

## CHAPTER 27

---

# COUNTERFACTUALS AND CASE STUDIES

---

JACK S. LEVY

IN Henry James's "The Jolly Corner," Spencer Bryden pursues the ghost of the man he might have been had he not left New York City three decades earlier for a more leisurely life abroad. In Dickens's "A Christmas Carol," Ebenezer Scrooge transforms his life after encountering the Ghost of Christmas Yet to Come and learning how the future would play out if he were to continue his old ways. Robert Frost is less explicit about what he would have encountered on "the road not taken," but he knows that the road less traveled "made all the difference."

Scholars frequently speculate about what might have been in history. Pascal famously wrote that "Cleopatra's nose: had it been shorter, the whole face of the world would have changed." It is often said that the First World War would not have occurred without the assassination of Archduke Ferdinand (Lebow 2007), that the Second World War would not have occurred without Hitler (Mueller 1991), and that the end of the cold war would have been significantly delayed without Gorbachev (English 2007). In a more general theoretical claim, Skocpol (1979) argues that without either peasant revolts or state breakdown, social revolutions will not occur.

Opposition politicians frequently invoke counterfactuals in arguing that if the current administration had acted differently, the country would have been better off. Hillary Clinton later defended her 2002 vote to authorize the president to use force in Iraq by saying that "if we knew then what we know now, there wouldn't have been a vote... and I certainly wouldn't have voted that way."<sup>1</sup>

<sup>1</sup> NBC, *Today* show, December 18, 2006.

Some historians are skeptical about the analytic utility of counterfactual analysis, and regard it as entertaining “after dinner history” (Ferguson 1999a, 15) or a “parlour game” (Carr 1964, 97), but not analytically sound scholarship. Croce (cited in Ferguson 1999a, 6) states that it is necessary “to exclude from history the ‘conditional’ which has no rightful place there. . . .” What is forbidden is . . . the “anti-historical and illogical ‘if.’” Fischer (1971) includes an index item for “counterfactual questions” in his book *Historians’ Fallacies*, but it says “see fictional questions.” Fischer describes Fogel’s (1964) pathbreaking counterfactual analysis of the impact of railroads on American economic development as a step “down the methodological rathole” and a return to “ancient metaphysical conundrums” (p. 18). Oakeshott (1966, 128–9) argues that if the historian were to consider “what *might* have happened” and treat “great events” or “turning points” as causally “decisive,” “the result is not merely bad or doubtful history, but the complete rejection of history . . . a monstrous incursion of science into the world of history.”

Other historians, and most social scientists, recognize that counterfactuals are unavoidable. They understand that “the study of history is a study of causes” (Carr 1964, 87), and they recognize that any causal statement involves assumptions about what did not happen but could have happened. Bueno de Mesquita (1996, 229) argues that in applied game theory “we cannot understand what happened in reality without understanding what did not happen but might have happened under other circumstances.” The historian Ferguson (1999a, 87) argues that “To understand . . . [history] as it actually was, we therefore need to understand *how it actually wasn’t*—but how, to contemporaries, it might have been.”

The question is how to validate counterfactual claims about what would have happened in a hypothetical or alternative world in which the hypothesized cause took on a different value, whether in a particular case or in a more general theoretical relationship. Whereas causal statements are in principle amenable to a direct empirical test, the same is not true for counterfactual statements, since the conditional upon which the counterfactual rests does not exist and cannot be fully realized in order to examine the effects that flow from it (Goodman 1983). In the absence of direct empirical confirmation, by what criteria can we say that some counterfactuals are more legitimate or more valid than others, and for what theoretical purposes? Since we cannot avoid counterfactuals, the question, in response to Oakeshott, is how to introduce science into the world of history in a way that enhances our understanding of history.

I focus on the methodologically normative use of counterfactual arguments to advance causal understanding of the social and political world. This differs from psychologists’ more descriptive focus on such cognitive science questions as how people actually use counterfactuals, how they judge the validity of counterfactual arguments, and how their cognitive and motivational biases affect those judgments and influence what kinds of counterfactuals they find most persuasive (Roese and Olson 1995; Tetlock and Belkin 1996; Tetlock and Lebow 2001).

Counterfactuals are relevant in any kind of causal analysis, but I focus primarily on the role of counterfactuals in case study analysis. This subject is particularly

important because qualitative-comparative researchers are more likely than quantitative scholars to posit necessary conditions, which automatically generate explicit counterfactuals (Goertz and Starr 2003).

After explaining why counterfactuals are important for history and social science, I identify the criteria by which we can evaluate the utility of counterfactuals for supporting idiographic and nomothetic causal claims. Throughout I assume that counterfactuals are best conceived as a method, to be used in conjunction with other methods, to generate and validate explanations about social and political behavior. The primary goal of studying what did not happen is to better help us to understand what did happen.

## 1 THE IMPORTANCE OF COUNTERFACTUALS

I follow Tetlock and Belkin (1996, 4) and define counterfactual as a “subjective conditional in which the antecedent is known or supposed for purposes of argument to be false.” It is a “contrary-to-fact” conditional that identifies a “possible” or “alternative” world in which the antecedent did not actually occur.

All causal statements imply some kind of counterfactual. A historical argument that a particular set of conditions, processes, and events caused, influenced, or contributed to a subsequent set of conditions, processes, and events implies that if the antecedent conditions had been different, the outcome would have been different. Similarly, a theoretical statement that  $x$  is a cause of  $y$  implies that if the value of  $x$  were different, the outcome  $y$  would be different.<sup>2</sup> The interpretative statement that British and French appeasement of Hitler contributed to the Second World War implies that if Britain and France had stood firm against Hitler the Second World War would not have occurred, or perhaps that it would have been considerably shorter and less destructive. The theoretical proposition that appeasement only encourages further aggression implies the counterfactual that a more hard-line strategy would reduce the likelihood of further aggression.

While all causal statements imply a counterfactual, some counterfactuals are more explicit than others. Historical interpretations and theoretical propositions that posit necessary conditions, which are fairly common in political science and in social science more generally (Goertz and Starr 2003), are particularly explicit about their counterfactual implications. The logical expression of necessary conditions—“if not  $x$  then not  $y$ ”—directly specifies the consequent of a counterfactual world. Necessary condition counterfactuals are central in “window of opportunity” models (Kingdon 1984), “powder keg” models (Goertz and Levy 2007),

<sup>2</sup> This is particularly clear if one says that  $c$  is a cause of  $e$  if the conditional probability that  $e$  occurs, given that  $c$  occurs, is greater than the unconditional probability that  $e$  occurs. For a discussion of statistical approaches to counterfactual analysis see Morgan and Winship (2007).

causal chains involving necessary conditions, and explanations based on path dependency and critical junctures (Mahoney 2000), all of which are common in political science.

Necessary conditions are also important because they are at the core of one of the two leading conceptualizations of causation:  $x$  is a cause of  $y$  if  $y$  would not occur in the absence of  $x$  (Lewis 1973).<sup>3</sup> The central role of necessary conditions in causal explanation, particularly in case studies, and the fact that the primary way of testing hypotheses involving necessary conditions is to analyze the validity of the counterfactual associated with it, add significantly to the importance of developing a valid methodology of counterfactual analysis.<sup>4</sup>

Counterfactuals play an essential role in common historiographical debates about whether a particular outcome was inevitable or contingent. The most effective way of supporting an argument that an outcome was contingent is to demonstrate that a slight change could easily have led to a different outcome. An effective way to demonstrate that an outcome was inevitable at a certain point is to demonstrate that there was no change in existing conditions that was both conceivable and capable of leading to a different outcome.

Some theoretical approaches are more explicit than others about their counterfactual implications. A good example is game theory. A game tree specifies exactly what happens if actors make different choices, how other actors react, and the set of possible outcomes. Actors' choices are causally dependent on their expectations of what would happen if they made other choices. "Off the equilibrium path" behavior is a counterfactual prediction.

Since causal statements—whether about particular historical events or about theoretical relationships between variables—imply counterfactuals, the validity of the counterfactual bears on the validity of the original causal proposition. Compelling evidence that Hitler would have initiated a war regardless of whether Britain or France had pursued a strategy of appeasement would falsify the claim that appeasement causally contributed to the Second World War.<sup>5</sup> Evidence that social revolutions sometimes occur in the absence of a peasant revolt and a state crisis would disconfirm Skocpol's (1979) theory of social revolution.

Thus the empirical validation of counterfactual statements is an important step in hypothesis testing. Evidence from several case studies is generally more persuasive than evidence from a single case study in testing a theoretical proposition

<sup>3</sup> The alternative conception of causation is the "regularity" model, which includes constant conjunction, temporal precedence, and nonspuriousness. This view is often traced to Hume, but in fact Hume emphasized both conceptions of causation (Goertz and Levy 2007).

<sup>4</sup> While all necessary conditions imply a specific counterfactual, not all counterfactuals imply the presence of necessary conditions. The statement "if not  $x$ , then  $y$  will still occur" posits a counterfactual, but without a necessary condition. Also, unlike statements of necessary conditions, statements of sufficient conditions do not imply a counterfactual, since "if  $x$  then  $y$ " says nothing about the consequences of "not  $x$ ." The widely accepted empirical generalization that a democratic dyad is sufficient for peace implies no counterfactual about what would happen if at least one of two states is not a democracy.

<sup>5</sup> Khong (1996) explores the complexities of validating this counterfactual. Ripsman and Levy (2007) argue that neither British nor French leaders expected that appeasement would avoid war.

(and quantitative researchers argue that statistical evidence is far superior to multiple case studies), but single case studies can also serve this purpose if the hypothesis posits necessary or sufficient conditions, if the hypothesis generates precise predictions, or if the proposition permits a most likely or least likely research design (George and Bennett 2005; Levy 2008).

This point relates to a larger principle about theory construction and research design. The widely accepted injunction to derive as many observable implications as possible from an historical interpretation or theoretical proposition, and to test them against the evidence (King, Keohane, and Verba 1994), applies to a theory's counterfactual implications as well as to its more direct implications.<sup>6</sup> *Ceteris paribus*, the more explicit the counterfactual implications of a theory, the better the theory. A theory that specifies the consequences of both  $x$  and not  $x$  tells us more about the empirical world than a theory that specifies only the consequences of  $x$ .

The preceding discussion emphasizes the utility of counterfactuals in testing both interpretations of individual cases and more general theoretical propositions. The first is more idiographic in orientation, and the second more nomothetic. Interpretations of individual cases can be either inductive or guided by theory.<sup>7</sup> Although theory-guided case studies are presumably more explicit about their counterfactual predictions, at least at the beginning of a research study,<sup>8</sup> inductively driven interpretive case studies can also end up suggesting what would have happened if certain things had been different. Ferguson's (1999b) argument about what would have happened had Britain "stood aside" in the First World War (probably no Bolshevik Revolution, no Second World War, no Holocaust) is based on an inductive case study.

In idiographic case studies, the role of counterfactual analysis is often to explore the question of whether history could have turned out differently, and how alternative worlds might have developed. Such "what if" scenarios include both relative short-term forecasts, such as the consequences of a failed assassination attempt in 1914, and long-term macrohistorical forecasts, such as the plausibility of alternative scenarios through which the rise of the West to a position of world dominance might have been blocked (Tetlock, Lebow, and Parker 2006).

Idiographic counterfactuals can also play a normative role—of passing moral judgment on individual leaders by asking whether or not they could have acted differently under the circumstances (Tetlock and Belkin 1996, 8). Presumably we could not blame a leader for a costly war if a counterfactual analysis were to provide persuasive evidence that the same war would have occurred if another leader had been in power. Neville Chamberlain is widely criticized for his policy of appeasing Hitler, but that

<sup>6</sup> Qualitative researchers qualify this argument by de-emphasizing the "as many" phrase and arguing that some observable implications are more important than others for theory testing.

<sup>7</sup> Scholars often mischaracterize the idiographic/nomothetic distinction. Idiographic refers to descriptions or explanations of the particular, whereas nomothetic refers to the construction or testing of general theoretical propositions (Levy 2001).

<sup>8</sup> In the analytic narrative research program (Bates et al. 1998), for example, game-theoretic models are used to guide individual case studies.

judgment rests on the counterfactual argument that another British leader would have acted differently and that the outcome would have been better for Britain, which some scholars have questioned.

Exploring a theory's counterfactual implications can also be useful for theory development. This is a deductive analysis that does not involve empirical case studies.<sup>9</sup> As Tetlock and Belkin (1996) note, "The goal is not historical understanding... [but] to pursue the logical implications of a theoretical framework." For this purpose, the counterfactual conditional itself need not necessarily be plausible, in the sense that we could imagine a path through which it might arise. Such nonplausible counterfactuals are often called "miracle counterfactuals" (Fearon 1996).

Any complete theory specifies the consequences if key causal variables were to take on other values. Economic models specify the consequences for the economy if the Federal Reserve Board were to raise interest rates by a full percent, even if such actions are quite implausible under the circumstances. Computer simulations are used to explore the consequences of possible worlds (including socially complex worlds) that are not really accessible through empirical methods (Cederman 1996). Similar models can be applied historically to trace the consequences of a particular set of initial conditions, however implausible they might be. To assess the contribution of the railroads to American economic growth, Fogel (1964) constructed a model of the American economy and explored how it might have developed in the absence of railroads.

Counterfactual thought experiments can also be used deductively for other purposes, including the pedagogical purposes of encouraging someone to think through the implications of his or her arguments or beliefs, confront uncomfortable arguments, and generally open his or her mind to new ideas. This may help reveal double standards in moral judgment, contradictions in causal beliefs, and the influence of cognitive or motivational biases. This mode of counterfactual analysis combines the descriptive and the normative, in that descriptive/causal knowledge about how people use counterfactuals in making judgments and decisions is used for the normative purpose of inducing them to think in less biased ways about causality and counterfactuals (Tetlock and Belkin 1996, 12–16; Weber 1996; Lebow 2000).

## 2 CRITERIA FOR EVALUATION

We now turn to the criteria by which we might evaluate the validity of counterfactual arguments in explaining cases or testing theoretical propositions. What kinds of counterfactuals are more legitimate than others for these purposes, recognizing

<sup>9</sup> In mathematics, an "indirect proof" or "proof by contradiction" assumes the theorem is false, traces the logical consequences, reveals a contradiction, and concludes that the theorem is true.

that different standards might be appropriate for different theoretical and descriptive purposes?<sup>10</sup>

Counterfactual propositions are similar to other theoretical propositions in that they have premises or initial conditions, hypothesized consequences, and a "covering law" or causal mechanisms explaining how the former leads to the latter. A theoretical proposition must be logically consistent and falsifiable in principle. We generally prefer theories that apply to broader empirical domains, make more (and more varied) predictions about the empirical world, generate substantial support from the empirical evidence, and are consistent with other well-accepted theories (Hempel 1966, ch. 4). Criteria for evaluating counterfactual arguments should be grounded in these widely accepted standards of theory evaluation. These are best organized around the categories of the clarity of the antecedents and consequents, the plausibility of the antecedent, and the conditional probability of the consequent given the antecedent.<sup>11</sup>

### 2.1 Clarity

Counterfactual arguments must include well-specified antecedents and consequents. The consequent should be more specific than "the outcome would have been different," which is only moderately helpful. Unless we specify *how* it would have been different, the statement is nearly nonfalsifiable and hence not particularly useful. For a counterfactual to be scientifically useful, the consequent must be clearly specified by the analyst, not left to the imagination of the reader.

Necessary condition counterfactuals are quite explicit, since they posit an antecedent of  $\sim x$  and a consequent of  $\sim y$ . To say that war would not have occurred in the absence of an assassination is a powerful statement. It would be more discriminating, and in many ways more useful, to specify whether the absence of war meant peace for several years or peace punctuated by ongoing crises and a high risk of war in the future. With respect to the antecedent, it might have made a difference, in terms of the likely response, if the failed assassination attempt had been discovered or not. Non-necessary condition counterfactuals can also be clear. The statement "if Hitler had been ousted in a coup, the Second World War would still have occurred" has well-specified antecedents and consequents.

While clarity is good, there is a trade-off between the specificity of the consequent and the probability of its occurrence. The more detailed a consequent, the more likely it is to be false. The probability of multiple outcomes is the product of their individual probabilities, and the probability of a single outcome is the product of the individual probabilities of each of the steps leading to it (Lebow 2000, 583). This suggests that

<sup>10</sup> The main difference between empirically oriented and deductively oriented counterfactuals is that the latter are not constrained by need for the antecedent to be plausible.

<sup>11</sup> This differs somewhat from the Tetlock and Belkin (1996, 16–31) and Lebow (2000, 577–85) conceptualizations.

counterfactuals that are either too detailed or the result of a long causal chain are likely to be false. Thus at the two extremes counterfactuals are either nonfalsifiable or false. Neither is particularly useful.<sup>12</sup> The best counterfactuals have specific but not too specific antecedents.

Relatedly, arguments suggesting that a particular counterfactual antecedent will lead to a particular consequent with a high degree of certainty are quite suspect. It would be problematic to invoke a counterfactual to demonstrate that history did not have to happen the way it did, only to predict a counterfactual consequent that deterministically flowed from the antecedent, since that would be supporting a statement of contingency with an argument based on determinism.

## 2.2 Plausibility of the Antecedent

It is widely agreed that for counterfactual arguments to be useful in exploring whether history might have turned out differently, the antecedent must be plausible or realistic as well as well specified. One must be able to imagine how that antecedent might arise. It is not useful to say that 1960s America would have been different if Abraham Lincoln had been president.

For the purposes of assessing causality, counterfactual analysis has the same general task as experimental, statistical, and comparative methods: to organize evidence to show that a change in the value of an outcome variable can be traced to the effects of a change in a causal variable, and not to changes in other variables. Just as experimental research manipulates one variable at a time in a controlled setting, and comparative research tries to select matched cases in which covariation with the outcome variable is limited to a single causal variable, counterfactual analysis ideally posits an alternative world that is identical to the real world in all theoretically relevant respects but one, in order to explore the consequences of that difference.

That is easier said than done, and the same problems that plague comparative research constrain counterfactual analysis as well: In a system of interconnected behavior, a change in one variable reverberates through the system and induces changes in many other variables, so that "we can never do merely one thing" (Hardin, cited in Jervis 1997, 10). In counterfactual thought experiments, as in a matching cases strategy, it is often quite difficult to hold everything else equal. Hence Lebow (2000) questions the utility of "surgical" counterfactuals.

An attempt to assess the impact of US nuclear superiority on the outcome of the Cuban Missile Crisis by imagining the crisis under conditions of Soviet strategic superiority would have to change too much history to be useful. Soviet superiority would be conceivable only if the US economy and technological capacity had been much weaker, if American society did not support a competitive military establishment, and so on. The presence of these other conditions would almost certainly

<sup>12</sup> Admittedly, an empirical confirmation of any detailed counterfactual prediction would be particularly compelling precisely because the *ex ante* probability was so low (Popper 1965).

have changed the world in other ways as well (the status of Berlin, for example), further complicating any effort to say that it was Soviet superiority, rather than these other changes, that caused the outcome. Moreover, a Soviet Union with strategic superiority would have had no incentive to put offensive missiles in Cuba in the first place.

