cases with high values on the IV, low values on the DV, and a close resemblance to the previously examined cases in other regards. Thus if the hypothesis that "economic downturns cause trade closure" has been tested using Europe 1929–39 as a case (high values on IV and DV), we would next look for cases when downturns occurred without closure. If we can find no high-IV–low-DV cases, this suggests that the conditions required for the theory's operation are abundant, and the theory

37. For an example of Delphi-method theory-making see Chapter 1, note 27. The Delphi method does not have much stature as a method for testing theories or explanations partly because the Delphi expert's discovery process cannot be replicated.

38. The method of agreement is too weak to bother with.

has broad applicability (or "external validity"). If we find such cases, we inspect them for points of difference with previously examined cases. Important antecedent conditions will appear as these points of cross-case difference.

2. *Congruence procedures.* The investigator measures the gap between the predicted and observed values on the dependent variable in a case and then looks for correlations between the size of the gap and values on other phenomena within the case. The investigator then nominates phenomena that correlate with the gap (that are scarce when the value on the DV is lower than the IV value warrants and are abundant when the DV value is higher than the IV value warrants) as possible antecedent conditions. Two formats are used.

The investigator can examine the outliers—those cases where the theory's posited cause is present but its predicted effect is absent. If we assume the theory is valid, this pattern indicates that an important antecedent condition is also notably absent. The missing antecedent condition can be identified among conditions that are often present but are absent in the outlier case.

The investigator can also explore cases with large within-case variance in the value on the dependent variable and constant high values on the independent variable. This pattern suggests that an important antecedent condition varies within the case.<sup>39</sup> It should announce itself as a factor that covaries with the DV.

3. *Process tracing.* The investigator traces backward the causal process by which the case outcome was produced, at each stage attempting to infer from the context what antecedent conditions the process requires.

4. *The Delphi method.* The investigator mines the views of case participants or others who experienced the case for possible

39. This pattern may also occur solely because of variance in the value of other variables that cause the DV.

antecedent conditions. They may have observed in person tell-tale dynamics that cannot be observed in retrospect by non-participants.

### Testing Antecedent Conditions with Case Studies

Antecedent conditions, like hypotheses, should be tested before they are lent credibility. Like hypotheses they can be tested three ways: controlled comparison, congruence procedures, and process tracing.

1. *Controlled comparison.* The investigator explores paired observations in two or more cases, asking if values on the pairs are congruent or incongruent with the premise that the antecedent condition magnifies the causal action of the independent variable on the dependent variable. For example, if values on the condition variable are higher in case *A* than case *B*, values on the dependent variable should also be higher, relative to values on the independent variable, in case *A* than *B*. If possible, the investigator selects cases according to an adaption of Mill's criteria for the method of difference: cases should have similar general characteristics, similar values on the IV, and different outcomes. If the condition variable (CV) has an impact, its values should covary with values on the DV.

2. *Congruence procedures.* Two congruence procedures are most useful for testing antecedent conditions. First, the investigator studies cases with extreme (high or low) values on the condition variable and a value greater than zero on the independent variable. A very high value on the CV should multiply the effects of the IV on the intervening variables (IntVs) and DV, moving their values above predicted ranges (with "predicted" meaning the value predicted by the test theory in light of the value on the IV in the case). A very low value on the CV should diminish the IV's

impact on IntVs and DV, lowering their values below predicted ranges.<sup>40</sup> Second, the investigator studies cases with large within-case variance on the value of the CV and little or no within-case variance on the IV. If the CV is important, the DV's value should covary with it.<sup>41</sup>

3. *Process tracing*. The investigator explores the chain of events or the decision-making process by which initial case conditions are translated into case outcomes. Antecedent conditions will leave footprints in this process: actors may refer to their importance and events will occur in a sequence that follows their appearance and disappearance.

### Explaining Cases

As noted in Chapter 1,<sup>42</sup> explanations for specific cases<sup>43</sup> are assessed by answering four questions:

1. Does the explanation exemplify a valid general theory (a covering law)? The specific explanation must exemplify a valid covering law. An explanation that rests on a false general theory falls.

2. Is the covering law's causal phenomenon present in the case? The explanation's causal phenomenon must be present in the case.

40. This test assumes that the catalytic effect of the CV is linear, expanding continuously as the value on the CV rises. It is inappropriate if logic suggests that the impact of the CV hits a threshold at some point, flattening out when the value of the CV rises above a certain level.

41. These two methods parallel the methods of congruence procedure type 1 (comparison to typical values) and type 2 (multiple within-case comparisons). See the discussion of congruence procedures in the section "Testing Theories with Case Studies" in this chapter.

42. See "How Can Specific Events Be Explained?" in Chapter 1.

43. Case-explaining studies are also called "explanatory," "interpretive," and "disciplined-configurative" case studies. Yin, *Case Study Research*, p. 5; Lijphart, "Comparative Politics and the Comparative Method," p. 692; Eckstein, "Case Study and Theory," pp. 99-104.

If not, the explanation falls. (Even if *A* is a confirmed cause of *B*, it cannot explain instances of *B* that occur when *A* is absent.)

3. Are the covering law's antecedent conditions met in the case? Theories cannot explain the outcomes of cases that omit their necessary antecedent conditions.

4. Are the covering law's intervening phenomena observed in the case? The phenomena that link the covering law's posited cause and effect should be evident in the case and appear in the proper order and location.

The logic of case-explaining parallels that of a pathologist doing an autopsy or a detective solving a crime. Specific explanations of the death (or crime) are evaluated by asking if they rest on a valid covering law, if the conditions for that covering law's operation—its cause and required antecedent conditions—are observed in the case at hand, and if telltale phenomena that signal its inner workings are also observed. A case-explaining inquiry does not test theories, although the evidence collected could also be used to check a theory's validity.

Political scientists seldom do case-explaining case studies, partly because they define the task of case-explaining as the domain of historians; however, historians often explain cases in a softer way than political scientists would. Their explanations are left vague, and the predictions they infer from these explanations are left unspecified, hence the meaning of their evidence is often ambiguous. The general theories that underlie their explanations are often deeply buried. As a result their explanations are hard to interpret and evaluate. This leaves wide latitude for political scientists to contribute to discussion of historical explanation.

### Strong vs. Weak Tests; Predictions and Tests

Strong tests are better than weak tests, and the results of strong tests carry more weight than the results of weak tests.

As noted in the section "Strong vs. Weak Tests" in Chapter 1, a strong test is one whose outcome is unlikely to result from any factor except the operation or failure of the theory. Strong tests evaluate predictions that are *certain* and *unique*. A *certain* prediction is an unequivocal forecast. The more certain the prediction, the stronger the test. A *unique* prediction is a forecast not made by other known theories. The more unique the prediction, the stronger the test.

When testing a theory, the investigator should select cases that enable the most strong tests. This calls for selection of cases about which the test theory makes certain or unique predictions (or both).

In writing up cases, authors should explain and justify the predictions they test. Interpretive disputes about case studies often arise from disputes about the fairness of the predictions tested. These disputes can be rationalized by offering a few words on why the prediction seems fair.

Authors should also comment on the strength of the tests performed. How unique and how certain were the predictions tested? Were the tests of the smoking-gun, hoop, doubly-decisive, or straw-in-the-wind variety?<sup>44</sup>

### Interpreting Contradictory Results

What should investigators do when tests produce contrary results—when theories pass some tests and flunk others? Answer: investigate further. Five procedures are appropriate:

1. Infer and test additional predictions, with a special eye toward finding "hoop" and "smoking-gun" tests. Such additional tests may resolve the confusion.
2. Double-check the accuracy of data used for past tests. Some

44. On these types of test see the section "Strong vs. Weak Tests" in Chapter 1.

may be wrong. If so, an unambiguous result may emerge from double-checking: all tests may now be passed or flunked.

3. Reconsider the predictions you inferred from the theory. Were they fair? Sometimes false flunks (or false passes) are reported because false predictions are tested.