As this example suggests, a counterfactual conditional is not complete in itself. It relies on other changes to sustain it. Goodman (1983) called these *connecting principles*, which themselves involve counterfactual propositions about the consequences of those other changes. Any counterfactual analysis needs to specify the secondary counterfactuals, or "enabling counterfactuals" (Lebow 2000), that must be introduced to sustain the primary counterfactual.

A good counterfactual requires that these connecting principles and enabling counterfactuals be specified with reasonable precision (like the primary counterfactual) and that they be consistent with each other and with the antecedent of the primary counterfactual.<sup>13</sup> Goodman (1983, 15) refers to this requirement of logical consistency as *cotenability*.

The counterfactual of Soviet strategic superiority in 1962 fails to satisfy the cotenability criterion. So does the argument that if Nixon had been president during the Cuban Missile Crisis, he would have ordered an air strike rather than a naval blockade. Lebow and Stein (1996) argue persuasively that if Nixon had been president he, unlike Kennedy, would have probably authorized the use of US forces in the Bay of Pigs operation, Castro would have been overthrown, and the Soviets would not have put offensive missiles in Cuba.

In his attempt to assess the likely development of the American economy in the absence of railroads, Fogel (1964) identified other developments that were likely to occur in the absence of railroads, including the introduction of the internal combustion engine and the automobile (along with its demands for iron and other materials) fifty years before its actual appearance. Elster (1978, 204–8) basically argues that this assumption of the early emergence of the automobile is not cotenable with the assumption of the delay of the railroad, since the technology upon which the automobile was based surely would have led to the railroad.<sup>14</sup>

For these reasons, most analysts accept Max Weber's (1949) argument that for the purposes of causal analysis the best counterfactual worlds to examine are those that require as few changes as possible in the real world. This is the "minimal rewrite of history" rule (Tetlock and Belkin 1996), which is defined in terms of the magnitude of the changes as well as their number.<sup>15</sup>

<sup>13</sup> Tetlock and Belkin (1996, 21) include consistency with the consequent, which I discuss in the next section on the conditional probability of the consequent given the antecedent.

<sup>14</sup> See also Tetlock and Belkin (1996, 22) and Lebow (2000, 582).

<sup>15</sup> Similarly, King and Zeng (2005) use statistical modeling to demonstrate that whereas the validity of counterfactual propositions positing relatively modest changes in the real world can be evaluated based on the data, "extreme counterfactuals" are quite sensitive to assumptions built into the model that have little to do with the data.

One example of a minimal rewrite counterfactual is the proposition that if George W. Bush had not won the 2000 election, the United States would not have gone to war in Iraq. This counterfactual involves a minimal rewrite of history because one does not have to change much—just a modest number of Florida voters reading their ballots correctly. So the counterfactual's antecedent is quite plausible.<sup>16</sup> The hypothesized consequent (President Al Gore not invading Iraq) is quite plausible but not certain, and the argument would have to include enabling counterfactuals such as how Gore would have reacted to 9/11.

Lebow's (2007) argument that if the assassination attempt against the Archduke had failed the First World War probably would not have occurred is another highly plausible, minimal rewrite counterfactual. The antecedent is quite easy to imagine, as Lebow explains in great detail. Indeed the *ex ante* probability of the alternative world was undoubtedly higher than that of the real world, which is another possible criterion for the evaluation of counterfactuals.<sup>17</sup> Lebow's analysis is particularly useful as an illustration of the value of specifying the *conditions and processes* through which one could get from the real world to the counterfactual world. Another good example of this is Mueller's (1991) detailed assessment of the identity and policy preferences of other possible German leaders who might have held the position of chancellor, to support his argument that in the absence of Hitler the Second World War would not have occurred.

Scholars will disagree, of course, on how minimal a rewrite of history has to be for the antecedent to be considered plausible. There is no single answer to this question, as there may be a trade-off between maximizing the plausibility of a counterfactual by minimizing the number of additional conditions necessary to sustain it, and selecting historically or theoretically meaningful counterfactuals. As the historian Schroeder (2007, 149–50) argues, “a major counterfactual . . . will change too much, and a minor one too little, to help us explain what really did happen and why, and why alternative scenarios failed to emerge.”<sup>18</sup> Still, counterfactuals relating to the 1914 assassination and the 2000 American election demonstrate that minimal rewrite counterfactuals can be consequential.

Ferguson (1999a, 86) offers his own answer of how we distinguish plausible and implausible counterfactuals: “We should consider as plausible or probable only those alternatives which we can show on the basis of contemporary evidence that contemporaries actually considered.” Ferguson further narrows the range of acceptable counterfactuals by restricting them to “hypothetical scenarios which contemporaries not only considered, but also committed to paper.”

The second criterion is unnecessarily restrictive. While paper records of actors' alternative options might provide particularly compelling evidence of choices not made, we should not exclude evidence based on oral interviews with first-hand

<sup>16</sup> By contrast, counterfactual claims about the consequences of George W. Bush not winning the 2004 election posit a more problematic antecedent.

<sup>17</sup> The validity of the hypothesized consequent is another, analytically distinct question which I return to later.

<sup>18</sup> See also Weber (1996) and Lebow (2000).

observers. Nor should we be precluded from exploring alternative histories in repressive political systems in which actors are afraid to commit their thoughts to paper or among nonelite groups for which there are no written records.

Even Ferguson's (1999a) first criterion might be too restrictive. It would exclude alternative choices that actors failed to consider, or failed to consider seriously, because of psychological biases, political constraints, or failure of imagination—alternative possibilities that other decision-makers, had they been in office, might have considered (assuming a change in personnel was a plausible antecedent). It is not clear, for example, whether Ferguson's criterion would allow us to consider the consequences of the US responding to the Soviet missiles in Cuba by doing nothing, since they gave that option little or no serious consideration. In addition, some options are formally considered only for political reasons but given no serious thought, as Janis (1982) suggests for many “devil's advocate” arguments.

Schroeder (2007, 151–2) agrees that the counterfactuals actually considered by the actors themselves are important, but recognizes that they are too restrictive. He argues that the historian needs to pose counterfactual questions of his own: “What other decisions and actions could the historical actors have made under the existing circumstances? To what extent did they recognize and consider these? What circumstances made these choices or alternative courses genuinely possible or merely specious and actually unreal? What might the alternative results of these choices have been?” These are precisely the right questions to ask, but even they might be too restrictive. They fail to give enough emphasis to counterfactual possibilities introduced by external events such as assassinations, battles or elections won or lost, fatal illness, personal tragedies, and other highly contingent events (Lebow 2000, 560).<sup>19</sup>

Game theory provides a still more restrictive set of criteria for the selection of acceptable counterfactuals from the enormous number of possible counterfactuals. First of all, game-theoretic models in extensive form specify exactly what will happen if actors were to make different decisions at various choice nodes. All sequences of choices “off the equilibrium path” are possible counterfactuals. But actors did not go down those paths for a reason. In a large class of games, however, there are multiple equilibria, multiple ways all actors could have behaved that were fully consistent with their interests given the constraints of the game. Each of these paths constitutes a fully acceptable counterfactual (Bueno de Mesquita 1996; Weingast 1996).<sup>20</sup>

The consideration of the alternatives that actors did consider or might have considered within a choice-theoretic framework raises an interesting asymmetry in counterfactual analysis. Both events and nonevents generate counterfactuals, but it is often easier to examine event-generated counterfactuals than those generated

<sup>19</sup> The outcomes of a small percentage of battles hinge on accidents, luck, insubordination, unexpected weather, and other contingencies, and the long-term political consequences of such reversals of outcome can be profound (Cowley 1999).

<sup>20</sup> The existence of alternative subgame perfect equilibria provides an excellent criterion for good counterfactuals where the specification of the game provides a reasonable fit to the situations faced by actors and choices they consider. We need to remember, however, that the simplifications necessary to create tractable games, including limits on the number of choices available to actors, may be too restrictive in some situations, including those involving enormous social complexity.

by nonevents. We can explore the consequences of a counterfactual failure of the assassination plot in 1914, since there is ample evidence from pre-assassination records of how Austrian and German leaders defined their decision problems and options, and thus how they might have acted. In the absence of an assassination, however, it would be far more difficult to explore the counterfactual possibility of an assassination of the Archduke, since presumably no one at the time gave serious thought to that possibility and what they would do if it happened, and we would have to introduce more secondary counterfactuals that would be somewhat speculative. Similarly, it is easier (though not easy) to explore counterfactuals created by the hypothetical failure of the assassination attempt against Kennedy than those that would have been created by the hypothetical success of the assassination attempt against Reagan. The *ex ante* probability, and hence counterfactual plausibility, of each of these antecedents is quite different.

If the 9/11 plot had failed and gone undetected, and if the US had not invaded Iraq (which is quite plausible but not certain), it would be quite difficult to undertake a counterfactual analysis, set in the absence of 9/11, of the possibility that a major terror attack against the United States originating from Afghanistan might trigger an American war against Iraq in 2003. In a world in which 9/11 (or something like it) had not occurred in 2001 (or soon thereafter), would any of our criteria for evaluating counterfactuals permit the characterization of a 9/11 attack as a plausible antecedent and an American invasion of Iraq as a likely consequent?<sup>21</sup>

### 2.3 The Conditional Plausibility of the Consequent

Thus far we have emphasized that the characteristics of a useful counterfactual include a well-specified antecedent and consequent and an antecedent or conditional that is plausible, involves a minimal rewrite of history, and is sustainable through conditions that are cotenable with each other and with the antecedent. The next question is whether the antecedent, along with the conditions that are necessary to support it, is likely to lead to the hypothesized consequent. The basic requirement, like those for any theoretical proposition, is that the causal linkages be clearly specified, logically complete, and consistent with the empirical evidence. It is admittedly difficult to separate theoretical and empirical criteria, since social science research is ideally characterized by an ongoing dialogue between theory and evidence (Levy 2007), but listing them separately is useful for our purposes here.

In many ways, the most important requirement for a good counterfactual, besides a plausible antecedent, is a good theory. As Fogel (1964, cited in Tetlock and Belkin 1996, 26) writes, "Counterfactual propositions... are merely inferences from hypothetico-deductive models." The greater the extent to which the hypothesized

<sup>21</sup> For an intriguing counterfactual analysis set in a counterfactual world, but one that involves a much more extended timeframe than the one posited here, see Lebow (2006). Imagining a world in which Mozart lived to sixty-five and in which neither of the world wars of the twentieth century occurred, Lebow considers the possibility that an early death of Mozart triggered a chain of historical happenings that led to these events, and considers the critiques and defenses of such a counterfactual.

causal mechanisms leading from a specific antecedent to a particular consequent are consistent with well-established and empirically confirmed theories, the greater the plausibility of the counterfactual. The better the theory—defined in terms of its logical coherence, precision, deductive support from well-established theories, and empirical support—and the more it makes explicit predictions about what will happen under a variety of counterfactual conditions, the better the counterfactual.

For the purposes of evaluating a particular counterfactual proposition, we are more interested in the predictive power of a relevant theory under the specific conditions defined by the antecedent than in its predictive power under a wide range of conditions, although the latter increases our confidence in the former. For this reason, a well-established empirical law, even in the absence of consensus regarding how to explain it theoretically, would also provide useful support for a counterfactual proposition. Admittedly, we have relatively few empirical laws in political science, but some propositions have far more empirical support than others.

Thus *consistency with the empirical evidence* is another criterion for a good counterfactual proposition. Tetlock and Belkin (1996, 27–30) include "consistency with well-established statistical generalizations" as one of their six criteria of a good counterfactual. I would construe evidence more broadly. While statistical (and experimental) evidence is highly desirable—assuming a sufficient number of comparable cases, variables that are operationalizable and measurable over those cases, accessible data, etc.—such evidence is not always available. Comparative studies and, less frequently, single case studies (if based on a compelling most or least likely design) can sometimes provide an adequate evidentiary basis to support a counterfactual proposition, and such evidence, when combined with statistical evidence, can often provide substantially better support than statistical evidence alone (George and Bennett 2005).

Although "direct" support for a specific counterfactual proposition is important, our confidence in its validity is enhanced if the counterfactual generates other theoretical predictions that are also supported. A counterfactual proposition has implications not only for the final outcome, but also for the intervening paths between antecedent and consequent, and those intervening predictions should be specified in advance and tested if at all possible. This is the criterion of *projectibility* (Tetlock and Belkin 1996, 30–1).

Let me return to the point I made above about the likelihood of a counterfactual being a function of the length of the causal chain leading to it. This is just the commonsense notion that short-term predictions are more plausible than long-term predictions, which applies to counterfactual propositions as well as to those with factual antecedents. Even where one finds short-term regularities, the compounding of small uncertainties generates enormous irregularity and unpredictability over the long term. Even deterministic processes can generate highly unpredictable outcomes if they are sensitive to initial conditions, as chaos theory demonstrates.

Fearon (1996) demonstrates this in a compelling way in his discussion of an extremely simple process involving cellular automata that follow rules of behavior that are well understood, precisely specified, but stochastic, with well-defined probabilities of moving from one state to another. He demonstrates how such processes generate local regularities but global unpredictability. Fearon's (1996) analysis leads

him to suggest a *proximity* criterion, so that we can assess plausibility “only where the counterfactuals involve causal mechanisms and regularities that are well understood and that are considered at a spatial and temporal range small enough that multiple mechanisms do not interact, yielding chaos” (p. 66).

A related reason for insisting on a proximity criterion is that even if the antecedent is plausible and there are good theoretical and empirical reasons to believe that the presence of the antecedent would lead to predicted consequents with fairly high probability, it is always possible that subsequent developments might return history to its original course, before it was diverted by the hypothesized antecedent. Lebow (2000, 584) labels these “second-order counterfactuals,” but I prefer the term *redirecting* counterfactuals, since primary counterfactuals generate numerous secondary counterfactuals and only some of them reroute history back to its original course.

Consider the argument that the Vietcong attack on the US military base at Pleiku in 1965 was a significant cause of US escalation in the Vietnam War. In the absence of that Vietcong attack, however, it is quite plausible to argue that another incident would have occurred, whether randomly or deliberately instigated for strategic advantage, and led to an American escalation. As Bundy argued, “Pleikus are like streetcars.” Wait long enough and one will come along.<sup>22</sup> This remark is often used in support of structural arguments and against arguments about the importance of contingency in historical processes.

Another example of a redirecting counterfactual arises with respect to Lebow’s (2007) own counterfactual argument that in the absence of the assassination of the Archduke the First World War probably would not have occurred. Lebow argues that in the absence of the assassination, existing trends in the balance of military power that favored Russia would have forced Moltke (or his successor as Chief of the German General Staff) to abandon the Schlieffen Plan and adopt instead a more defensive strategy, which would have eliminated the incentives for preventive or preemptive military action in any crisis that might arise.

This counterfactual is clear and the antecedent constitutes a minimal rewrite. The causal linkages to the consequent are not plausible, however, because they would have probably generated self-refuting strategic behavior. Lebow’s assessment of military trends is right on the mark. German military and political leaders would almost certainly have accepted his vision of a 1917 world, and precisely for that reason they never would have allowed that world to come about. The same fears for the future that generated Germany’s preventive motivation for war in 1914 (which Lebow implicitly acknowledges) would have led German leaders to initiate or provoke the preventive war they thought they needed before Russia grew too strong (Fischer 1967). This critique, of course, generates its own counterfactual, which would need to be evaluated against Lebow’s counterfactual. Fortunately, the expansive literature on the First World War provides enough evidence to resolve this debate, even if not with complete certainty.

<sup>22</sup> National Archives and Records Administration, Lyndon Baines Johnson Library, oral interviews of Frederick W. Flott, <[http://webstorage1.mcpa.virginia.edu/library/mc/poh/transcripts/flott\\_frederick\\_1984\\_0927.pdf](http://webstorage1.mcpa.virginia.edu/library/mc/poh/transcripts/flott_frederick_1984_0927.pdf)>.

### 3 CONCLUSION

All causal statements generate counterfactuals about what would happen if certain variables were to take on different values, and all nonexperimental methodologies must deal with this in one way or the other. I focus here on the role of counterfactuals in case studies. I argue that if counterfactuals are made explicit and used according to scientifically acceptable rules of inference, a study of history as it really wasn’t can help us understand history as it really was (to borrow from Ferguson 1999a).

Counterfactuals serve different theoretical and descriptive purposes. Different theoretical goals and normative values call for different trade-offs among various research objectives, and consequently there is no single set of methodological criteria applicable to all counterfactuals. Leaving aside the use of counterfactual thought experiments to stimulate the imagination, which is useful but which follows a different set of rules, I suggest criteria for the evaluation of counterfactuals for the purposes of explaining historical cases or using cases to assess more general theoretical propositions.

The analyst should clearly specify a counterfactual’s antecedent, consequent, and the causal linkages between them. Counterfactuals should change as few aspects of the real world as possible in order to isolate their causal effects (the “minimal-rewrite-of-history” rule). The analyst should specify both the “supporting conditions” or “connecting principles” that are needed to sustain the primary counterfactual and the secondary counterfactuals that lead from the antecedent (and its supporting conditions) to the consequent. The consequent should be more specific than “the outcome would have been different,” but not so specific that it becomes implausible, given the fact that the probability of a highly specific outcome is far less than the probability of a range of outcomes.

Counterfactual analysis, perhaps even more so than other kinds of analysis, is a theory-driven process. We cannot directly trace the consequences of an unobservable antecedent, so we must rely on theoretical knowledge. The stronger the theory, and the more the analyst can resort to theory or empirical laws to justify his or her hypothesized causal mechanisms linking the antecedent to the consequent, the better the counterfactual. Counterfactuals that generate additional observable implications and that can themselves be “tested” against the evidence provide additional confidence in the validity of the counterfactual. This is the “projectibility” criterion.

Logical consistency is another important criterion. The supporting conditions for a primary factual must be consistent with each other and with the antecedent. Their linkages to the consequent must also be theoretically sensible. It does little good to posit an antecedent that could only lead to the consequent if supported by connecting principles and secondary counterfactuals that were themselves inconsistent with the antecedent. Researchers should be particularly alert to strategic behavior that might return history to its original path (“redirecting counterfactuals”)—where actors recognize that the hypothesized consequences of the antecedent are both accurate and undesirable, and act to head off those outcomes.



It would also be inconsistent to try to support an argument that a particular outcome was contingent by demonstrating that a small change in the situation would have invariably led to a different outcome—one cannot support an argument for contingency by invoking a deterministic process. To the extent possible, based on existing theory and empirical evidence, the author should give a sense of the likelihood that the consequent would have followed from the antecedent.

Counterfactual analysis, like any causal analysis, must recognize that uncertainty about initial conditions and the pervasiveness of stochastic behavior limits even the best of our theories to relatively short-term predictions. After that, too many things interact in too many unpredictable ways. The longer the causal chain, the lower the probability of occurrence of the outcome at the end of the chain, particularly given the relative absence of sufficient conditions in social theory. This suggests temporal and spatial “proximity” as an additional criterion for evaluating counterfactuals.