4. Replicate your tests using new cases. Replication may produce more consistent results.

5. Repair the theory in ways that enable it to pass flunked tests, by limiting the scope of its claims or by removing flunked explanatory hypotheses. This can salvage a damaged theory (although the salvaged product is now a different, narrower theory).

### Case-Selection Criteria

Practitioners of case studies have produced neither a comprehensive catalog of possible case-study research designs<sup>45</sup> nor a comprehensive list of case-selection methods. Accordingly, I have made my own list of useful case-selection criteria.<sup>46</sup> My list (of eleven criteria) does not exhaust the logical possibilities, but it includes all that seem strong to me. Specifically, I argue that the following case attributes are possible reasons for case selection: (1) data richness; (2) extreme values on the independent variable, dependent variable, or condition variable; (3) large within-case variance in values on the independent, dependent, or condition variables; (4) divergence of predictions made of the case by competing theories; (5) the resemblance of case background conditions to the conditions of current policy problems; (6) prototypicality of case background conditions; (7) appropriateness for controlled comparison with other cases (mainly using Mill's method of difference); (8) outlier character; (9) intrinsic impor-

45. Noting this failure is Yin, *Case Study Research*, p. 18.

46. These criteria evolved from discussions with Andy Bennett, Tom Christensen, Chaim Kaufmann, Jack Snyder, and Steve Walt and include their ideas.

tance; (10) appropriateness for replication of previous tests; and (11) appropriateness for performing a previously omitted type of test.

This list reflects two general criteria for case selection:

First, investigators should select cases that best serve the purpose of their inquiry. As noted earlier, there are five purposes for case studies: testing theories, creating theories, identifying possible antecedent conditions that theories require to operate, testing the importance of these antecedent conditions, and explaining cases of intrinsic importance. The selection criterion that is most appropriate differs from purpose to purpose, hence investigators should be clear in their purpose before they select cases.<sup>47</sup> Some of the following selection criteria are appropriate for most purposes, but some serve only one or two purposes. Hence investigators should take care to match criteria and purpose. (See the table at the end of this chapter, for a summary of matches and mismatches between mission and case-selection criteria.)

Case-selection criteria should therefore differ with the stage at which the investigation stands. Investigators first seek to infer theories, then to test theories, then to test their range (or "external validity") by inferring and testing antecedent conditions. Rules for case selection vary across these tasks, and hence vary with the stage of the inquiry.

Second, when testing theories investigators should select cases to maximize the strength and number of tests they let the investigator perform. The best case selection allows the most strong tests (tests of predictions that are certain and/or unique) with the least research effort.<sup>48</sup>

47. This means you cannot know what cases are best to select until you frame your questions. Decisions on case selection are premature before you know what you want to know.

48. Some of what follows repeats remarks made earlier on inferring and testing theories and antecedent conditions, since methods of case selection are an aspect of general methods of inquiry.

1. *Select data-rich cases.* We learn more from case studies that let us answer more questions about the case. The more data we have the more questions we can answer. Hence more tests are possible, hence data-rich cases are preferred, other things being equal.<sup>49</sup>

Selecting cases for data-richness is especially appropriate if you plan to infer or test theories using process tracing, since process tracing requires a great deal of data.

Data richness can take several forms. Abundant archival data may be available. Participants in the case may be alive and available for interviews. Other scholars may have studied the case for their own purposes and done much of the legwork for you.

2. *Select cases with extreme (high or low) values on the independent variable (IV), the dependent variable (DV), or the condition variable (CV).*<sup>50</sup> Under this method we select cases in which the study variable (the variable whose causes or effects we seek to establish) is present in unusually large quantities or unusually small quantities.<sup>51</sup>

To test a theory, select cases with extreme values on the independent variable. Such cases offer strong tests because the theory's predictions about the case are certain and unique (as noted above in this chapter.)

It is often argued that one should select cases that are representative or typical of the universe of cases. The "extreme value on the IV" method of case selection argues the opposite, that cases that are atypical in their endowment with the independent variable teach us the most.<sup>52</sup>

49. Yin, *Case Study Research*, p. 40, concurs.

50. Recommending this selection criteria is Eckstein, "Case Study and Theory," pp. 119–20.

51. This "extreme value" method of case selection is closely akin to method 7, controlled comparison. The difference is only that with method 7 cases are selected to ease explicit cross-case comparisons, whereas here cases are selected to ease implicit comparisons to normal conditions. Such comparisons are clearest if within-case values on IV and DV contrast clearly with their normal values.

52. Thus I chose the 1914 case to test offense-defense theory (which posits that



Some also argue that selecting cases for extreme values on the IV sets up weak tests because passage of the test is likely: high IV values should elevate some DV values even if the theory operates only weakly, hence the test is easy to pass. This view rests, however, on a false definition of "strong test." A strong test is one whose outcome is unlikely to result from any factor other than the operation or failure of the theory. According to this definition, a test using a case selected for extreme value on the IV is a strong test. We should expect extreme results in such a test.<sup>53</sup> If they occur, these extreme results are unlikely to stem from other factors. If they do not occur, this is unlikely to stem from any cause other than the theory's failure. Hence cases with extreme IV values are laboratories for strong tests.

To make a theory, select cases with extreme values on the study variable. If values on the study variable are very high, its causes (or effects, if these are sought) should be present in unusual abundance, hence these causes (or effects) should stand out against the background of the case more clearly. This makes them easier to

---

war is more likely when conquest is believed easy) partly because 1914 was the heyday of the "cult of the offensive," a remarkable European elite belief that conquest was easy. Shortly before 1914 this belief reached heights never seen before or since. Because it was extreme it should have had extreme effects, if it ever has any effects. Hence these effects should have been clearly visible in 1914, standing out starkly from the European political landscape, and should have appeared in quantities unlikely to be produced by measurement error or by the action of other causes.

The predictions of offense-defense theory in 1914 therefore are both certain and unique. The predicted effects are too large to ascribe to measurement error or to other causes (hence the predictions are unique). The absence of these effects likewise could not be plausibly blamed on measurement error or the overriding effects of other phenomena, since these could not mask or override such large effects (hence the predictions are certain). Hence the test posed by the 1914 case is strong. Offense-defense theory would be strongly corroborated by passing it and badly damaged by flunking it. See Van Evera, *Causes of War* (Ithaca, N.Y.: Cornell University Press, in press), vol. 1, chap. 7.

53. I assume there is no threshold effect, that the impact of the IV on the DV does not peter out above a certain threshold. With threshold effects this selection criteria is less useful.

spot. Likewise, if values on the study variable are unusually low its causes (or effects) should be made more striking by their absence.

To infer antecedent conditions, select cases with extreme and opposite values on the IV and DV—specifically, with very high values on IVs and very low values on DVs. These are cases where the theory's posited cause is abundantly present but the predicted effect is notably absent. For example, to infer conditions required for literacy to cause democracy, we should select highly literate societies with authoritarian regimes. To infer conditions required for economic depression to cause war, we should select cases where deep depressions occurred but no war resulted, and so on. Such a pattern indicates that an important antecedent condition is also notably absent. The missing condition can be identified among conditions that are normally present but are absent in the studied case.<sup>54</sup>

To test a candidate antecedent condition (a condition that a theory requires to operate or that magnifies its action), select cases with extreme values on the condition variable. A high value on the CV should multiply the effect of the IV on intervening variables (IntVs) and DV. A low value on the CV should leave the IV with little impact on IntVs and DV. In both instances the predicted results are pronounced and hence less likely to arise from measurement error or the actions of a third variable.<sup>55</sup>

54. Under the selection methods discussed in the previous four paragraphs, cases are chosen to highlight the contrast between observed and normal values on the IV and DV. Here, in contrast, cases are chosen to highlight the contrast between observed values on IV and DV. Sharp contrasts are sought in both cases, although the sought contrasts differ in nature.