While the application of these criteria might make “after dinner history” somewhat less entertaining, it disciplines the use of counterfactuals and provides an additional methodological tool for evaluating causality in a nonexperimental world in which many confounding variables interact in unpredictable ways.

## REFERENCES

- BATES, R., GREIF, A., LEVI, M., ROSENTHAL, J.-L., and WEINGAST, B. 1998. *Analytic Narratives*. Princeton, NJ: Princeton University Press.
- BUENO DE MESQUITA, B. 1996. Counterfactuals and international affairs: some insights from game theory. Pp. 211–29 in *Counterfactual Thought Experiments in World Politics*, ed. P. E. Tetlock and A. Belkin. Princeton, NJ: Princeton University Press.
- CARR, E. H. 1964. *What is History?* Harmondsworth: Penguin.
- CEDERMAN, L.-E. 1996. Rerunning history: counterfactual simulation in world politics. Pp. 247–67 in *Counterfactual Thought Experiments in World Politics*, ed. P. E. Tetlock and A. Belkin. Princeton, NJ: Princeton University Press.
- COWLEY, R. 1999. *What If?* New York: G. P. Putnam's Sons.
- ELSTER, J. 1978. *Logic and Society*. New York: John Wiley and Sons.
- ENGLISH, R. 2007. Perestroika without politics: how realism misunderstands the Cold War's end. Pp. 237–60 in *Explaining War and Peace: Case Studies and Necessary Condition Counterfactuals*, ed. G. Goertz and J. S. Levy. New York: Routledge.
- FEARON, J. D. 1996. Causes and counterfactuals in social science: exploring an analogy between cellular automata and historical processes. Pp. 39–67 in *Counterfactual Thought Experiments in World Politics*, ed. P. E. Tetlock and A. Belkin. Princeton, NJ: Princeton University Press.
- FERGUSON, N. 1999a. Virtual history: toward a “chaotic” theory of the past. Pp. 1–90 in *Virtual History: Alternatives and Counterfactuals*, ed. N. Ferguson. New York: Basic Books.
- 1999b. *The Pity of War: Explaining World War I*. New York: Basic Books.
- FISCHER, D. 1971. *Historians' Fallacies: Toward a Logic of Historical Thought*. London: Routledge and Kegan Paul.
- FISCHER, F. 1967. *Germany's Aims in the First World War*. New York: Norton.
- FOGEL, R. 1964. *Railroads and American Economic Growth: Essays in Econometric History*. Baltimore: Johns Hopkins University Press.
- GEORGE, A. L., and BENNETT, A. 2005. *Case Studies and Theory Development in the Social Sciences*. Cambridge, Mass.: MIT Press.
- GOERTZ, G., and LEVY, J. S. 2007. Causal explanation, necessary conditions, and case studies. Pp. 9–45 in *Explaining War and Peace: Case Studies and Necessary Condition Counterfactuals*, ed. G. Goertz and J. S. Levy. New York: Routledge.
- and STARR, H. (eds.) 2003. *Necessary Conditions: Theory, Methodology, and Applications*. Lanham, Md.: Rowman and Littlefield.
- GOODMAN, N. 1983. *Fact, Fiction, and Forecast*, 4th edn. Cambridge, Mass.: Harvard University Press.
- HEMPEL, C. G. 1966. *Philosophy of Natural Science*. Englewood Cliffs, NJ: Prentice Hall.
- JANIS, I. L. 1982. *Groupthink*, 2nd rev. edn. Boston: Houghton Mifflin.
- JERVIS, R. 1997. *System Effects*. Princeton, NJ: Princeton University Press.
- KHONG, Y. F. 1996. Confronting Hitler and its consequences. Pp. 95–118 in *Counterfactual Thought Experiments in World Politics*, ed. P. E. Tetlock and A. Belkin. Princeton, NJ: Princeton University Press.
- KING, G., KEOHANE, R., and VERBA, S. 1994. *Designing Social Inquiry*. Princeton, NJ: Princeton University Press.
- and ZENG, L. 2006. The dangers of extreme counterfactuals. *Political Analysis*, 14: 131–59.
- KINGDON, J. 1984. *Agendas, Alternatives, and Public Policies*. Boston: Little, Brown.
- LEBOW, R. N. 2000. What's so different about a counterfactual? *World Politics*, 52: 550–85.
- 2006. If Mozart had died at your age: psychological versus statistical inference. *Political Psychology*, 27: 157–72.
- 2007. Contingency, catalysts, and nonlinear change: the origins of World War I. Pp. 85–111 in *Explaining War and Peace: Case Studies and Necessary Condition Counterfactuals*, ed. G. Goertz and J. S. Levy. New York: Routledge.
- and STEIN, J. G. 1996. Back to the past: counterfactuals and the Cuban missile crisis. Pp. 119–48 in *Counterfactual Thought Experiments in World Politics*, ed. P. E. Tetlock and A. Belkin. Princeton, NJ: Princeton University Press.
- LEVY, J. S. 2001. Explaining events and testing theories: history, political science, and the analysis of international relations. Pp. 39–83 in *Bridges and Boundaries: Historians, Political Scientists, and the Study of International Relations*, ed. C. Elman and M. F. Elman. Cambridge, Mass.: MIT Press.
- 2007. Theory, evidence, and politics in the evolution of research programs. Pp. 177–97 in *Theory and Evidence in Comparative Politics and International Relations*, ed. R. N. Lebow and M. Lichbach. New York: Palgrave Macmillan.
- 2008. Case studies: types, designs, and logics of inference. *Conflict Management and Peace Science*, 25: 1–18.
- LEWIS, D. 1973. *Counterfactuals*. Cambridge, Mass.: Harvard University Press.
- MAHONEY, J. 2000. Path dependence in historical sociology. *Theory and Society*, 29: 507–48.
- MORGAN, S. L., and WINSHIP, C. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. New York: Cambridge University Press.
- MUELLER, J. 1991. Changing attitudes towards war: the impact of the First World War. *British Journal of Political Science*, 21: 1–28.
- OAKESHOTT, M. 1966. *Experience and its Modes*. London: Cambridge University Press.
- POPPER, K. 1965. *The Logic of Scientific Discovery*. New York: Harper Torchbacks.
- RIPSMAN, N. M., and LEVY, J. S. 2007. The preventive war that never happened: Britain, France, and the rise of Germany in the 1930s. *Security Studies*, 16: 32–67.

- ROESE, N. J., and OLSON, J. M. 1995. *What Might Have Been: The Social Psychology of Counterfactual Thinking*. Mahwah, NJ: Lawrence Erlbaum.
- SCHROEDER, P. W. 2007. Necessary conditions and World War I as an unavoidable war. Pp. 147–93 in *Explaining War and Peace: Case Studies and Necessary Condition Counterfactuals*, ed. G. Goertz and J. S. Levy. New York: Routledge.
- SKOCPOL, T. 1979. *States and Social Revolutions*. Cambridge: Cambridge University Press.
- TETLOCK, P. E., and BELKIN, A. 1996. Counterfactual thought experiments in world politics: logical, methodological, and psychological perspectives. Pp. 1–38 in *Counterfactual Thought Experiments in World Politics*, ed. P. E. Tetlock and A. Belkin. Princeton, NJ: Princeton University Press.
- and LEBOW, R. N. 2001. Poking counterfactual holes in covering laws: cognitive styles and historical reasoning. *American Political Science Review*, 95: 829–43.
- and PARKER, G. (eds.) 2006. *Unmaking the West: "What If?" Scenarios that Rewrite World History*. Ann Arbor: University of Michigan Press.
- WEBER, M. 1949. *The Methodology of the Social Sciences*. Glencoe, Ill.: Free Press.
- WEBER, S. 1996. Counterfactuals, past and future. Pp. 268–88 in *Counterfactual Thought Experiments in World Politics*, ed. P. E. Tetlock and A. Belkin. Princeton, NJ: Princeton University Press.
- WEINGAST, B. R. 1996. Off-the-path behavior: a game-theoretic approach to counterfactuals and its implications for political and historical analysis. Pp. 230–43 in *Counterfactual Thought Experiments in World Politics*, ed. P. E. Tetlock and A. Belkin. Princeton, NJ: Princeton University Press.

## CHAPTER 28

---

# CASE SELECTION FOR CASE-STUDY ANALYSIS: QUALITATIVE AND QUANTITATIVE TECHNIQUES

---

JOHN GERRING

CASE-study analysis focuses on one or several cases that are expected to provide insight into a larger population. This presents the researcher with a formidable problem of case selection: Which cases should she or he choose?

In large-sample research, the task of case selection is usually handled by some version of randomization. However, in case-study research the sample is small (by definition) and this makes random sampling problematic, for any given sample may be wildly unrepresentative. Moreover, there is no guarantee that a few cases, chosen randomly, will provide leverage into the research question of interest.

In order to isolate a sample of cases that both reproduces the relevant causal features of a larger universe (representativeness) and provides variation along the dimensions of theoretical interest (causal leverage), case selection for very small samples must employ purposive (nonrandom) selection procedures. Nine such methods are discussed in this chapter, each of which may be identified with a distinct case-study

“type:” *typical, diverse, extreme, deviant, influential, crucial, pathway, most-similar, and most-different*. Table 28.1 summarizes each type, including its general definition, a technique for locating it within a population of potential cases, its uses, and its probable representativeness.

While each of these techniques is normally practiced on one or several cases (the diverse, most-similar, and most-different methods require at least two), all *may* employ additional cases—with the proviso that, at some point, they will no longer offer an opportunity for in-depth analysis and will thus no longer be “case studies” in the usual sense (Gerring 2007, ch. 2). It will also be seen that small-*N* case-selection procedures rest, at least implicitly, upon an analysis of a larger population of potential cases (as does randomization). The case(s) identified for intensive study is chosen from a population and the reasons for this choice hinge upon the way in which it is situated within that population. This is the origin of the terminology—typical, diverse, extreme, et al. It follows that case-selection procedures in case-study research may build upon prior cross-case analysis and that they depend, at the very least, upon certain assumptions about the broader population.

In certain circumstances, the case-selection procedure may be structured by a quantitative analysis of the larger population. Here, several caveats must be satisfied. First, the inference must pertain to more than a few dozen cases; otherwise, statistical analysis is problematic. Second, relevant data must be available for that population, or a significant sample of that population, on key variables, and the researcher must feel reasonably confident in the accuracy and conceptual validity of these variables. Third, all the standard assumptions of statistical research (e.g. identification, specification, robustness) must be carefully considered, and wherever possible, tested. I shall not dilate further on these familiar issues except to warn the researcher against the unreflective use of statistical techniques.<sup>1</sup> When these requirements are *not* met, the researcher must employ a qualitative approach to case selection.

The point of this chapter is to elucidate general principles that might guide the process of case selection in case-study research, building upon earlier work by Harry Eckstein, Arend Lijphart, and others. Sometimes, these principles can be applied in a quantitative framework and sometimes they are limited to a qualitative framework. In either case, the logic of case selection remains quite similar, whether practiced in small-*N* or large-*N* contexts.

Before we begin, a bit of notation is necessary. In this chapter “*N*” refers to cases, not observations. Here, I am concerned primarily with causal inference, rather than inferences that are descriptive or predictive in nature. Thus, all hypotheses involve at least one independent variable (*X*) and one dependent variable (*Y*). For convenience, I shall label the causal factor of special theoretical interest  $X_1$ , and the control variable, or vector of controls (if there are any),  $X_2$ . If the writer is concerned to explain a puzzling outcome, but has no preconceptions about its causes, then the research will be described as *Y-centered*. If a researcher is concerned to

<sup>1</sup> Gujarati (2003); Kennedy (2003). Interestingly, the potential of cross-case statistics in helping to choose cases for in-depth analysis is recognized in some of the earliest discussions of the case-study method (e.g. Queen 1928, 226).

Table 28.1. Techniques of case selection

1. Typical
  - *Definition*: Cases (1 or more) are typical examples of some cross-case relationship.
  - *Cross-case technique*: A low-residual case (on-lier).
  - *Uses*: Hypothesis-testing.
  - *Representativeness*: By definition, the typical case is representative.
2. Diverse
  - *Definition*: Cases (2 or more) illuminate the full range of variation on  $X_1$ ,  $Y$ , or  $X_1/Y$ .
  - *Cross-case technique*: Diversity may be calculated by (a) categorical values of  $X_1$  or  $Y$  (e.g. Jewish, Catholic, Protestant), (b) standard deviations of  $X_1$  or  $Y$  (if continuous), (c) combinations of values (e.g. based on cross-tabulations, factor analysis, or discriminant analysis).
  - *Uses*: Hypothesis generating or hypothesis testing.
  - *Representativeness*: Diverse cases are likely to be representative in the minimal sense of representing the full variation of the population (though they might not mirror the *distribution* of that variation in the population).
3. Extreme
  - *Definition*: Cases (1 or more) exemplify extreme or unusual values of  $X_1$  or  $Y$  relative to some univariate distribution.
  - *Cross-case technique*: A case lying many standard deviations away from the mean of  $X_1$  or  $Y$ .
  - *Uses*: Hypothesis-generating (open-ended probe of  $X_1$  or  $Y$ ).
  - *Representativeness*: Achievable only in comparison with a larger sample of cases.
4. Deviant
  - *Definition*: Cases (1 or more) deviate from some cross-case relationship.
  - *Cross-case technique*: A high-residual case (outlier).
  - *Uses*: Hypothesis-generating (to develop new explanations for  $Y$ ).
  - *Representativeness*: After the case study is conducted it may be corroborated by a cross-case test, which includes a general hypothesis (a new variable) based on the case-study research. If the case is now an on-lier, it may be considered representative of the new relationship.
5. Influential
  - *Definition*: Cases (1 or more) with influential configurations of the independent variables.
  - *Cross-case technique*: Hat matrix or Cook's Distance.
  - *Uses*: Hypothesis-testing (to verify the status of cases that may influence the results of a cross-case analysis).
  - *Representativeness*: Not pertinent, given the goals of the influential-case study.
6. Crucial
  - *Definition*: Cases (1 or more) are most or least likely to exhibit a given outcome.
  - *Cross-case technique*: Qualitative assessment of relative crucialness.
  - *Uses*: Hypothesis-testing (confirmatory or disconfirmatory).
  - *Representativeness*: Often difficult to assess.
7. Pathway
  - *Definition*: Cases (1 or more) that embody a distinct causal path from  $X_1$  to  $Y$ .
  - *Cross-case technique*: Cross-tab (for categorical variables) or residual analysis (for continuous variables).
  - *Uses*: Hypothesis-testing (to probe causal mechanisms).
  - *Representativeness*: May be tested by examining residuals for the chosen cases.

(cont.)

Table 28.1. (Continued)

- |  |
|--|
| 8. Most-similar  |
| ○ <i>Definition</i> : Cases (2 or more) are similar on specified variables other than $X_1$ and/or $Y$ .         |
| ○ <i>Cross-case technique</i> : Matching.  |
| ○ <i>Uses</i> : Hypothesis-generating or hypothesis-testing.   |
| ○ <i>Representativeness</i> : May be tested by examining residuals for the chosen cases.                         |
| 9. Most-different  |
| ○ <i>Definition</i> : Cases (2 or more) are different on specified variables other than $X_1$ and $Y$ .          |
| ○ <i>Cross-case technique</i> : The inverse of the most-similar method of large- $N$ case selection (see above). |
| ○ <i>Uses</i> : Hypothesis-generating or hypothesis-testing (eliminating deterministic causes).                  |
| ○ <i>Representativeness</i> : May be tested by examining residuals for the chosen cases.                         |

investigate the effects of a particular cause, with no preconceptions about what these effects might be, the research will be described as *X-centered*. If a researcher is concerned to investigate a particular causal relationship, the research will be described as *X<sub>1</sub>/Y-centered*, for it connects a particular cause with a particular outcome.<sup>2</sup> *X*- or *Y*-centered research is exploratory; its purpose is to generate new hypotheses. *X<sub>1</sub>/Y-centered* research, by contrast, is confirmatory/disconfirmatory; its purpose is to test an existing hypothesis.

## 1 TYPICAL CASE

In order for a focused case study to provide insight into a broader phenomenon it must be representative of a broader set of cases. It is in this context that one may speak of a *typical-case* approach to case selection. The typical case exemplifies what is considered to be a typical set of values, given some general understanding of a phenomenon. By construction, the typical case is also a representative case.

Some typical cases serve an exploratory role. Here, the author chooses a case based upon a set of descriptive characteristics and then probes for causal relationships. Robert and Helen Lynd (1929/1956) selected a single city "to be as representative as possible of contemporary American life." Specifically, they were looking for a city with

1) a temperate climate; 2) a sufficiently rapid rate of growth to ensure the presence of a plentiful assortment of the growing pains accompanying contemporary social change; 3) an industrial culture with modern, high-speed machine production; 4) the absence of dominance of the city's industry by a single plant (i.e., not a one-industry town); 5) a substantial local artistic

<sup>2</sup> This expands on Mill (1843/1872, 253), who wrote of scientific enquiry as twofold: "either inquiries into the cause of a given effect or into the effects or properties of a given cause."

life to balance its industrial activity...; and 6) the absence of any outstanding peculiarities or acute local problems which would mark the city off from the midchannel sort of American community. (Lynd and Lynd 1929/1956, quoted in Yin 2004, 29–30)

After examining a number of options the Lynds decided that Muncie, Indiana, was more representative than, or at least as representative as, other mid-sized cities in America, thus qualifying as a typical case.

This is an inductive approach to case selection. Note that typicality may be understood according to the mean, median, or mode on a particular dimension; there may be multiple dimensions (as in the foregoing example); and each may be differently weighted (some dimensions may be more important than others). Where the selection criteria are multidimensional and a large sample of potential cases is in play, some form of factor analysis may be useful in identifying the most-typical case(s).

However, the more common employment of the typical-case method involves a *causal* model of some phenomenon of theoretical interest. Here, the researcher has identified a particular outcome ( $Y$ ), and perhaps a specific  $X_1/Y$  hypothesis, which she wishes to investigate. In order to do so, she looks for a typical example of that causal relationship. Intuitively, one imagines that a case selected according to the mean values of all parameters must be a typical case relative to some causal relationship. However, this is by no means assured.

Suppose that the Lynds were primarily interested in explaining feelings of trust/distrust among members of different social classes (one of the implicit research goals of the *Middletown* study). This outcome is likely to be affected by many factors, only some of which are included in their six selection criteria. So choosing cases with respect to a causal hypothesis involves, first of all, identifying the relevant parameters. It involves, secondly, the selection of a case that has a "typical" value relative to the overall causal model; it is well explained. Cases with untypical scores on a particular dimension (e.g. very high or very low) may still be typical examples of a causal relationship. Indeed, they may be more typical than cases whose values lie close to the mean. Thus, a descriptive understanding of typicality is quite different from a causal understanding of typicality. Since it is the latter version that is more common, I shall adopt this understanding of typicality in the remainder of the discussion.

From a qualitative perspective, causal typicality involves the selection of a case that conforms to expectations about some general causal relationship. It performs as expected. In a quantitative setting, this notion is measured by the size of a case's residual in a large- $N$  cross-case model. Typical cases lie on or near the regression line; their residuals are small. Insofar as the model is correctly specified, the size of a case's residual (i.e. the number of standard deviations that separate the actual value from the fitted value) provides a helpful clue to how representative that case is likely to be. "Outliers" are unlikely to be representative of the target population.

Of course, just because a case has a low residual does not *necessarily* mean that it is a representative case (with respect to the causal relationship of interest). Indeed,

the issue of case representativeness is an issue that can never be definitively settled. When one refers to a "typical case" one is saying, in effect, that the *probability* of a case's representativeness is high, relative to other cases. This test of typicality is misleading if the statistical model is mis-specified. And it provides little insurance against errors that are purely stochastic. A case may lie directly on the regression line but still be, in some important respect, atypical. For example, it might have an odd combination of values; the interaction of variables might be different from other cases; or additional causal mechanisms might be at work. For this reason, it is important to supplement a statistical analysis of cases with evidence drawn from the case in question (the case study itself) and with our deductive knowledge of the world. One should never judge a case solely by its residual. Yet, all other things being equal, a case with a low residual is less likely to be unusual than a case with a high residual, and to this extent the method of case selection outlined here may be a helpful guide to case-study researchers faced with a large number of potential cases.

By way of conclusion, it should be noted that because the typical case embodies a typical value on some set of causally relevant dimensions, the variance of interest to the researcher must lie *within* that case. Specifically, the typical case of some phenomenon may be helpful in exploring causal mechanisms and in solving identification problems (e.g. endogeneity between  $X_1$  and  $Y$ , an omitted variable that may account for  $X_1$  and  $Y$ , or some other spurious causal association). Depending upon the results of the case study, the author may confirm an existing hypothesis, disconfirm that hypothesis, or reframe it in a way that is consistent with the findings of the case study. These are the uses of the typical-case study.

## 2 DIVERSE CASES

A second case-selection strategy has as its primary objective the achievement of maximum variance along relevant dimensions. I refer to this as a *diverse-case* method. For obvious reasons, this method requires the selection of a set of cases—at minimum, two—which are intended to represent the full range of values characterizing  $X_1$ ,  $Y$ , or some particular  $X_1/Y$  relationship.<sup>3</sup>

Where the individual variable of interest is categorical (on/off, red/black/blue, Jewish/Protestant/Catholic), the identification of diversity is readily apparent. The investigator simply chooses one case from each category. For a continuous variable,

<sup>3</sup> This method has not received much attention on the part of qualitative methodologists; hence, the absence of a generally recognized name. It bears some resemblance to J. S. Mill's Joint Method of Agreement and Difference (Mill 1843/1872), which is to say a mixture of most-similar and most-different analysis, as discussed below. Patton (2002, 234) employs the concept of "maximum variation (heterogeneity) sampling."

the choices are not so obvious. However, the researcher usually chooses both extreme values (high and low), and perhaps the mean or median as well. The researcher may also look for break-points in the distribution that seem to correspond to categorical differences among cases. Or she may follow a theoretical hunch about which threshold values count, i.e. which are likely to produce different values on  $Y$ .

Another sort of diverse case takes account of the values of multiple variables (i.e. a vector), rather than a single variable. If these variables are categorical, the identification of causal types rests upon the intersection of each category. Two dichotomous variables produce a matrix with four cells. Three trichotomous variables produce a matrix of eight cells. And so forth. If all variables are deemed relevant to the analysis, the selection of diverse cases mandates the selection of one case drawn from within each cell. Let us say that an outcome is thought to be affected by sex, race (black/white), and marital status. Here, a diverse-case strategy of case selection would identify one case within each of these intersecting cells—a total of eight cases. Things become slightly more complicated when one or more of the factors is continuous, rather than categorical. Here, the diversity of case values do not fall neatly into cells. Rather, these cells must be created by fiat—e.g. high, medium, low.

It will be seen that where multiple variables are under consideration, the logic of diverse-case analysis rests upon the logic of *typological* theorizing—where different combinations of variables are assumed to have effects on an outcome that vary across types (Elman 2005; George and Bennett 2005, 235; Lazarsfeld and Barton 1951). George and Smoke, for example, wish to explore different types of deterrence failure—by "fait accompli," by "limited probe," and by "controlled pressure." Consequently, they wish to find cases that exemplify each type of causal mechanism.<sup>4</sup>

Diversity may thus refer to a range of variation on  $X$  or  $Y$ , or to a particular combination of causal factors (with or without a consideration of the outcome). In each instance, the goal of case selection is to capture the full range of variation along the dimension(s) of interest.

Since diversity can mean many things, its employment in a large- $N$  setting is necessarily dependent upon how this key term is defined. If it is understood to pertain only to a single variable ( $X_1$  or  $Y$ ), then the task is fairly simple. A categorical variable mandates the choice of at least one case from each category—two if dichotomous, three if trichotomous, and so forth. A continuous variable suggests the choice of at least one "high" and "low" value, and perhaps one drawn from the mean or median. But other choices might also be justified, according to one's hunch about the underlying causal relationship or according to natural thresholds found in the data, which may be grouped into discrete categories. Single-variable traits are usually easy to discover in a large- $N$  setting through descriptive statistics or through visual inspection of the data.

<sup>4</sup> More precisely, George and Smoke (1974, 534, 522–36, ch. 18; see also discussion in Collier and Mahoney 1996, 78) set out to investigate causal pathways and discovered, through the course of their investigation of many cases, these three causal types. Yet, for our purposes what is important is that the final sample includes at least one representative of each "type."

Where diversity refers to particular *combinations* of variables, the relevant cross-case technique is some version of stratified random sampling (in a probabilistic setting) or Qualitative Comparative Analysis (in a deterministic setting) (Ragin 2000). If the researcher suspects that a causal relationship is affected not only by combinations of factors but also by their *sequencing*, then the technique of analysis must incorporate temporal elements (Abbott 2001; Abbott and Forrest 1986; Abbott and Tsay 2000). Thus, the method of identifying causal types rests upon whatever method of identifying causal relationships is employed in the large-*N* sample.

Note that the identification of distinct case types is intended to identify groups of cases that are internally homogeneous (in all respects that might affect the causal relationship of interest). Thus, the choice of cases within each group should not be problematic, and may be accomplished through random sampling or purposive case selection. However, if there is suspected diversity within each category, then measures should be taken to assure that the chosen cases are typical of each category. A case study should not focus on an atypical member of a subgroup.

Indeed, considerations of diversity and typicality often go together. Thus, in a study of globalization and social welfare systems, Duane Swank (2002) first identifies three distinctive groups of welfare states: “universalistic” (social democratic), “corporatist conservative,” and “liberal.” Next, he looks within each group to find the most-typical cases. He decides that the Nordic countries are more typical of the universalistic model than the Netherlands since the latter has “some characteristics of the occupationally based program structure and a political context of Christian Democratic-led governments typical of the corporatist conservative nations” (Swank 2002, 11; see also Esping-Andersen 1990). Thus, the Nordic countries are chosen as representative cases within the universalistic case type, and are accompanied in the case-study portion of his analysis by other cases chosen to represent the other welfare state types (corporatist conservative and liberal).

Evidently, when a sample encompasses a full range of variation on relevant parameters one is likely to enhance the representativeness of that sample (relative to some population). This is a distinct advantage. Of course, the inclusion of a full range of variation may distort the actual distribution of cases across this spectrum. If there are more “high” cases than “low” cases in a population and the researcher chooses only one high case and one low case, the resulting sample of two is not perfectly representative. Even so, the diverse-case method probably has stronger claims to representativeness than any other small-*N* sample (including the standalone typical case). The selection of diverse cases has the additional advantage of introducing variation on the key variables of interest. A set of diverse cases is, by definition, a set of cases that encompasses a range of high and low values on relevant dimensions. There is, therefore, much to recommend this method of case selection. I suspect that these advantages are commonly understood and are applied on an intuitive level by case-study researchers. However, the lack of a recognizable name—and an explicit methodological defense—has made it difficult for case-study researchers to utilize this method of case selection, and to do so in an explicit and self-conscious fashion. Neologism has its uses.

### 3 EXTREME CASE

The *extreme-case* method selects a case because of its extreme value on an independent ( $X_1$ ) or dependent ( $Y$ ) variable of interest. Thus, studies of domestic violence may choose to focus on extreme instances of abuse (Browne 1987). Studies of altruism may focus on those rare individuals who risked their lives to help others (e.g. Holocaust resisters) (Monroe 1996). Studies of ethnic politics may focus on the most heterogeneous societies (e.g. Papua New Guinea) in order to better understand the role of ethnicity in a democratic setting (Reilly 2000–1). Studies of industrial policy often focus on the most successful countries (i.e. the NICS) (Deyo 1987). And so forth.<sup>5</sup>

Often an extreme case corresponds to a case that is considered to be prototypical or paradigmatic of some phenomena of interest. This is because concepts are often defined by their extremes, i.e. their ideal types. Italian Fascism defines the concept of Fascism, in part, because it offered the most extreme example of that phenomenon. However, the *methodological* value of this case, and others like it, derives from its extremity (along some dimension of interest), not its theoretical status or its status in the literature on a subject.

The notion of “extreme” may now be defined more precisely. An extreme value is an observation that lies far away from the mean of a given distribution. This may be measured (if there are sufficient observations) by a case’s “*Z* score”—the number of standard deviations between a case and the mean value for that sample. Extreme cases have high *Z* scores, and for this reason may serve as useful subjects for intensive analysis.

For a continuous variable, the distance from the mean may be in either direction (positive or negative). For a dichotomous variable (present/absent), extremeness may be interpreted as *unusual*. If most cases are positive along a given dimension, then a negative case constitutes an extreme case. If most cases are negative, then a positive case constitutes an extreme case. It should be clear that researchers are not simply concerned with cases where something “happened,” but also with cases where something did not. It is the rareness of the value that makes a case valuable, in this context, not its positive or negative value.<sup>6</sup> Thus, if one is studying state capacity, a case of state failure is probably more informative than a case of state endurance simply because the former is more unusual. Similarly, if one is interested in incest taboos a culture where the incest taboo is absent or weak is probably more useful than a culture where it is present or strong. Fascism is more important than nonfascism. And so forth. There is a good reason, therefore, why case studies of revolution tend to focus on “revolutionary” cases. Theda Skocpol (1979) had much more to learn from France than from Austro-Hungary since France was more unusual than Austro-Hungary within the population of nation states that Skocpol was

<sup>5</sup> For further examples see Collier and Mahoney (1996); Geddes (1990); Tendler (1997).

<sup>6</sup> Traditionally, methodologists have conceptualized cases as having “positive” or “negative” values (e.g. Emigh 1997; Mahoney and Goertz 2004; Ragin 2000, 60; 2004, 126).

concerned to explain. The reason is quite simple: There are fewer revolutionary cases than nonrevolutionary cases; thus, the variation that we explore as a clue to causal relationships is encapsulated in these cases, against a background of nonrevolutionary cases.

Note that the extreme-case method of case selection appears to violate the social science folk wisdom warning us not to “select on the dependent variable.”<sup>7</sup> Selecting cases on the dependent variable is indeed problematic if a number of cases are chosen, all of which lie on one end of a variable’s spectrum (they are all positive or negative), and if the researcher then subjects this sample to cross-case analysis as if it were representative of a population.<sup>8</sup> Results for this sort of analysis would almost assuredly be biased. Moreover, there will be little variation to explain since the values of each case are explicitly constrained.

However, this is not the proper employment of the extreme-case method. (It is more appropriately labeled an extreme-*sample* method.) The extreme-case method actually refers back to a larger sample of cases that lie in the background of the analysis and provide a full range of variation as well as a more representative picture of the population. It is a self-conscious attempt to *maximize* variance on the dimension of interest, not to minimize it. If this population of cases is well understood—either through the author’s own cross-case analysis, through the work of others, or through common sense—then a researcher may justify the selection of a single case exemplifying an extreme value for within-case analysis. If not, the researcher may be well advised to follow a diverse-case method, as discussed above.

By way of conclusion, let us return to the problem of representativeness. It will be seen that an extreme case may be typical or deviant. There is simply no way to tell because the researcher has not yet specified an  $X_1/Y$  causal proposition. Once such a causal proposition has been specified one may then ask whether the case in question is similar to some population of cases in all respects that might affect the  $X_1/Y$  relationship of interest (i.e. unit homogeneous). It is at this point that it becomes possible to say, within the context of a cross-case statistical model, whether a case lies near to, or far from, the regression line. However, this sort of analysis means that the researcher is no longer pursuing an extreme-case method. The extreme-case method is purely exploratory—a way of probing possible causes of  $Y$ , or possible effects of  $X$ , in an open-ended fashion. If the researcher has some notion of what additional factors might affect the outcome of interest, or of what relationship the causal factor of interest might have with  $Y$ , then she ought to pursue one of the other methods explored in this chapter. This also implies that an extreme-case method may transform into a different kind of approach as a study evolves; that is, as a more specific hypothesis comes to light. Useful extreme cases at the outset of a study may prove less useful at a later stage of analysis.

<sup>7</sup> Geddes (1990); King, Keohane, and Verba (1994). See also discussion in Brady and Collier (2004); Collier and Mahoney (1996); Rogowski (1995).

<sup>8</sup> The exception would be a circumstance in which the researcher intends to disprove a deterministic argument (Dion 1998).

## 4 DEVIANT CASE

The *deviant-case* method selects that case(s) which, by reference to some general understanding of a topic (either a specific theory or common sense), demonstrates a surprising value. It is thus the contrary of the typical case. Barbara Geddes (2003) notes the importance of deviant cases in medical science, where researchers are habitually focused on that which is “pathological” (according to standard theory and practice). The *New England Journal of Medicine*, one of the premier journals of the field, carries a regular feature entitled Case Records of the Massachusetts General Hospital. These articles bear titles like the following: “An 80-Year-Old Woman with Sudden Unilateral Blindness” or “A 76-Year-Old Man with Fever, Dyspnea, Pulmonary Infiltrates, Pleural Effusions, and Confusion.”<sup>9</sup> Another interesting example drawn from the field of medicine concerns the extensive study now devoted to a small number of persons who seem resistant to the AIDS virus (Buchbinder and Vittinghoff 1999; Haynes, Pantaleo, and Fauci 1996). Why are they resistant? What is different about these people? What can we learn about AIDS in other patients by observing people who have built-in resistance to this disease?

Likewise, in psychology and sociology case studies may be comprised of deviant (in the social sense) persons or groups. In economics, case studies may consist of countries or businesses that overperform (e.g. Botswana; Microsoft) or underperform (e.g. Britain through most of the twentieth century; Sears in recent decades) relative to some set of expectations. In political science, case studies may focus on countries where the welfare state is more developed (e.g. Sweden) or less developed (e.g. the United States) than one would expect, given a set of general expectations about welfare state development. The deviant case is closely linked to the investigation of theoretical anomalies. Indeed, to say deviant is to imply “anomalous.”<sup>10</sup>

Note that while extreme cases are judged relative to the mean of a single distribution (the distribution of values along a single variable), deviant cases are judged relative to some general model of causal relations. The deviant-case method selects cases which, by reference to some (presumably) general relationship, demonstrate a surprising value. They are “deviant” in that they are poorly explained by the multivariate model. The important point is that deviant-ness can only be assessed relative to the general (quantitative or qualitative) model. This means that the relative deviant-ness of a case is likely to change whenever the general model is altered. For example, the United States is a deviant welfare state when this outcome is gauged relative to societal wealth. But it is less deviant—and perhaps not deviant at all—when certain additional (political and societal) factors are included in the model, as discussed in

<sup>9</sup> Geddes (2003, 131). For other examples of casework from the annals of medicine see “Clinical reports” in the *Lancet*, “Case studies” in *Canadian Medical Association Journal*, and various issues of the *Journal of Obstetrics and Gynecology*, often devoted to clinical cases (discussed in Jenicek 2001, 7). For examples from the subfield of comparative politics see Kazancigil (1994).

<sup>10</sup> For a discussion of the important role of anomalies in the development of scientific theorizing see Elman (2003); Lakatos (1978). For examples of deviant-case research designs in the social sciences see Amenta (1991); Coppedge (2004); Eckstein (1975); Emigh (1997); Kendall and Wolf (1949/1955).

the epilogue. Deviance is model dependent. Thus, when discussing the concept of the deviant case it is helpful to ask the following question: *Relative to what general model* (or set of background factors) is Case A deviant?

Conceptually, we have said that the deviant case is the logical contrary of the typical case. This translates into a directly contrasting statistical measurement. While the typical case is one with a low residual (in some general model of causal relations), a deviant case is one with a high residual. This means, following our previous discussion, that the deviant case is likely to be an *unrepresentative* case, and in this respect appears to violate the supposition that case-study samples should seek to reproduce features of a larger population.

However, it must be borne in mind that the primary purpose of a deviant-case analysis is to probe for new—but as yet unspecified—explanations. (If the purpose is to disprove an extant theory I shall refer to the study as crucial-case, as discussed below.) The researcher hopes that causal processes identified within the deviant case will illustrate some causal factor that is applicable to other (more or less deviant) cases. This means that a deviant-case study usually culminates in a general proposition, one that may be applied to other cases in the population. Once this general proposition has been introduced into the overall model, the expectation is that the chosen case will no longer be an outlier. Indeed, the hope is that it will now be *typical*, as judged by its small residual in the adjusted model. (The exception would be a circumstance in which a case's outcome is deemed to be "accidental," and therefore inexplicable by any general model.)

This feature of the deviant-case study should help to resolve questions about its representativeness. Even if it is not possible to measure the new causal factor (and thus to introduce it into a large-*N* cross-case model), it may still be plausible to assert (based on general knowledge of the phenomenon) that the chosen case is representative of a broader population.

## 5 INFLUENTIAL CASE

Sometimes, the choice of a case is motivated solely by the need to verify the assumptions behind a general model of causal relations. Here, the analyst attempts to provide a rationale for disregarding a problematic case or a set of problematic cases. That is to say, she attempts to show why apparent deviations from the norm are not *really* deviant, or do not challenge the core of the theory, once the circumstances of the special case or cases are fully understood. A cross-case analysis may, after all, be marred by several classes of problems including measurement error, specification error, errors in establishing proper boundaries for the inference (the scope of the argument), and stochastic error (fluctuations in the phenomenon under study that are treated as random, given available theoretical resources). If poorly fitting cases

can be explained away by reference to these kinds of problems, then the theory of interest is that much stronger. This sort of deviant-case analysis answers the question, "What about Case A (or cases of type A)? How does that, seemingly disconfirming, case fit the model?"