55. This selection strategy needs adjustment if the impact of antecedent conditions follows sharp thresholds—if the IV requires some value on the CV to cause the DV but further increases in the value on the CV have no effects. Thus seed and fertilizer cannot cause grass to grow without some accompanying rainfall, but beyond a certain point enough rain is enough, and too much will drown the grass. In such cases we should select cases for very low values on the CV and test the prediction that the IV should lack causal power in such cases. Selecting cases

3. *Select cases with large within-case variance in the value on the independent variable, dependent variable, or condition variable across time or space.*<sup>56</sup>

To test a theory, select cases with large within-case variance in the value on the independent variable. Theories make predictions about the impact of variance in the value on the IV, hence variance in the IV's value generates predictions, hence the more within-case variance in the IV's value, the more predictions we have to test. Such variance takes the form of diachronic or synchronous change in the value on the IV—that is, change over time within the period covered by the case, or diversity on the value of the IV across regions, groups, organizations, or individuals present in the case.

Selecting cases for within-case IV-value variance is especially appropriate if you plan to use a congruence procedure type 2 for testing, since type 2 congruence procedures rely on observing within-case variance.

To make a theory, select cases with large within-case variance in the value on the study variable. The causes and effects of the study variable should also vary widely in such a case, in step with the study variable. This makes them easier to spot against the case background. Candidate causes and effects will announce themselves as case characteristics that vary with the SV's value—that is, as factors present when the value on the SV is high, missing when it is low.

To infer a theory's antecedent conditions, select cases with large within-case variance in the value on the DV and constant high values on the IV. Such a case contains some observations where the relative value on the IV and DV match the theory's predictions (high on IV and DV), and some where it does not (high

with very large values on the CV is unfruitful, since very large values predict the same results as moderate values.

56. Obliquely recommending this selection criteria is Eckstein, "Case Study and Theory," pp. 119, 126.

on IV, low on DV). Candidate antecedent conditions will announce themselves as factors that are more abundant when relative values match predictions—that is, when DV values are higher.

To test a candidate antecedent condition, select cases with large within-case variance on the value of the CV. If the CV is important, the DV's value should be higher, relative to the IV's value, when the CV is abundant than when it is scarce.

4. *Select cases about which competing theories make opposite predictions.* This selection method is appropriate if you are more interested in testing the relative power of the two theories than testing a theory against the null hypothesis (that is, if you prefer to arrange a Lakatosian "three-cornered fight" over a "two-cornered fight").<sup>57</sup>

If you are testing the relative power of two theories, choose cases about which they make opposite predictions, for instance, a case with opposite within-case variance in the values on the two IVs (values on one IV fall over time and values on the other rise over time). The DV should covary with the stronger IV.

If you are testing the relative power of two antecedent conditions, choose cases where the IV is present and the CVs show opposite within-case variance (e.g., values on one CV fall over time, values on the other rise over time). The DV should covary with the stronger CV.

For best results, select cases that allow tests of predictions that are unique and certain as well as opposite.

5. *Select cases that resemble current situations of policy concern.*<sup>58</sup> A theory inferred from or tested in a case that resembles a second case will more often "travel" to that second case—that is, operate in the second case as well. Hence policy prescriptions deduced from the first case can be more safely applied to the second.

57. On two- and three-cornered fights see note 43 to Chapter 1.

58. Jack Snyder recommends this criterion.

Scholars interested in offering policy prescriptions should therefore study cases whose background characteristics parallel the characteristics of current or future policy problems.

A study of health policy in Minnesota yields more reliable prescriptions for health policy in Wisconsin than a study of health policy in Burkina Faso. Theories that operate in Burkina Faso may well require conditions absent in Wisconsin, hence prescriptions deduced from these theories will prove unsound in Wisconsin. This is less likely of theories that operate in Minnesota, since Minnesota and Wisconsin are similar in many ways.

6. *Select cases with prototypical background characteristics.* One might select cases with average or typical background conditions, on the grounds that theories that pass the tests these cases pose are more likely to "travel" well, applying widely to other cases.

This selection method is sometimes appropriate but is overused. If one is seeking theories with wide applicability, it is often more appropriate to follow selection method 5, "select cases that resemble current situations of policy concern," since that method offers a better guarantee that corroborated theories will apply to other important situations. Method 6 selects theories that apply widely; method 5 selects theories that less widely overall but more widely to important circumstances. The latter goal is often more important.