Because its underlying purpose is different from the usual deviant-case study, I offer a new term for this method. The *influential case* is a case that casts doubt upon a theory, and for that reason warrants close inspection. This investigation may reveal, after all, that the theory is validated—perhaps in some slightly altered form. In this guise, the influential case is the "case that proves the rule." In other instances, the influential-case analysis may contribute to disconfirming, or reconceptualizing, a theory. The key point is that the value of the case is judged relative to some extant cross-case model.

A simple version of influential-case analysis involves the confirmation of a key case's score on some critical dimension. This is essentially a question of measurement. Sometimes cases are poorly explained simply because they are poorly understood. A close examination of a particular context may reveal that an apparently falsifying case has been miscoded. If so, the initial challenge presented by that case to some general theory has been obviated.

However, the more usual employment of the influential-case method culminates in a substantive reinterpretation of the case—perhaps even of the general model. It is not just a question of measurement. Consider Thomas Ertman's (1997) study of state building in Western Europe, as summarized by Gerardo Munck. This study argues

that the interaction of a) the type of local government during the first period of statebuilding, with b) the timing of increases in geopolitical competition, strongly influences the kind of regime and state that emerge. [Ertman] tests this hypothesis against the historical experience of Europe and finds that most countries fit his predictions. Denmark, however, is a major exception. In Denmark, sustained geopolitical competition began relatively late and local government at the beginning of the statebuilding period was generally participatory, which should have led the country to develop "patrimonial constitutionalism." But in fact, it developed "bureaucratic absolutism." Ertman carefully explores the process through which Denmark came to have a bureaucratic absolutist state and finds that Denmark had the early marks of a patrimonial constitutionalist state. However, the country was pushed off this developmental path by the influence of German knights, who entered Denmark and brought with them German institutions of local government. Ertman then traces the causal process through which these imported institutions pushed Denmark to develop bureaucratic absolutism, concluding that this development was caused by a factor well outside his explanatory framework.

(Munck 2004, 118)

Ertman's overall framework is confirmed insofar as he has been able to show, by an in-depth discussion of Denmark, that the causal processes stipulated by the general theory hold even in this apparently disconfirming case. Denmark is still deviant, but it is so because of "contingent historical circumstances" that are exogenous to the theory (Ertman 1997, 316).

Evidently, the influential-case analysis is similar to the deviant-case analysis. Both focus on outliers. However, as we shall see, they focus on different kinds of outliers.



Moreover, the animating goals of these two research designs are quite different. The influential-case study begins with the aim of confirming a general model, while the deviant-case study has the aim of generating a new hypothesis that modifies an existing general model. The confusion stems from the fact that the same case study may fulfill both objectives—qualifying a general model and, at the same time, confirming its core hypothesis.

Thus, in their study of Roberto Michels's "iron law of oligarchy," Lipset, Trow, and Coleman (1956) choose to focus on an organization—the International Typographical Union—that appears to violate the central presupposition. The ITU, as noted by one of the authors, has "a long-term two-party system with free elections and frequent turnover in office" and is thus anything but oligarchic (Lipset 1959, 70). As such, it calls into question Michels's grand generalization about organizational behavior. The authors explain this curious result by the extraordinarily high level of education among the members of this union. Michels's law is shown to be true for most organizations, but not all. It is true, with qualifications. Note that the respecification of the original model (in effect, Lipset, Trow, and Coleman introduce a new control variable or boundary condition) involves the exploration of a new hypothesis. In this instance, therefore, the use of an influential case to confirm an existing theory is quite similar to the use of a deviant case to explore a new theory.

In a quantitative idiom, influential cases are those that, if counterfactually assigned a different value on the dependent variable, would most substantially change the resulting estimates. They may or may not be outliers (high-residual cases). Two quantitative measures of influence are commonly applied in regression diagnostics (Belsey, Kuh, and Welsch 2004). The first, often referred to as the leverage of a case, derives from what is called the *hat matrix*. Based solely on each case's scores on the independent variables, the hat matrix tells us how much a change in (or a measurement error on) the dependent variable for that case would affect the overall regression line. The second is *Cook's distance*, a measure of the extent to which the estimates of all the parameters would change if a given case were omitted from the analysis. Cases with a large leverage or Cook's distance contribute quite a lot to the inferences drawn from a cross-case analysis. In this sense, such cases are vital for maintaining analytic conclusions. Discovering a significant measurement error on the dependent variable or an important omitted variable for such a case may dramatically revise estimates of the overall relationships. Hence, it may be quite sensible to select influential cases for in-depth study.

Note that the use of an influential-case strategy of case selection is limited to instances in which a researcher has reason to be concerned that her results are being driven by one or a few cases. This is most likely to be true in small to moderate-sized samples. Where  $N$  is very large—greater than 1,000, let us say—it is extremely unlikely that a small set of cases (much less an individual case) will play an "influential" role. Of course, there may be influential *sets* of cases, e.g. countries within a particular continent or cultural region, or persons of Irish extraction. Sets of influential observations are often problematic in a time-series cross-section data-set where each

unit (e.g. country) contains multiple observations (through time), and hence may have a strong influence on aggregate results. Still, the general rule is: the larger the sample, the less important individual cases are likely to be and, hence, the less likely a researcher is to use an influential-case approach to case selection.

## 6 CRUCIAL CASE

Of all the extant methods of case selection perhaps the most storied—and certainly the most controversial—is the *crucial-case* method, introduced to the social science world several decades ago by Harry Eckstein. In his seminal essay, Eckstein (1975, 118) describes the crucial case as one "that *must closely fit* a theory if one is to have confidence in the theory's validity, or, conversely, *must not fit* equally well any rule contrary to that proposed." A case is crucial in a somewhat weaker—but much more common—sense when it is most, or least, likely to fulfill a theoretical prediction. A "most-likely" case is one that, on all dimensions *except* the dimension of theoretical interest, is predicted to achieve a certain outcome, and yet does not. It is therefore used to disconfirm a theory. A "least-likely" case is one that, on all dimensions *except* the dimension of theoretical interest, is predicted not to achieve a certain outcome, and yet does so. It is therefore used to confirm a theory. In all formulations, the crucial-case offers a most-difficult test for an argument, and hence provides what is perhaps the strongest sort of evidence possible in a nonexperimental, single-case setting.

Since the publication of Eckstein's influential essay, the crucial-case approach has been claimed in a multitude of studies across several social science disciplines and has come to be recognized as a staple of the case-study method.<sup>11</sup> Yet the idea of any single case playing a crucial (or "critical") role is not widely accepted among most methodologists (e.g. Sekhon 2004). (Even its progenitor seems to have had doubts.)

Let us begin with the confirmatory (a.k.a. least-likely) crucial case. The implicit logic of this research design may be summarized as follows. Given a set of facts, we are asked to contemplate the probability that a given theory is true. While the facts matter, to be sure, the effectiveness of this sort of research also rests upon the formal properties of the theory in question. Specifically, the degree to which a theory is amenable to confirmation is contingent upon how many predictions can be derived from the theory and on how "risky" each individual prediction is. In Popper's (1963, 36) words, "Confirmations should count only if they are the result of *risky predictions*; that is to say, if, unenlightened by the theory in question, we should have expected an

<sup>11</sup> For examples of the crucial-case method see Bennett, Lepgold, and Unger (1994); Desch (2002); Goodin and Smitsman (2000); Kemp (1986); Reilly and Phillpot (2003). For general discussion see George and Bennett (2005); Levy (2002); Stinchcombe (1968, 24–8).

event which was incompatible with the theory—and event which would have refuted the theory. Every ‘good’ scientific theory is a prohibition; it forbids certain things to happen. The more a theory forbids, the better it is” (see also Popper 1934/1968). A risky prediction is therefore one that is highly precise and determinate, and therefore unlikely to be achieved by the product of other causal factors (external to the theory of interest) or through stochastic processes. A theory produces many such predictions if it is fully elaborated, issuing predictions not only on the central outcome of interest but also on specific causal mechanisms, and if it is broad in purview. (The notion of riskiness may also be conceptualized within the Popperian lexicon as *degrees of falsifiability*.)

These points can also be articulated in Bayesian terms. Colin Howson and Peter Urbach explain: “The degree to which  $h$  [a hypothesis] is confirmed by  $e$  [a set of evidence] depends... on the extent to which  $P(e|h)$  exceeds  $P(e)$ , that is, on how much more probable  $e$  is relative to the hypothesis and background assumptions than it is relative just to background assumptions.” Again, “confirmation is correlated with how much more probable the evidence is if the hypothesis is true than if it is false” (Howson and Urbach 1989, 86). Thus, the stranger the prediction offered by a theory—relative to what we would normally expect—the greater the degree of confirmation that will be afforded by the evidence. As an intuitive example, Howson and Urbach (1989, 86) offer the following:

If a soothsayer predicts that you will meet a dark stranger sometime and you do in fact, your faith in his powers of precognition would not be much enhanced: you would probably continue to think his predictions were just the result of guesswork. However, if the prediction also gave the correct number of hairs on the head of that stranger, your previous scepticism would no doubt be severely shaken.

While these Popperian/Bayesian notions<sup>12</sup> are relevant to all empirical research designs, they are especially relevant to case-study research designs, for in these settings a single case (or, at most, a small number of cases) is required to bear a heavy burden of proof. It should be no surprise, therefore, that Popper’s idea of “riskiness” was to be appropriated by case-study researchers like Harry Eckstein to validate the enterprise of single-case analysis. (Although Eckstein does not cite Popper the intellectual lineage is clear.) Riskiness, here, is analogous to what is usually referred to as a “most-difficult” research design, which in a case-study research design would be understood as a “least-likely” case. Note also that the distinction between a “must-fit” case and a least-likely case—that, in the event, actually does fit the terms of a theory—is a matter of degree. Cases are more or less crucial for confirming theories. The point is that, in some circumstances, a paucity of empirical evidence may be compensated by the riskiness of the theory.

The crucial-case research design is, perforce, a highly deductive enterprise; much depends on the quality of the theory under investigation. It follows that the theories most amenable to crucial-case analysis are those which are lawlike in their precision,

<sup>12</sup> A third position, which purports to be neither Popperian or Bayesian, has been articulated by Mayo (1996, ch. 6). From this perspective, the same idea is articulated as a matter of “severe tests.”

degree of elaboration, consistency, and scope. The more a theory attains the status of a causal law, the easier it will be to confirm, or to disconfirm, with a single case. Indeed, risky predictions are common in natural science fields such as physics, which in turn served as the template for the deductive-nomological (“covering-law”) model of science that influenced Eckstein and others in the postwar decades (e.g. Hempel 1942).

A frequently cited example is the first important empirical demonstration of the theory of relativity, which took the form of a single-event prediction on the occasion of the May 29, 1919, solar eclipse (Eckstein 1975; Popper 1963). Stephen Van Evera (1997, 66–7) describes the impact of this prediction on the validation of Einstein’s theory.

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.

The strength of this test is the extraordinary fit between the theory and a set of facts found in a single case, and the corresponding lack of fit between all other theories and this set of facts. Einstein offered an explanation of a particular set of anomalous findings that no other existing theory could make sense of. Of course, one must assume that there was no—or limited—measurement error. And one must assume that the phenomenon of interest is largely invariant; light does not bend differently at different times and places (except in ways that can be understood through the theory of relativity). And one must assume, finally, that the theory itself makes sense on other grounds (other than the case of special interest); it is a plausible general theory. If one is willing to accept these a priori assumptions, then the 1919 “case study” provides a very strong confirmation of the theory. It is difficult to imagine a stronger proof of the theory from within an observational (nonexperimental) setting.

In social science settings, by contrast, one does not commonly find single-case studies offering knockout evidence for a theory. This is, in my view, largely a product of the looseness (the underspecification) of most social science theories. George and Bennett point out that while the thesis of the democratic peace is as close to a “law” as social science has yet seen, it cannot be confirmed (or refuted) by looking at specific causal mechanisms because the causal pathways mandated by the theory are multiple and diverse. Under the circumstances, no single-case test can offer strong confirmation of the theory (George and Bennett 2005, 209).

However, if one adopts a softer version of the crucial-case method—the least-likely (most difficult) case—then possibilities abound. Indeed, I suspect that, *implicitly*, most case-study work that makes a positive argument focusing on a single case (without a corresponding cross-case analysis) relies largely on the logic of the least-likely case. Rarely is this logic made explicit, except perhaps in a passing phrase or two. Yet the deductive logic of the “risky” prediction is central to the case-study enterprise.

Whether a case study is convincing or not often rests on the reader's evaluation of how strong the evidence for an argument might be, and this in turn—wherever cross-case evidence is limited and no manipulated treatment can be devised—rests upon an estimation of the degree of “fit” between a theory and the evidence at hand, as discussed.

Lily Tsai's (2007) investigation of governance at the village level in China employs several in-depth case studies of villages which are chosen (in part) because of their least-likely status relative to the theory of interest. Tsai's hypothesis is that villages with greater social solidarity (based on preexisting religious or familial networks) will develop a higher level of social trust and mutual obligation and, as a result, will experience better governance. Crucial cases, therefore, are villages that evidence a high level of social solidarity but which, along other dimensions, would be judged least likely to develop good governance, e.g. they are poor, isolated, and lack democratic institutions or accountability mechanisms from above. “Li Settlement,” in Fujian province, is such a case. The fact that this impoverished village nonetheless boasts an impressive set of infrastructural accomplishments such as paved roads with drainage ditches (a rarity in rural China) suggests that something rather unusual is going on here. Because her case is carefully chosen to eliminate rival explanations, Tsai's conclusions about the special role of social solidarity are difficult to gainsay. How else is one to explain this otherwise anomalous result? This is the strength of the least-likely case, where all other plausible causal factors for an outcome have been minimized.<sup>13</sup>

Jack Levy (2002, 144) refers to this, evocatively, as a “Sinatra inference:” if it can make it here, it can make it anywhere (see also Khong 1992, 49; Sagan 1995, 49; Shafer 1988, 14–6). Thus, if social solidarity has the hypothesized effect in Li Settlement it should have the same effect in more propitious settings (e.g. where there is greater economic surplus). The same implicit logic informs many case-study analyses where the intent of the study is to confirm a hypothesis on the basis of a single case.

Another sort of crucial case is employed for the purpose of *disconfirming* a causal hypothesis. A central Popperian insight is that it is easier to disconfirm an inference than to confirm that same inference. (Indeed, Popper doubted that any inference could be fully confirmed, and for this reason preferred the term “corroborate.”) This is particularly true of case-study research designs, where evidence is limited to one or several cases. The key proviso is that the theory under investigation must take a consistent (a.k.a. invariant, deterministic) form, even if its predictions are not terrifically precise, well elaborated, or broad.

As it happens, there are a fair number of invariant propositions floating around the social science disciplines (Goertz and Levy forthcoming; Goertz and Starr 2003). It used to be argued, for example, that political stability would occur only in countries that are relatively homogeneous, or where existing heterogeneities are mitigated by

<sup>13</sup> It should be noted that Tsai's conclusions do not rest solely on this crucial case. Indeed, she employs a broad range of methodological tools, encompassing case-study and cross-case methods.

cross-cutting cleavages (Almond 1956; Bentley 1908/1967; Lipset 1960/1963; Truman 1951). Arend Lijphart's (1968) study of the Netherlands, a peaceful country with reinforcing social cleavages, is commonly viewed as refuting this theory on the basis of a single in-depth case analysis.<sup>14</sup>

Granted, it may be questioned whether presumed invariant theories are *really* invariant; perhaps they are better understood as probabilistic. Perhaps, that is, the theory of cross-cutting cleavages is still true, probabilistically, despite the apparent Dutch exception. Or perhaps the theory is still true, deterministically, within a subset of cases that does not include the Netherlands. (This sort of claim seems unlikely in this particular instance, but it is quite plausible in many others.) Or perhaps the theory is in need of reframing; it is true, deterministically, but applies only to cross-cutting *ethnic/racial* cleavages, not to cleavages that are primarily religious. One can quibble over what it means to “disconfirm” a theory. The point is that the crucial case has, in all these circumstances, provided important updating of a theoretical prior.

Heretofore, I have treated causal factors as dichotomous. Countries have either reinforcing or cross-cutting cleavages and they have regimes that are either peaceful or conflictual. Evidently, these sorts of parameters are often matters of degree. In this reading of the theory, cases are *more or less* crucial. Accordingly, the most useful—i.e. most crucial—case for Lijphart's purpose is one that has the most segregated social groups and the most peaceful and democratic track record. In these respects, the Netherlands was a very good choice. Indeed, the degree of disconfirmation offered by this case study is probably greater than the degree of disconfirmation that might have been provided by other cases such as India or Papua New Guinea—countries where social peace has not always been secure. The point is that where variables are continuous rather than dichotomous it is possible to evaluate potential cases in terms of their *degree of crucialness*.

Note that the crucial-case method of case-selection, whether employed in a confirmatory or disconfirmatory mode, cannot be employed in a large-*N* context. This is because an explicit cross-case model would render the crucial-case study redundant. Once one identifies the relevant parameters and the scores of all cases on those parameters, one has in effect constructed a cross-case model that confirms or disconfirms the theory in question. The case study is thenceforth irrelevant, at least as a means of decisive confirmation or disconfirmation.<sup>15</sup> It remains highly relevant as a means of exploring causal mechanisms, of course. Yet, because this objective is quite different from that which is usually associated with the term, I enlist a new term for this technique.

<sup>14</sup> See also the discussion in Eckstein (1975) and Lijphart (1969). For additional examples of case studies disconfirming general propositions of a deterministic nature see Allen (1965); Lipset, Trow, and Coleman (1956); Njolstad (1990); Reilly (2000–1); and discussion in Dion (1998); Rogowski (1995).

<sup>15</sup> Granted, insofar as case-study analysis provides a window into causal mechanisms, and causal mechanisms are integral to a given theory, a single case may be enlisted to confirm or disconfirm a proposition. However, if the case study upholds a posited pattern of *X/Y* covariation, and finds fault only with the stipulated causal mechanism, it would be more accurate to say that the study forces the *reformulation* of a given theory, rather than its confirmation or disconfirmation. See further discussion in the following section.

## 7 PATHWAY CASE

One of the most important functions of case-study research is the elucidation of causal mechanisms. But which sort of case is most useful for this purpose? Although all case studies presumably shed light on causal mechanisms, not all cases are equally transparent. In situations where a causal hypothesis is clear and has already been confirmed by cross-case analysis, researchers are well advised to focus on a case where the causal effect of  $X_1$  on  $Y$  can be isolated from other potentially confounding factors ( $X_2$ ). I shall call this a *pathway case* to indicate its uniquely penetrating insight into causal mechanisms. In contrast to the crucial case, this sort of method is practicable only in circumstances where cross-case covariational patterns are well studied and where the mechanism linking  $X_1$  and  $Y$  remains dim. Because the pathway case builds on prior cross-case analysis, the problem of case selection must be situated within that sample. There is no standalone pathway case.

The logic of the pathway case is clearest in situations of causal sufficiency—where a causal factor of interest,  $X_1$ , is sufficient by itself (though perhaps not necessary) to account for  $Y$ 's value (0 or 1). The other causes of  $Y$ , about which we need make no assumptions, are designated as a vector,  $X_2$ .

Note that wherever various causal factors are substitutable for one another, each factor is conceptualized (individually) as sufficient (Braumoeller 2003). Thus, situations of causal equifinality presume causal sufficiency on the part of each factor or set of conjoint factors. An example is provided by the literature on democratization, which stipulates three main avenues of regime change: leadership-initiated reform, a controlled opening to opposition, or the collapse of an authoritarian regime (Colomer 1991). The case-study format constrains us to analyze one at a time, so let us limit our scope to the first one—leadership-initiated reform. So considered, a causal-pathway case would be one with the following features: (a) democratization, (b) leadership-initiated reform, (c) no controlled opening to the opposition, (d) no collapse of the previous authoritarian regime, and (e) no other extraneous factors that might affect the process of democratization. In a case of this type, the causal mechanisms by which leadership-initiated reform may lead to democratization will be easiest to study. Note that it is not necessary to assume that leadership-initiated reform *always* leads to democratization; it may or may not be a deterministic cause. But it is necessary to assume that leadership-initiated reform can *sometimes* lead to democratization on its own (given certain background features).

Now let us move from these examples to a general-purpose model. For heuristic purposes, let us presume that all variables in that model are dichotomous (coded as 0 or 1) and that the model is complete (all causes of  $Y$  are included). All causal relationships will be coded so as to be positive:  $X_1$  and  $Y$  covary as do  $X_2$  and  $Y$ . This allows us to visualize a range of possible combinations at a glance.

Recall that the pathway case is always focused, by definition, on a single causal factor, denoted  $X_1$ . (The researcher's focus may shift to other causal factors, but may only focus on one causal factor at a time.) In this scenario, and regardless of how

Table 28.2. Pathway case with dichotomous causal factors

Case types	$X_1$	$X_2$	$Y$
A	1	1	1
B	0	0	0
C	0	1	1
D	0	0	1
E	1	0	0
F	1	1	0
G	1	0	1
H	0	1	0

Notes:  $X_1$  = the variable of theoretical interest.  $X_2$  = a vector of controls (a score of 0 indicates that all control variables have a score of 0, while a score of 1 indicates that all control variables have a score of 1).  $Y$  = the outcome of interest. A–H = case types (the  $N$  for each case type is indeterminate). G, H = possible pathway cases. Sample size = indeterminate.

Assumptions: (a) all variables can be coded dichotomously (a binary coding of the concept is valid); (b) all independent variables are positively correlated with  $Y$  in the general case; (c)  $X_1$  is (at least sometimes) a sufficient cause of  $Y$ .

many additional causes of  $Y$  there might be (denoted  $X_2$ , a vector of controls), there are only eight *relevant* case types, as illustrated in Table 28.2. Identifying these case types is a relatively simple matter, and can be accomplished in a small- $N$  sample by the construction of a truth-table (modeled after Table 28.2) or in a large- $N$  sample by the use of cross-tabs.

Note that the total number of combinations of values depends on the number of control variables, which we have represented with a single vector,  $X_2$ . If this vector consists of a single variable then there are only eight case types. If this vector consists of two variables ( $X_{2a}$ ,  $X_{2b}$ ) then the total number of possible combinations increases from eight ( $2^3$ ) to sixteen ( $2^4$ ). And so forth. However, none of these combinations is relevant for present purposes except those where  $X_{2a}$  and  $X_{2b}$  have the *same value* (0 or 1). “Mixed” cases are not causal pathway cases, for reasons that should become clear.

The pathway case, following the logic of the crucial case, is one where the causal factor of interest,  $X_1$ , correctly predicts  $Y$  while all other possible causes of  $Y$  (represented by the vector,  $X_2$ ) make “wrong” predictions. If  $X_1$  is—at least in some circumstances—a sufficient cause of  $Y$ , then it is these sorts of cases that should be most useful for tracing causal mechanisms. There are only two such cases in Table 28.2—G and H. In all other cases, the mechanism running from  $X_1$  to  $Y$  would be difficult to discern either because  $X_1$  and  $Y$  are not correlated in the usual way (constituting an unusual case, in the terms of our hypothesis) or because other confounding factors ( $X_2$ ) intrude. In case A, for example, the positive value on  $Y$

could be a product of  $X_1$  or  $X_2$ . An in-depth examination of this case is not likely to be very revealing.

Keep in mind that because the researcher already knows from her cross-case examination what the general causal relationships are, she knows (prior to the case-study investigation) what constitutes a correct or incorrect prediction. In the crucial-case method, by contrast, these expectations are deductive rather than empirical. This is what differentiates the two methods. And this is why the causal pathway case is useful principally for elucidating causal mechanisms rather than verifying or falsifying general propositions (which are already more or less apparent from the cross-case evidence). Of course, we must leave open the possibility that the investigation of causal mechanisms would invalidate a general claim, if that claim is utterly contingent upon a specific set of causal mechanisms and the case study shows that no such mechanisms are present. However, this is rather unlikely in most social science settings. Usually, the result of such a finding will be a reformulation of the causal processes by which  $X_1$  causes  $Y$ —or, alternatively, a realization that the case under investigation is aberrant (atypical of the general population of cases).

Sometimes, the research question is framed as a unidirectional cause: one is interested in why 0 becomes 1 (or vice versa) but not in why 1 becomes 0. In our previous example, we asked why democracies fail, *not* why countries become democratic or authoritarian. So framed, there can be only one type of causal-pathway case. (Whether regime failure is coded as 0 or 1 is a matter of taste.) Where researchers are interested in bidirectional causality—a movement from 0 to 1 as well as from 1 to 0—there are two possible causal-pathway cases, G and H. In practice, however, one of these case types is almost always more useful than the other. Thus, it seems reasonable to employ the term “pathway case” in the singular. In order to determine which of these two case types will be more useful for intensive analysis the researcher should look to see whether each case type exhibits desirable features such as: (a) a rare (unusual) value on  $X_1$  or  $Y$  (designated “extreme” in our previous discussion), (b) observable temporal variation in  $X_1$ , (c) an  $X_1/Y$  relationship that is easier to study (it has more visible features; it is more transparent), or (d) a lower residual (thus indicating a more typical case, within the terms of the general model). Usually, the choice between G and H is intuitively obvious.

Now, let us consider a scenario in which all (or most) variables of concern to the model are continuous, rather than dichotomous. Here, the job of case selection is considerably more complex, for causal “sufficiency” (in the usual sense) cannot be invoked. It is no longer plausible to assume that a given cause can be entirely partitioned, i.e. rival factors eliminated. However, the search for a pathway case may still be viable. What we are looking for in this scenario is a case that satisfies two criteria: (1) it is not an outlier (or at least not an extreme outlier) in the general model and (2) its score on the outcome ( $Y$ ) is strongly influenced by the theoretical variable of interest ( $X_1$ ), taking all other factors into account ( $X_2$ ). In this sort of case it should be easiest to “see” the causal mechanisms that lie between  $X_1$  and  $Y$ .

Achieving the second desiderata requires a bit of manipulation. In order to determine which (nonoutlier) cases are most strongly affected by  $X_1$ , given all the other

parameters in the model, one must compare the size of the residuals for each case in a reduced form model,  $Y = \text{Constant} + X_2 + \text{Res}_{\text{reduced}}$ , with the size of the residuals for each case in a full model,  $Y = \text{Constant} + X_2 + X_1 + \text{Res}_{\text{full}}$ . The pathway case is that case, or set of cases, which shows the greatest difference between the residual for the reduced-form model and the full model ( $\Delta\text{Residual}$ ). Thus,

$$\text{Pathway} = |\text{Res}_{\text{reduced}} - \text{Res}_{\text{full}}|, \text{ if } |\text{Res}_{\text{reduced}}| > |\text{Res}_{\text{full}}|. \quad (1)$$

Note that the residual for a case must be smaller in the full model than in the reduced-form model; otherwise, the addition of the variable of interest ( $X_1$ ) pulls the case away from the regression line. We want to find a case where the addition of  $X_1$  pushes the case towards the regression line, i.e. it helps to “explain” that case.

As an example, let us suppose that we are interested in exploring the effect of mineral wealth on the prospects for democracy in a society. According to a good deal of work on this subject, countries with a bounty of natural resources—particularly oil—are less likely to democratize (or once having undergone a democratic transition, are more likely to revert to authoritarian rule) (Barro 1999; Humphreys 2005; Ross 2001). The cross-country evidence is robust. Yet as is often the case, the causal mechanisms remain rather obscure. In order to better understand this phenomenon it may be worthwhile to exploit the findings of cross-country regression models in order to identify a country whose regime type (i.e. its democracy “score” on some general index) is strongly affected by its natural-research wealth, all other things held constant. An analysis of this sort identifies two countries—the United Arab Emirates and Kuwait—with high  $\Delta$  Residual values and modest residuals in the full model (signifying that these cases are not outliers). Researchers seeking to explore the effect of oil wealth on regime type might do well to focus on these two cases since their patterns of democracy cannot be well explained by other factors—e.g. economic development, religion, European influence, or ethnic fractionalization. The presence of oil wealth in these countries would appear to have a strong *independent* effect on the prospects for democratization in these cases, an effect that is well modeled by general theory and by the available cross-case evidence.

To reiterate, the logic of causal “elimination” is much more compelling where variables are dichotomous and where causal sufficiency can be assumed ( $X_1$  is sufficient by itself, at least in some circumstances, to cause  $Y$ ). Where variables are continuous, the strategy of the pathway case is more dubious, for potentially confounding causal factors ( $X_2$ ) cannot be neatly partitioned. Even so, we have indicated why the selection of a pathway case may be a logical approach to case-study analysis in many circumstances.

The exceptions may be briefly noted. Sometimes, where all variables in a model are dichotomous, there are no pathway cases, i.e. no cases of type G or H (in Table 28.2). This is known as the “empty cell” problem, or a problem of severe causal multicollinearity. The universe of observational data does not always oblige us with cases that allow us to independently test a given hypothesis. Where variables are continuous, the analogous problem is that of a causal variable of interest ( $X_1$ ) that

has only minimal effects on the outcome of interest. That is, its role in the general model is quite minor. In these situations, the only cases that are strongly affected by  $X_1$ —if there are any at all—may be extreme outliers, and these sorts of cases are not properly regarded as providing confirmatory evidence for a proposition, for reasons that are abundantly clear by now.

Finally, it should be clarified that the identification of a causal pathway case does not obviate the utility of exploring other cases. One might, for example, want to compare both sorts of potential pathway cases—G and H—with each other. Many other combinations suggest themselves. However, this sort of multi-case investigation moves beyond the logic of the causal-pathway case.

## 8 MOST-SIMILAR CASES

The most-similar method employs a minimum of two cases.<sup>16</sup> In its purest form, the chosen pair of cases is similar in all respects except the variable(s) of interest. If the study is *exploratory* (i.e. hypothesis generating), the researcher looks for cases that differ on the outcome of theoretical interest but are similar on various factors that might have contributed to that outcome, as illustrated in Table 28.3 (A). This is a common form of case selection at the initial stage of research. Often, fruitful analysis begins with an apparent anomaly: two cases are apparently quite similar, and yet demonstrate surprisingly different outcomes. The hope is that intensive study of these cases will reveal one—or at most several—factors that differ across these cases. These differing factors ( $X_1$ ) are looked upon as putative causes. At this stage, the research may be described by the second diagram in Table 28.3 (B). Sometimes, a researcher begins with a strong hypothesis, in which case her research design is confirmatory (hypothesis testing) from the get-go. That is, she strives to identify cases that exhibit different outcomes, different scores on the factor of interest, and similar scores on all other possible causal factors, as illustrated in the second (hypothesis-testing) diagram in Table 28.3 (B).

The point is that the purpose of a most-similar research design, and hence its basic setup, often changes as a researcher moves from an exploratory to a confirmatory mode of analysis. However, regardless of where one begins, the results, when published, look like a hypothesis-testing research design. Question marks have been removed: (A) becomes (B) in Table 28.3.

As an example, let us consider Leon Epstein's classic study of party cohesion, which focuses on two "most-similar" countries, the United States and Canada. Canada has highly disciplined parties whose members vote together on the floor of the House of

<sup>16</sup> Sometimes, the most-similar method is known as the "method of difference," after its inventor (Mill 1843/1872). For later treatments see Cohen and Nagel (1934); Eggan (1954); Gerring (2001, ch. 9); Lijphart (1971; 1975); Meckstroth (1975); Przeworski and Teune (1970); Skocpol and Somers (1980).

Table 28.3. Most-similar analysis with two case types

Case types	$X_1$	$X_2$	$Y$
<i>(A) Hypothesis-generating (Y-centered):</i>			
A	?	0	1
B	?	0	0
<i>(B) Hypothesis-testing (<math>X_1/Y</math>-centered):</i>			
A	1	0	1
B	0	0	0

$X_1$  = the variable of theoretical interest.  $X_2$  = a vector of controls.  $Y$  = the outcome of interest.

Commons while the United States has weak, undisciplined parties, whose members often defect on floor votes in Congress. In explaining these divergent outcomes, persistent over many years, Epstein first discusses possible causal factors that are held more or less constant across the two cases. Both the United States and Canada inherited English political cultures, both have large territories and heterogeneous populations, both are federal, and both have fairly loose party structures with strong regional bases and a weak center. These are the "control" variables. Where they differ is in one constitutional feature: Canada is parliamentary while the United States is presidential. And it is this institutional difference that Epstein identifies as the crucial (differentiating) cause. (For further examples of the most-similar method see Brenner 1976; Hamilton 1977; Lipset 1968; Miguel 2004; Moulder 1977; Posner 2004.)

Several caveats apply to any most-similar analysis (in addition to the usual set of assumptions applying to all case-study analysis). First, each causal factor is understood as having an independent and additive effect on the outcome; there are no "interaction" effects. Second, one must code cases dichotomously (high/low, present/absent). This is straightforward if the underlying variables are also dichotomous (e.g. federal/unitary). However, it is often the case that variables of concern in the model are continuous (e.g. party cohesion). In this setting, the researcher must "dichotomize" the scoring of cases so as to simplify the two-case analysis. (Some flexibility is admissible on the vector of controls ( $X_2$ ) that are "held constant" across the cases. Nonidentity is tolerable if the deviation runs counter to the predicted hypothesis. For example, Epstein describes both the United States and Canada as having strong regional bases of power, a factor that is probably more significant in recent Canadian history than in recent American history. However, because regional bases of power should lead to weaker parties, rather than stronger parties, this element of nonidentity does not challenge Epstein's conclusions. Indeed, it sets up a most-difficult research scenario, as discussed above.)

In one respect the requirements for case control are not so stringent. Specifically, it is not usually necessary to *measure* control variables (at least not with a high degree of precision) in order to control for them. If two countries can be assumed

to have similar cultural heritages one needn't worry about constructing variables to measure that heritage. One can simply assert that, whatever they are, they are more or less constant across the two cases. This is similar to the technique employed in a randomized experiment, where the researcher typically does not attempt to measure all the factors that might affect the causal relationship of interest. She assumes, rather, that these unknown factors have been neutralized across the treatment and control groups by randomization or by the choice of a sample that is internally homogeneous.

The most useful statistical tool for identifying cases for in-depth analysis in a most-similar setting is probably some variety of matching strategy—e.g. exact matching, approximate matching, or propensity-score matching.<sup>17</sup> The product of this procedure is a set of matched cases that can be compared in whatever way the researcher deems appropriate. These are the “most-similar” cases. Rosenbaum and Silber (2001, 223) summarize:

Unlike model-based adjustments, where [individuals] vanish and are replaced by the coefficients of a model, in matching, ostensibly comparable patterns are compared directly, one by one. Modern matching methods involve statistical modeling and combinatorial algorithms, but the end result is a collection of pairs or sets of people who look comparable, at least on average. In matching, people retain their integrity as people, so they can be examined and their stories can be told individually.

Matching, conclude the authors, “facilitates, rather than inhibits, thick description” (Rosenbaum and Silber 2001, 223).

In principle, the same matching techniques that have been used successfully in observational studies of medical treatments might also be adapted to the study of nation states, political parties, cities, or indeed any traditional paired cases in the social sciences. Indeed, the current popularity of matching among statisticians—relative, that is, to garden-variety regression models—rests upon what qualitative researchers would recognize as a “case-based” approach to causal analysis. If Rosenbaum and Silber are correct, it may be perfectly reasonable to appropriate this large- $N$  method of analysis for case-study purposes.

As with other methods of case selection, the most-similar method is prone to problems of nonrepresentativeness. If employed in a qualitative fashion (without a systematic cross-case selection strategy), potential biases in the chosen case must be addressed in a speculative way. If the researcher employs a matching technique of case selection within a large- $N$  sample, the problem of potential bias can be addressed by assuring the choice of cases that are not extreme outliers, as judged by their residuals in the full model. Most-similar cases should also be “typical” cases, though some scope for deviance around the regression line may be acceptable for purposes of finding a good fit among cases.

<sup>17</sup> For good introductions see Ho et al. (2004); Morgan and Harding (2005); Rosenbaum (2004); Rosenbaum and Silber (2001). For a discussion of matching procedures in Stata see Abadie et al. (2001).

Table 28.4. Most-different analysis with two cases

Case types	$X_1$	$X_{2a}$	$X_{2b}$	$X_{2c}$	$X_{2d}$	$Y$
A	1	1	0	1	0	1
B	1	0	1	0	1	1

$X_1$  = the variable of theoretical interest.  $X_{2a-d}$  = a vector of controls.  
 $Y$  = the outcome of interest.

## 9 MOST-DIFFERENT CASES

A final case-selection method is the reverse image of the previous method. Here, variation on independent variables is prized, while variation on the outcome is eschewed. Rather than looking for cases that are most-similar, one looks for cases that are *most-different*. Specifically, the researcher tries to identify cases where just one independent variable ( $X_1$ ), as well as the dependent variable ( $Y$ ), covary, while all other plausible factors ( $X_{2a-d}$ ) show different values.<sup>18</sup>

The simplest form of this two-case comparison is illustrated in Table 28.4. Cases A and B are deemed “most different,” though they are similar in two essential respects—the causal variable of interest and the outcome.

As an example, I follow Marc Howard's (2003) recent work, which explores the enduring impact of Communism on civil society.<sup>19</sup> Cross-national surveys show a strong correlation between former Communist regimes and low social capital, controlling for a variety of possible confounders. It is a strong result. Howard wonders why this relationship is so strong and why it persists, and perhaps even strengthens, in countries that are no longer socialist or authoritarian. In order to answer this question, he focuses on two most-different cases, Russia and East Germany. These two countries were quite different—in all ways *other than* their Communist experience—prior to the Soviet era, during the Soviet era (since East Germany received substantial subsidies from West Germany), and in the post-Soviet era, as East Germany was absorbed into West Germany. Yet, they both score near the bottom of various cross-national indices intended to measure the prevalence of civic engagement in the current era. Thus, Howard's (2003, 6–9) case selection procedure meets the requirements of the most-different research design: Variance is found on all (or most) dimensions

<sup>18</sup> The most-different method is also sometimes referred to as the “method of agreement,” following its inventor, J. S. Mill (1843/1872). See also DeFelice (1986); Gerring (2001, 212–14); Lijphart (1971; 1975); Meckstroth (1975); Przeworski and Teune (1970); Skocpol and Somers (1980). For examples of this method see Collier and Collier (1991/2002); Converse and Dupeux (1962); Karl (1997); Moore (1966); Skocpol (1979); Yashar (2005, 23). However, most of these studies are described as *combining* most-similar and most-different methods.