7. *Select cases that are well matched for controlled cross-case comparisons.* One can select cases to allow their pairing for controlled comparison, that is, for the method of difference (cases have similar characteristics and different values on the study variable) or the method of agreement (cases have different characteristics and similar values on the study variable). The method of difference, being the stronger of the two, is usually preferred.

Controlled-comparison criteria can be applied to select one or more cases. A single case can be selected with an eye to comparing it to existing case studies that others have already researched and written. Specifically, if we plan to perform a method-of-difference

comparison, we select a new case whose characteristics resemble those of an already-studied case but which has different values on the study variable. Multiple cases can be selected with an eye to comparing them (in other words, if we plan to perform a method-of-difference comparison, we select cases with similar characteristics and diverse study-variable values) or with existing cases (we select cases with characteristics similar to those of an already-studied case but with different values on the study variable).

To test a theory using a controlled-comparison method, select cases with similar characteristics and different values on the study variable (that is, select for the method of difference). The theory passes the test if study reveals that values on IV and DV correspond across the cases. If, for example, the IV has a higher value in case 1 than case 2, the DV should also have a higher value in case 1 than case 2.

Remember, however, the method of difference is a fairly weak instrument for theory testing (and the method of agreement is even weaker). Hence other selection criteria should have higher priority for theory-testers.

To make a theory, select cases with similar characteristics and different values on the study variable (for method-of-difference comparison) or cases with different characteristics and similar values on the study variable (for method-of-agreement comparison).

Candidate causes or effects will announce themselves as differences in the characteristics of the compared cases when the method of difference is used. They announce themselves as similarities in the characteristics of compared cases when the method of agreement is used.

The method of difference is preferred when the characteristics of available cases are quite homogeneous (most things about most cases are similar). The method of agreement is preferred when the characteristics of available cases are quite heterogeneous (most things about most cases are different).

To infer antecedent conditions, select cases using variants of the method of difference or the method of agreement.

For the method of difference, choose cases with (1) similar values on the IV; (2) similar case characteristics; and (3) different values on the DV. Candidate antecedent conditions will announce themselves as differences in the characteristics of the compared cases.

For the method of agreement, choose cases with (1) similar values on the IV; (2) different case characteristics; and (3) similar values on the DV. Candidate antecedent conditions will announce themselves as similarities in the characteristics of the compared cases.

When testing a candidate antecedent condition, select cases with (1) similar values on the IV and (2) different values on the DV. The condition passes its test if values on the CV correspond with values on the DV across cases.

8. *Select outlier cases.* Here the investigator selects cases that are poorly explained by existing theories, on the assumption that unknown causes explain their outcomes and can be identified by examining the case. We select cases where the values on the dependent variable are high and its known causes are absent. Candidate new causes will announce themselves as unusual characteristics of these cases and as characteristics that correspond with the DV within the case.

To make a theory, select cases where the DV's known causes are scarce yet the DV is abundantly present. This suggests that unknown causes are operating in the case and that study of the case may reveal them.

To infer an antecedent condition, select cases where the DV's known causes are abundant yet the DV is scarce or absent. This suggests that unknown antecedent conditions are absent in the case, and that study of the case may identify them.

9. *Select cases of intrinsic importance.* Selecting cases of intrinsic human or historical importance (World War I, World War II, the

Holocaust) is appropriate if our object is to explain the course of history. We select such cases with a nod to their data-richness (there is little point in studying cases where the record is too thin to answer our questions), but mainly in accord with the magnitude of their human consequences.

10. *Select for test replication.* Thorough theory testing requires repeating initial tests to corroborate their results. When doing this, we choose cases for their appropriateness as laboratories to replicate previous tests.<sup>59</sup> This approach considers multiple cases as multiple experiments. Test replication, not cross-case comparison, is the goal of later studies in the series.

A replication can be exact or inexact (a "quasi-replication"). An exact replication repeats a previous test exactly with a similar case. A quasi-replication (which is far more common) repeats a previous test with some alteration to the research design.<sup>60</sup> Cases are selected by means of the same selection criteria used to select the case(s) for the test being replicated.