<sup>19</sup> In the following discussion I treat the terms social capital, civil society, and civic engagement interchangeably.

aside from the key factor of interest (Communism) and the outcome (civic engagement).

What leverage is brought to the analysis from this approach? Howard's case studies combine evidence drawn from mass surveys and from in-depth interviews of small, stratified samples of Russians and East Germans. (This is a good illustration, incidentally, of how quantitative and qualitative evidence can be fruitfully combined in the intensive study of several cases.) The product of this analysis is the identification of three causal pathways that, Howard (2003, 122) claims, help to explain the lagged status of civil society in post-Communist polities: "the mistrust of communist organizations, the persistence of friendship networks, and the disappointment with post-communism." Simply put, Howard (2003, 145) concludes, "a great number of citizens in Russia and Eastern Germany feel a strong and lingering sense of distrust of any kind of public organization, a general satisfaction with their own personal networks (accompanied by a sense of deteriorating relations within society overall), and disappointment in the developments of post-communism."

The strength of this most-different case analysis is that the results obtained in East Germany and Russia should also apply in other post-Communist polities (e.g. Lithuania, Poland, Bulgaria, Albania). By choosing a heterogeneous sample, Howard solves the problem of representativeness in his restricted sample. However, this sample is demonstrably *not* representative across the population of the inference, which is intended to cover all countries of the world.

More problematic is the lack of variation on key causal factors of interest—Communism and its putative causal pathways. For this reason, it is difficult to reach conclusions about the causal status of these factors on the basis of the most-different analysis alone. It is possible, that is, that the three causal pathways identified by Howard also operate within polities that never experienced Communist rule.

Nor does it seem possible to conclusively eliminate rival hypotheses on the basis of this most-different analysis. Indeed, this is not Howard's intention. He wishes merely to show that whatever influence on civil society might be attributed to economic, cultural, and other factors does not exhaust this subject.

My considered judgment is that the most-different research design provides minimal leverage into the problem of why Communist systems appear to suppress civic engagement, years after their disappearance. Fortunately, this is not the only research design employed by Howard in his admirable study. Indeed, the author employs two other small-*N* cross-case methods, as well as a large-*N* cross-country statistical analysis. These methods do most of the analytic work. East Germany may be regarded as a causal pathway case (see above). It has all the attributes normally assumed to foster civic engagement (e.g. a growing economy, multiparty competition, civil liberties, a free press, close association with Western European culture and politics), but nonetheless shows little or no improvement on this dimension during the post-transition era (Howard 2003, 8). It is plausible to attribute this lack of change to its Communist past, as Howard does, in which case East Germany should be a fruitful case for the investigation of causal mechanisms. The contrast between East and West Germany provides a most-similar analysis since the two polities share

virtually everything except a Communist past. This variation is also deftly exploited by Howard.

I do not wish to dismiss the most-different research method entirely. Surely, Howard's findings are stronger with the intensive analysis of Russia than they would be without. Yet his book would not stand securely on the empirical foundation provided by most-different analysis alone. If one strips away the pathway-case (East Germany) and the most-similar analysis (East/West Germany) there is little left upon which to base an analysis of causal relations (aside from the large-*N* cross-national analysis). Indeed, most scholars who employ the most-different method do so in conjunction with other methods.<sup>20</sup> It is rarely, if ever, a standalone method.<sup>21</sup>

Generalizing from this discussion of Marc Howard's work, I offer the following summary remarks on the most-different method of case analysis. (I leave aside issues faced by *all* case-study analyses, issues that are explored in Gerring 2007.)

Let us begin with a methodological obstacle that is faced by both Millean styles of analysis—the necessity of dichotomizing every variable in the analysis. Recall that, as with most-similar analysis, differences across cases must generally be sizeable enough to be interpretable in an essentially dichotomous fashion (e.g. high/low, present/absent) and similarities must be close enough to be understood as essentially identical (e.g. high/high, present/present). Otherwise the results of a Millean style analysis are not interpretable. The problem of "degrees" is deadly if the variables under consideration are, by nature, continuous (e.g. GDP). This is a particular concern in Howard's analysis, where East Germany scores somewhat higher than Russia in civic engagement; they are both low, but Russia is quite a bit lower. Howard assumes that this divergence is minimal enough to be understood as a difference of degrees rather than of kinds, a judgment that might be questioned. In these respects, most-different analysis is no more secure—but also no less—than most-similar analysis.

In one respect, most-different analysis is superior to most-similar analysis. If the coding assumptions are sound, the most-different research design may be quite useful for *eliminating necessary causes*. Causal factors that do not appear across the chosen cases—e.g.  $X_{2a-d}$  in Table 28.4—are evidently unnecessary for the production of *Y*. However, it does not follow that the most-different method is the *best* method for eliminating necessary causes. Note that the defining feature of this method is the

<sup>20</sup> E.g. Collier and Collier (1991/2002); Karl (1997); Moore (1966); Skocpol (1979); Yashar (2005, 23). Karl (1997), which affects to be a most-different system analysis (20), is a particularly clear example of this. Her study, focused ostensibly on petro-states (states with large oil reserves), makes two sorts of inferences. The first concerns the (usually) obstructive role of oil in political and economic development. The second sort of inference concerns variation *within* the population of petro-states, showing that some countries (e.g. Norway, Indonesia) manage to avoid the pathologies brought on elsewhere by oil resources. When attempting to explain the constraining role of oil on petro-states, Karl usually relies on contrasts between petro-states and nonpetro-states (e.g. ch. 10). Only when attempting to explain differences among petro-states does she restrict her sample to petro-states. In my opinion, very little use is made of the most-different research design.

<sup>21</sup> This was recognized, at least implicitly, by Mill (1843/1872, 258–9). Skepticism has been echoed by methodologists in the intervening years (e.g. Cohen and Nagel 1934, 251–6; Gerring 2001; Skocpol and Somers 1980). Indeed, explicit defenses of the most-different method are rare (but see DeFelice 1986).



shared element across cases— $X_1$  in Table 28.4. This feature does not help one to eliminate necessary causes. Indeed, if one were focused solely on eliminating necessary causes one would presumably seek out cases that register the same outcomes and have maximum diversity on other attributes. In Table 28.4, this would be a set of cases that satisfy conditions  $X_{2a-d}$ , but not  $X_1$ . Thus, even the presumed strength of the most-different analysis is not so strong.

Usually, case-study analysis is focused on the identification (or clarification) of causal relations, not the elimination of possible causes. In this setting, the most-different technique is useful, but only if assumptions of causal uniqueness hold. By “causal uniqueness,” I mean a situation in which a given outcome is the product of only one cause:  $Y$  cannot occur except in the presence of  $X$ .  $X$  is necessary, and in some situations (given certain background conditions) sufficient, to cause  $Y$ .<sup>22</sup>

Consider the following hypothetical example. Suppose that a new disease, about which little is known, has appeared in Country A. There are hundreds of infected persons across dozens of affected communities in that country. In Country B, located at the other end of the world, several new cases of the disease surface in a single community. In this setting, we can imagine two sorts of Milleean analyses. The first examines two similar communities within Country A, one of which has developed the disease and the other of which has not. This is the most-similar style of case comparison, and focuses accordingly on the identification of a difference between the two cases that might account for variation across the sample. A second approach focuses on communities where the disease has appeared across the two countries and searches for any similarities that might account for these similar outcomes. This is the most-different research design.

Both are plausible approaches to this particular problem, and we can imagine epidemiologists employing them simultaneously. However, the most-different design demands stronger assumptions about the underlying factors at work. It supposes that the disease arises from the *same cause* in any setting. This is often a reasonable operating assumption when one is dealing with natural phenomena, though there are certainly many exceptions. Death, for example, has many causes. For this reason, it would not occur to us to look for most-different cases of high mortality around the world. In order for the most-different research design to effectively identify a causal factor at work in a given outcome, the researcher must assume that  $X_1$ —the factor held constant across the diverse cases—is the only possible cause of  $Y$  (see Table 28.4). This assumption rarely holds in social-scientific settings. Most outcomes of interest to anthropologists, economists, political scientists, and sociologists have *multiple* causes. There are many ways to win an election, to build a welfare state, to get into a war, to overthrow a government, or—returning to Marc Howard’s work—to build a strong civil society. And it is for this reason that most-different analysis is rarely applied in social science work and, where applied, is rarely convincing.

If this seems a tad severe, there is a more charitable way of approaching the most-different method. Arguably, this is not a pure “method” at all but merely a

<sup>22</sup> Another way of stating this is to say that  $X$  is a “nontrivial necessary condition” of  $Y$ .

supplement, a way of incorporating diversity in the sub-sample of cases that provide the unusual outcome of interest. If the unusual outcome is revolutions, one might wish to encompass a wide variety of revolutions in one’s analysis. If the unusual outcome is post-Communist civil society, it seems appropriate to include a diverse set of post-Communist polities in one’s sample of case studies, as Marc Howard does. From this perspective, the most-different method (so-called) might be better labeled a *diverse-case* method, as explored above.

## 10 CONCLUSIONS

In order to be a case of something broader than itself, the chosen case must be representative (in some respects) of a larger population. Otherwise—if it is purely idiosyncratic (“unique”)—it is uninformative about anything lying outside the borders of the case itself. A study based on a nonrepresentative sample has no (or very little) external validity. To be sure, no phenomenon is purely idiosyncratic; the notion of a unique case is a matter that would be difficult to define. One is concerned, as always, with matters of degree. Cases are *more or less* representative of some broader phenomenon and, on that score, may be considered better or worse subjects for intensive analysis. (The one exception, as noted, is the influential case.)

Of all the problems besetting case-study analysis, perhaps the most persistent—and the most persistently bemoaned—is the problem of sample bias (Achen and Snidal 1989; Collier and Mahoney 1996; Geddes 1990; King, Keohane, and Verba 1994; Rohlfing 2004; Sekhon 2004). Lisa Martin (1992, 5) finds that the overemphasis of international relations scholars on a few well-known cases of economic sanctions—most of which failed to elicit any change in the sanctioned country—“has distorted analysts’ view of the dynamics and characteristics of economic sanctions.” Barbara Geddes (1990) charges that many analyses of industrial policy have focused exclusively on the most successful cases—primarily the East Asian NICs—leading to biased inferences. Anna Breman and Carolyn Shelton (2001) show that case-study work on the question of structural adjustment is systematically biased insofar as researchers tend to focus on disaster cases—those where structural adjustment is associated with very poor health and human development outcomes. These cases, often located in sub-Saharan Africa, are by no means representative of the entire population. Consequently, scholarship on the question of structural adjustment is highly skewed in a particular ideological direction (against neoliberalism) (see also Gerring, Thacker, and Moreno 2005).

These examples might be multiplied many times. Indeed, for many topics the most-studied cases are acknowledged to be less than representative. It is worth reflecting upon the fact that our knowledge of the world is heavily colored by a few “big” (populous, rich, powerful) countries, and that a good portion of the disciplines of economics, political science, and sociology are built upon scholars’ familiarity with

the economics, political science, and sociology of one country, the United States.<sup>23</sup> Case-study work is particularly prone to problems of investigator bias since so much rides on the researcher's selection of one (or a few) cases. Even if the investigator is unbiased, her sample may still be biased simply by virtue of "random" error (which may be understood as measurement error, error in the data-generation process, or as an underlying causal feature of the universe).

There are only two situations in which a case-study researcher need not be concerned with the representativeness of her chosen case. The first is the influential case research design, where a case is chosen because of its possible influence on a cross-case model, and hence is not expected to be representative of a larger sample. The second is the deviant-case method, where the chosen case is employed to confirm a broader cross-case argument to which the case stands as an apparent exception. Yet even here the chosen case is expected to be representative of a broader set of cases—those, in particular, that are poorly explained by the extant model.

In all other circumstances, cases must be representative of the population of interest in whatever ways might be relevant to the proposition in question. Note that where a researcher is attempting to disconfirm a deterministic proposition the question of representativeness is perhaps more appropriately understood as a question of classification: Is the chosen case appropriately classified as a member of the designated population? If so, then it is fodder for a disconfirming case study.

If the researcher is attempting to confirm a deterministic proposition, or to make probabilistic arguments about a causal relationship, then the problem of representativeness is of the more usual sort: Is case A unit-homogeneous relative to other cases in the population? This is not an easy matter to test. However, in a large-*N* context the residual for that case (in whatever model the researcher has greatest confidence in) is a reasonable place to start. Of course, this test is only as good as the model at hand. Any incorrect specifications or incorrect modeling procedures will likely bias the results and give an incorrect assessment of each case's "typicality." In addition, there is the possibility of stochastic error, errors that cannot be modeled in a general framework. Given the explanatory weight that individual cases are asked to bear in a case-study analysis, it is wise to consider more than just the residual test of representativeness. Deductive logic and an in-depth knowledge of the case in question are often more reliable tools than the results of a cross-case model.

In any case, there is no dispensing with the question. Case studies (with the two exceptions already noted) rest upon an assumed synecdoche: The case should stand for a population. If this is not true, or if there is reason to doubt this assumption, then the utility of the case study is brought severely into question.

Fortunately, there is some safety in numbers. Insofar as case-study evidence is combined with cross-case evidence the issue of sample bias is mitigated. Indeed, the suspicion of case-study work that one finds in the social sciences today is, in my view, a product of a too-literal interpretation of the case-study method. A case study *tout*

<sup>23</sup> Wahlke (1979, 13) writes of the failings of the "behavioralist" mode of political science analysis: "It rarely aims at generalization; research efforts have been confined essentially to case studies of single political systems, most of them dealing . . . with the American system."

*court* is thought to mean a case study *tout seul*. Insofar as case studies and cross-case studies can be enlisted within the same investigation (either in the same study or by reference to other studies in the same subfield), problems of representativeness are less worrisome. This is the virtue of cross-level work, a.k.a. "triangulation."

## 11 AMBIGUITIES

Before concluding, I wish to draw attention to two ambiguities in case-selection strategies in case-study research. The first concerns the admixture of several case-selection strategies. The second concerns the changing status of a case as a study proceeds.

Some case studies follow only one strategy of case selection. They are *typical*, *diverse*, *extreme*, *deviant*, *influential*, *crucial*, *pathway*, *most-similar*, or *most-different* research designs, as discussed. However, many case studies mix and match among these case-selection strategies. Indeed, insofar as all case studies seek representative samples, they are always in search of "typical" cases. Thus, it is common for writers to declare that their case is, for example, both extreme and typical; it has an extreme value on  $X_1$  or  $Y$  but is not, in other respects, idiosyncratic. There is not much that one can say about these combinations of strategies except that, where the cases allow for a variety of empirical strategies, there is no reason not to pursue them. And where the same cases can serve several functions at once (without further effort on the researcher's part), there is little cost to a multi-pronged approach to case analysis.

The second issue that deserves emphasis is the changing status of a case during the course of a researcher's investigation—which may last for years, if not decades. The problem is acute wherever a researcher begins in an exploratory mode and proceeds to hypothesis-testing (that is, she develops a specific  $X_1/Y$  proposition) or where the operative hypothesis or key control variable changes (a new causal factor is discovered or another outcome becomes the focus of analysis). Things change. And it is the mark of a good researcher to keep her mind open to new evidence and new insights. Too often, methodological discussions give the misleading impression that hypotheses are clear and remain fixed over the course of a study's development. Nothing could be further from the truth. The unofficial transcripts of academia—accessible in informal settings, where researchers let their guards down (particularly if inebriated)—are filled with stories about dead-ends, unexpected findings, and drastically revised theory chapters. It would be interesting, in this vein, to compare published work with dissertation prospectuses and fellowship applications. I doubt if the correlation between these two stages of research is particularly strong.

Research, after all, is about discovery, not simply the verification or falsification of static hypotheses. That said, it is also true that research on a particular topic should move from hypothesis generating to hypothesis-testing. This marks the progress of a

field, and of a scholar's own work. As a rule, research that begins with an open-ended (*X*- or *Y*-centered) analysis should conclude with a determinate  $X_1/Y$  hypothesis.

The problem is that research strategies that are ideal for exploration are not always ideal for confirmation. The extreme-case method is inherently exploratory since there is no clear causal hypothesis; the researcher is concerned merely to explore variation on a single dimension (*X* or *Y*). Other methods can be employed in either an open-ended (exploratory) or a hypothesis-testing (confirmatory/disconfirmatory) mode. The difficulty is that once the researcher has arrived at a determinate hypothesis the originally chosen research design may no longer appear to be so well designed.

This is unfortunate, but inevitable. One cannot construct the perfect research design until (a) one has a specific hypothesis and (b) one is reasonably certain about what one is going to find "out there" in the empirical world. This is particularly true of observational research designs, but it also applies to many experimental research designs: Usually, there is a "good" (informative) finding, and a finding that is less insightful. In short, the perfect case-study research design is usually apparent only *ex post facto*.

There are three ways to handle this. One can explain, straightforwardly, that the initial research was undertaken in an exploratory fashion, and therefore not constructed to test the specific hypothesis that is—now—the primary argument. Alternatively, one can try to redesign the study after the new (or revised) hypothesis has been formulated. This may require additional field research or perhaps the integration of additional cases or variables that can be obtained through secondary sources or through consultation of experts. A final approach is to simply jettison, or de-emphasize, the portion of research that no longer addresses the (revised) key hypothesis. A three-case study may become a two-case study, and so forth. Lost time and effort are the costs of this downsizing.

In the event, practical considerations will probably determine which of these three strategies, or combinations of strategies, is to be followed. (They are not mutually exclusive.) The point to remember is that revision of one's cross-case research design is *normal* and perhaps to be expected. Not all twists and turns on the meandering trail of truth can be anticipated.

## 12 ARE THERE OTHER METHODS OF CASE SELECTION?

At the outset of this chapter I summarized the task of case selection as a matter of achieving two objectives: representativeness (typicality) and variation (causal leverage). Evidently, there are other objectives as well. For example, one wishes to identify cases that are *independent* of each other. If chosen cases are affected by

each other (sometimes known as Galton's problem or a problem of diffusion), this problem must be corrected before analysis can take place. I have neglected this issue because it is usually apparent to the researcher and, in any case, there are no simple techniques that might be utilized to correct for such biases. (For further discussion of this and other factors impinging upon case selection see Gerring 2001, 178–81.)

I have also disregarded *pragmatic/logistical* issues that might affect case selection. Evidently, case selection is often influenced by a researcher's familiarity with the language of a country, a personal entrée into that locale, special access to important data, or funding that covers one archive rather than another. Pragmatic considerations are often—and quite rightly—decisive in the case-selection process.

A final consideration concerns the *theoretical prominence* of a particular case within the literature on a subject. Researchers are sometimes obliged to study cases that have received extensive attention in previous studies. These are sometimes referred to as "paradigmatic" cases or "exemplars" (Flyvbjerg 2004, 427).

However, neither pragmatic/logistical utility nor theoretical prominence qualifies as a *methodological* factor in case selection. That is, these features of a case have no bearing on the validity of the findings stemming from a study. As such, it is appropriate to grant these issues a peripheral status in this chapter.

One final caveat must be issued. While it is traditional to distinguish among the tasks of case selection and case analysis, a close look at these processes shows them to be indistinct and overlapping. One cannot choose a case without considering the sort of analysis that it might be subjected to, and vice versa. Thus, the reader should consider choosing cases by employing the nine techniques laid out in this chapter *along with* any considerations that might be introduced by virtue of a case's quasi-experimental qualities, a topic taken up elsewhere (Gerring 2007, ch. 6).

## REFERENCES

- ABADIE, A., DRUKKER, D., HERR, J. L., and IMBENS, G. W. 2001. Implementing matching estimators for average treatment effects in Stata. *Stata Journal*, 1: 1–18.
- ABBOTT, A. 2001. *Time Matters: On Theory and Method*. Chicago: University of Chicago Press.
- and TSAY, A. 2000. Sequence analysis and optimal matching methods in sociology. *Sociological Methods and Research*, 29: 3–33.
- and FORREST, J. 1986. Optimal matching methods for historical sequences. *Journal of Interdisciplinary History*, 16: 471–94.
- ACHEN, C. H., and SNIDAL, D. 1989. Rational deterrence theory and comparative case studies. *World Politics*, 41: 143–69.
- ALLEN, W. S. 1965. *The Nazi Seizure of Power: The Experience of a Single German Town, 1930–1935*. New York: Watts.
- ALMOND, G. A. 1956. Comparative political systems. *Journal of Politics*, 18: 391–409.

- AMENTA, E. 1991. Making the most of a case study: theories of the welfare state and the American experience. Pp. 172–94 in *Issues and Alternatives in Comparative Social Research*, ed. C. C. Ragin. Leiden: E. J. Brill.
- BARRO, R. J. 1999. Determinants of democracy. *Journal of Political Economy*, 107: 158–83.
- BELSEY, D. A., KUH, E., and WELSCH, R. E. 2004. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.
- BENNETT, A., LEPGOLD, J., and UNGER, D. 1994. Burden-sharing in the Persian Gulf War. *International Organization*, 48: 39–75.
- BENTLEY, A. 1908/1967. *The Process of Government*. Cambridge, Mass.: Harvard University Press.
- BRADY, H. E., and COLLIER, D. (eds.) 2004. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, Md.: Rowman and Littlefield.
- BRAUMOELLER, B. F. 2003. Causal complexity and the study of politics. *Political Analysis*, 11: 209–33.
- BREMAN, A., and SHELTON, C. 2001. Structural adjustment and health: a literature review of the debate, its role-players and presented empirical evidence. CMH Working Paper Series, Paper No. WG6: 6. WHO, Commission on Macroeconomics and Health.
- BRENNER, R. 1976. Agrarian class structure and economic development in pre-industrial Europe. *Past and Present*, 70: 30–75.
- BROWNE, A. 1987. *When Battered Women Kill*. New York: Free Press.
- BUCHBINDER, S., and VITTINGHOFF, E. 1999. HIV-infected long-term nonprogressors: epidemiology, mechanisms of delayed progression, and clinical and research implications. *Microbes Infect*, 1: 1113–20.
- COHEN, M. R., and NAGEL, E. 1934. *An Introduction to Logic and Scientific Method*. New York: Harcourt, Brace and Company.
- COLLIER, D., and MAHONEY, J. 1996. Insights and pitfalls: selection bias in qualitative research. *World Politics*, 49: 56–91.
- COLLIER, R. B., and COLLIER, D. 1991/2002. *Shaping the Political Arena: Critical Junctures, the Labor Movement, and Regime Dynamics in Latin America*. Notre Dame, Ind.: University of Notre Dame Press.
- COLOMER, J. M. 1991. Transitions by agreement: modeling the Spanish way. *American Political Science Review*, 85: 1283–302.
- CONVERSE, P. E., and DUPEUX, G. 1962. Politicization of the electorate in France and the United States. *Public Opinion Quarterly*, 16: 1–23.
- COPPEDGE, M. J. 2004. The conditional impact of the economy on democracy in Latin America. Presented at the conference “Democratic Advancements and Setbacks: What Have We Learnt?,” Uppsala University, June 11–13.
- DEFELICE, E. G. 1986. Causal inference and comparative methods. *Comparative Political Studies*, 19: 415–37.
- DESCH, M. C. 2002. Democracy and victory: why regime type hardly matters. *International Security*, 27: 5–47.
- DEYO, F. (ed.) 1987. *The Political Economy of the New Asian Industrialism*. Ithaca, NY: Cornell University Press.
- DION, D. 1998. Evidence and inference in the comparative case study. *Comparative Politics*, 30: 127–45.
- ECKSTEIN, H. 1975. Case studies and theory in political science. In *Handbook of Political Science*, vii: *Political Science: Scope and Theory*, ed. F. I. Greenstein and N. W. Polsby. Reading, Mass.: Addison-Wesley.

- EGGAN, F. 1954. Social anthropology and the method of controlled comparison. *American Anthropologist*, 56: 743–63.
- ELMAN, C. 2003. Lessons from Lakatos. In *Progress in International Relations Theory: Appraising the Field*, ed. C. Elman and M. F. Elman. Cambridge, Mass.: MIT Press.
- 2005. Explanatory typologies in qualitative studies of international politics. *International Organization*, 59: 293–326.
- EMIGH, R. 1997. The power of negative thinking: the use of negative case methodology in the development of sociological theory. *Theory and Society*, 26: 649–84.
- EPSTEIN, L. D. 1964. A comparative study of Canadian parties. *American Political Science Review*, 58: 46–59.
- ERTMAN, T. 1997. *Birth of the Leviathan: Building States and Regimes in Medieval and Early Modern Europe*. Cambridge: Cambridge University Press.
- ESPING-ANDERSEN, G. 1990. *The Three Worlds of Welfare Capitalism*. Princeton, NJ: Princeton University Press.
- FLYVBJERG, B. 2004. Five misunderstandings about case-study research. Pp. 420–34 in *Qualitative Research Practice*, ed. C. Seale, G. Gobo, J. F. Gubrium, and D. Silverman. London: Sage.
- GEDDES, B. 1990. How the cases you choose affect the answers you get: selection bias in comparative politics. In *Political Analysis*, vol. ii, ed. J. A. Stimson. Ann Arbor: University of Michigan Press.
- 2003. *Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics*. Ann Arbor: University of Michigan Press.
- GEORGE, A. L., and BENNETT, A. 2005. *Case Studies and Theory Development*. Cambridge, Mass.: MIT Press.
- and SMOKE, R. 1974. *Deterrence in American Foreign Policy: Theory and Practice*. New York: Columbia University Press.
- GERRING, J. 2001. *Social Science Methodology: A Criterial Framework*. Cambridge: Cambridge University Press.
- 2007. *Case Study Research: Principles and Practices*. Cambridge: Cambridge University Press.
- THACKER, S., and MORENO, C. 2005. Do neoliberal policies save lives? Unpublished manuscript.
- GOERTZ, G., and STARR, H. (eds.) 2003. *Necessary Conditions: Theory, Methodology and Applications*. New York: Rowman and Littlefield.
- and LEVY, J. (eds.) forthcoming. Causal explanations, necessary conditions, and case studies: World War I and the end of the Cold War. Manuscript.
- GOODIN, R. E., and SMITSMAN, A. 2000. Placing welfare states: the Netherlands as a crucial test case. *Journal of Comparative Policy Analysis*, 2: 39–64.
- GUJARATI, D. N. 2003. *Basic Econometrics*, 4th edn. New York: McGraw-Hill.
- HAMILTON, G. G. 1977. Chinese consumption of foreign commodities: a comparative perspective. *American Sociological Review*, 42: 877–91.
- HAYNES, B. F., PANTALEO, G., and FAUCI, A. S. 1996. Toward an understanding of the correlates of protective immunity to HIV infection. *Science*, 271: 324–8.
- HEMPEL, C. G. 1942. The function of general laws in history. *Journal of Philosophy*, 39: 35–48.
- HO, D. E., IMAI, K., KING, G., and STUART, E. A. 2004. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Manuscript.
- HOWARD, M. M. 2003. *The Weakness of Civil Society in Post-Communist Europe*. Cambridge: Cambridge University Press.

- HOWSON, C., and URBACH, P. 1989. *Scientific Reasoning: The Bayesian Approach*. La Salle, Ill.: Open Court.
- HUMPHREYS, M. 2005. Natural resources, conflict, and conflict resolution: uncovering the mechanisms. *Journal of Conflict Resolution*, 49: 508–37.
- JENICEK, M. 2001. *Clinical Case Reporting in Evidence-Based Medicine*, 2nd edn. Oxford: Oxford University Press.
- KARL, T. L. 1997. *The Paradox of Plenty: Oil Booms and Petro-states*. Berkeley: University of California Press.
- KAZANCIGIL, A. 1994. The deviant case in comparative analysis: high stateness in comparative analysis. Pp. 213–38 in *Comparing Nations: Concepts, Strategies, Substance*, ed. M. Dogan and A. Kazancigil. Cambridge: Blackwell.
- KEMP, K. A. 1986. Race, ethnicity, class and urban spatial conflict: Chicago as a crucial case. *Urban Studies*, 23: 197–208.
- KENDALL, P. L., and WOLF, K. M. 1949/1955. The analysis of deviant cases in communications research. In *Communications Research, 1948–1949*, ed. P. F. Lazarsfeld and F. N. Stanton. New York: Harper and Brothers. Reprinted as pp. 167–70 in *The Language of Social Research*, ed. P. F. Lazarsfeld and M. Rosenberg. New York: Free Press.
- KENNEDY, C. H. 2005. *Single-case Designs for Educational Research*. Boston: Allyn and Bacon.
- KENNEDY, P. 2003. *A Guide to Econometrics*, 5th edn. Cambridge, Mass.: MIT Press.
- KHONG, Y. F. 1992. *Analogies at War: Korea, Munich, Dien Bien Phu, and the Vietnam Decisions of 1965*. Princeton, NJ: Princeton University Press.
- KING, G., KEOHANE, R. O., and VERBA, S. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
- LAKATOS, I. 1978. *The Methodology of Scientific Research Programmes*. Cambridge: Cambridge University Press.
- LAZARSPFELD, P. F., and BARTON, A. H. 1951. Qualitative measurement in the social sciences: classification, typologies, and indices. In *The Policy Sciences*, ed. D. Lerner and H. D. Lasswell. Stanford, Calif.: Stanford University Press.
- LEVY, J. S. 2002. Qualitative methods in international relations. In *Evaluating Methodology in International Studies*, ed. F. P. Harvey and M. Brecher. Ann Arbor: University of Michigan Press.
- LIJPHART, A. 1968. *The Politics of Accommodation: Pluralism and Democracy in the Netherlands*. Berkeley: University of California Press.
- 1969. Consociational democracy. *World Politics*, 21: 207–25.
- 1971. Comparative politics and the comparative method. *American Political Science Review*, 65: 682–93.
- 1975. The comparable cases strategy in comparative research. *Comparative Political Studies*, 8: 158–77.
- LIPSET, S. M. 1959. Some social requisites of democracy: economic development and political development. *American Political Science Review*, 53: 69–105.
- 1960/1963. *Political Man: The Social Bases of Politics*. Garden City, NY: Anchor.
- 1968. *Agrarian Socialism: The Cooperative Commonwealth Federation in Saskatchewan. A Study in Political Sociology*. Garden City, NY: Doubleday.
- TROW, M. A., and COLEMAN, J. S. 1956. *Union Democracy: The Internal Politics of the International Typographical Union*. New York: Free Press.
- LYND, R. S., and LYND, H. M. 1929/1956. *Middletown: A Study in American Culture*. New York: Harcourt, Brace.
- MAHONEY, J., and GOERTZ, G. 2004. The possibility principle: choosing negative cases in comparative research. *American Political Science Review*, 98: 653–69.

- MARTIN, L. L. 1992. *Coercive Cooperation: Explaining Multilateral Economic Sanctions*. Princeton, NJ: Princeton University Press.
- MAYO, D. G. 1996. *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.
- MECKSTROTH, T. 1975. “Most different systems” and “most similar systems:” a study in the logic of comparative inquiry. *Comparative Political Studies*, 8: 133–77.
- MIGUEL, E. 2004. Tribe or nation: nation-building and public goods in Kenya versus Tanzania. *World Politics*, 56: 327–62.
- MILL, J. S. 1843/1872. *The System of Logic*, 8th edn. London: Longmans, Green.
- MONROE, K. R. 1996. *The Heart of Altruism: Perceptions of a Common Humanity*. Princeton, NJ: Princeton University Press.
- MOORE, B., JR. 1966. *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*. Boston: Beacon Press.
- MORGAN, S. L., and HARDING, D. J. 2005. Matching estimators of causal effects: from stratification and weighting to practical data analysis routines. Manuscript.
- MOULDER, F. V. 1977. *Japan, China and the Modern World Economy: Toward a Reinterpretation of East Asian Development ca. 1600 to ca. 1918*. Cambridge: Cambridge University Press.
- MUNCK, G. L. 2004. Tools for qualitative research. Pp. 105–21 in *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, ed. H. E. Brady and D. Collier. Lanham, Md.: Rowman and Littlefield.
- NJOLSTAD, O. 1990. Learning from history? Case studies and the limits to theory-building. Pp. 220–46 in *Arms Races: Technological and Political Dynamics*, ed. O. Njolstad. Thousand Oaks, Calif.: Sage.
- PATTON, M. Q. 2002. *Qualitative Evaluation and Research Methods*. Newbury Park, Calif.: Sage.
- POPPER, K. 1934/1968. *The Logic of Scientific Discovery*. New York: Harper and Row.
- 1963. *Conjectures and Refutations*. London: Routledge and Kegan Paul.
- POSNER, D. 2004. The political salience of cultural difference: why Chewas and Tumbukas are allies in Zambia and adversaries in Malawi. *American Political Science Review*, 98: 529–46.
- PRZEWSKI, A., and TEUNE, H. 1970. *The Logic of Comparative Social Inquiry*. New York: John Wiley.
- QUEEN, S. 1928. Round table on the case study in sociological research. *Publications of the American Sociological Society, Papers and Proceedings*, 22: 225–7.
- RAGIN, C. C. 2000. *Fuzzy-set Social Science*. Chicago: University of Chicago Press.
- 2004. Turning the tables. Pp. 123–38 in *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, ed. H. E. Brady and D. Collier. Lanham, Md.: Rowman and Littlefield.
- REILLY, B. 2000–1. Democracy, ethnic fragmentation, and internal conflict: confused theories, faulty data, and the “crucial case” of Papua New Guinea. *International Security*, 25: 162–85.
- and PHILLIPOT, R. 2003. “Making democracy work” in Papua New Guinea: social capital and provincial development in an ethnically fragmented society. *Asian Survey*, 42: 906–27.
- ROGOWSKI, R. 1995. The role of theory and anomaly in social-scientific inference. *American Political Science Review*, 89: 467–70.
- ROHLFING, I. 2004. Have you chosen the right case? Uncertainty in case selection for single case studies. Working Paper, International University, Bremen.
- ROSENBAUM, P. R. 2004. Matching in observational studies. In *Applied Bayesian Modeling and Causal Inference from an Incomplete-data Perspective*, ed. A. Gelman and X.-L. Meng. New York: John Wiley.
- and SILBER, J. H. 2001. Matching and thick description in an observational study of mortality after surgery. *Biostatistics*, 2: 217–32.
- ROSS, M. 2001. Does oil hinder democracy? *World Politics*, 53: 325–61.

- SAGAN, S. D. 1995. *Limits of Safety: Organizations, Accidents, and Nuclear Weapons*. Princeton, NJ: Princeton University Press.
- SEKHON, J. S. 2004. Quality meets quantity: case studies, conditional probability and counterfactuals. *Perspectives in Politics*, 2: 281–93.
- SHAFFER, M. D. 1988. *Deadly Paradigms: The Failure of U.S. Counterinsurgency Policy*. Princeton, NJ: Princeton University Press.
- SKOCPOL, T. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. Cambridge: Cambridge University Press.
- and SOMERS, M. 1980. The uses of comparative history in macrosocial inquiry. *Comparative Studies in Society and History*, 22: 147–97.
- STINCHCOMBE, A. L. 1968. *Constructing Social Theories*. New York: Harcourt, Brace.
- SWANK, D. H. 2002. *Global Capital, Political Institutions, and Policy Change in Developed Welfare States*. Cambridge: Cambridge University Press.
- TENDLER, J. 1997. *Good Government in the Tropics*. Baltimore: Johns Hopkins University Press.
- TRUMAN, D. B. 1951. *The Governmental Process*. New York: Alfred A. Knopf.
- TSAI, L. 2007. *Accountability without Democracy: How Solidary Groups Provide Public Goods in Rural China*. Cambridge: Cambridge University Press.
- VAN EVERA, S. 1997. *Guide to Methods for Students of Political Science*. Ithaca, NY: Cornell University Press.
- WAHLKE, J. C. 1979. Pre-behavioralism in political science. *American Political Science Review*, 73: 9–31.
- YASHAR, D. J. 2005. *Contesting Citizenship in Latin America: The Rise of Indigenous Movements and the Postliberal Challenge*. Cambridge: Cambridge University Press.
- YIN, R. K. 2004. *Case Study Anthology*. Thousand Oaks, Calif.: Sage.

## CHAPTER 29

---

INTERVIEWING  
AND QUALITATIVE  
FIELD METHODS:  
PRAGMATISM AND  
PRACTICALITIES

---

BRIAN C. RATHBUN

INTENSIVE interviewing is a powerful, but unfortunately underused tool in political science methodology. Where it is used, it is generally to add a little color to otherwise stiff accounts. Rarely do researchers talk to more than a handful of respondents. There are numerous practical reasons for this. Gaining access to interview subjects, particularly elites, is often difficult. Interviewing is costly as it often entails traveling great distances, sometimes across national borders. Interviewing often requires tremendous personal investment in language training that might not seem worth it. It is often a risky strategy. Even after the hurdles of access and travel are overcome, informants might reveal little. However, these obstacles cannot fully explain why more political scientists do not utilize interviewing in their research as a major source of data, or even as a supplement to quantitative analysis or archival records.

I maintain that there are two reasons why interviewing is often underused. First, interviewing often runs afoul of methodological tendencies in the discipline. Certain precepts of what I call the naive versions of behavioralism and rationalism make many skeptical about interviewing. Naive behavioralism objects to the status of data derived from interviewing as it is by nature subjective and imprecise, and therefore subject to