11. *Select for previously omitted types of tests.* If a theory has already faced one kind of test, it may now be appropriate to subject it to a another kind of test. For example, if a theory has already faced congruence procedure case-study tests it may now be appropriate to subject it to a process-tracing case-study test. We select such cases for their utility as process-tracing test beds.

Can the investigator test a theory with the same case from which it was inferred? As I noted in Chapter 1,<sup>61</sup> this practice is criticized on grounds that such tests lack integrity. The criticism rests on a preference for blind testing. The assumption is that data not used to infer a theory are less well known to an investigator

59. Making this point is Yin, *Case Study Research*, pp. 45-50.

60. On quasi-replication see Edward S. Balian, *How to Design, Analyze, and Write Doctoral or Masters Research*, 2d ed. (Lanham, Md.: University Press of America, 1988), pp. 12-13.

61. See the discussion of blind testing (item 3) in the section "Methodology Myths" in Chapter 1.

than used data, hence the investigator using unused data is less tempted to sample the data selectively.

Prohibiting the reuse of theory-inspiring cases for theory testing is not feasible in practice, however, and would cause a loss of good evidence. Other barriers against test-fudging—for example, infusing social science professions with high standards of honesty—are more practical.

Table 1, below, summarizes matches and mismatches of study missions and case-selection criteria.

Table 1. Eleven Case Selection Criteria: When Is Each Appropriate?

Case selection criteria	When testing theories?	When inferring theories?	When inferring antecedent conditions?	When testing antecedent conditions?	When studying cases of intrinsic importance?
1. Data richness	Yes	Yes	Yes	Yes	Maybe
2. Extreme values on IV, DV, or CV	Yes (on IV)	Yes (on SV)	Yes (high on IV, low on DV)	Yes (on CV)	No
3. Large within-case variance in values on IV, DV, or CV	Yes (on IV)	Yes (on SV)	Yes (on DV)	Yes (on CV)	No
4. Competing theories make divergent predictions about the case	Yes	No	No	Yes	No
5. Resemblance to current policy-problem cases	Yes	Yes	Yes	Yes	No
6. Prototypical case characteristics	Sometimes	Sometimes	No	No	No
7. Matched for cross-case controlled comparison (namely, method of difference or agreement)	Seldom	Yes	Yes	Seldom	No
8. Outcome unexplained by other theories (thus, an "outlier" case)	No	Yes	Yes	No	No
9. Intrinsic importance	No	No	No	No	Yes
10. Good case for replicating previous tests	Yes	No	No	Yes	No
11. Allows a new type of test	Yes	No	No	Yes	No

## CHAPTER 3

### What Is a Political Science Dissertation?

Dissertations in political science<sup>1</sup> can perform seven principal missions. This gives rise to seven types of dissertation, one for each mission. Most dissertations perform several of these missions, and thus are hybrids, but it is still useful to consider possible ideal-type dissertations.

1. A *theory-proposing* dissertation advances new hypotheses. A deductive argument for these hypotheses is advanced. Examples may be offered to illustrate these hypotheses and to demonstrate their plausibility, but strong empirical tests are not performed.<sup>2</sup>

1. As my examples suggest, what follows was drafted for students in the subfield of international relations and security affairs. It should also apply to other political science subfields, however, with the exception of political philosophy. Apologies to those in other subfields for my IR-centric examples.

2. Examples of theory-proposing works include Robert Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976); Robert Jervis, "Cooperation Under the Security Dilemma," *World Politics* 30 (January 1978): 167–214; Kenneth N. Waltz, *Theory of International Politics* (Reading, Mass.: Addison-Wesley, 1979); Geoffrey Blainey, *The Causes of War*, 3d ed. (New York: Free Press, 1988); Thomas C. Schelling, *Arms and Influence* (New Haven: Yale University Press, 1966); Thomas C. Schelling, *The Strategy of Conflict* (New York: Oxford University Press, 1960); Robert Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984); Carl von Clausewitz, *On War* (Prince-