

Advance Praise for *Case Study Research*:

Prior work on the case-study method focuses centrally on data-gathering and data-aggregating techniques. This book is concerned with the underlying logic that informs or should inform – the case-study approach to social science. It addresses issues of research design that will help scholars across a wide range of disciplines. Indeed, the advice offered here will be of value to researchers who employ various methods, and not just case studies. Strongly recommended.”

– David Collier, *University of California, Berkeley*

A brilliantly compact yet rigorous guide to getting the most from case-study analysis. Would be required reading for all students planning to do comparative or single case-study research.”

– Jack A. Goldstone, *George Mason University*

Case Study Research is a superb cross-disciplinary treatment of the contribution of case studies to causal analysis in the social sciences. It provides a particularly good discussion of what a case study is, how it relates to other social science methodologies, how to evaluate the relative strengths and limitations of case study and cross-case analysis, and how to deal with some of the practical problems confronting case study methodology. The book will be essential reading for anyone who does case studies or who wants to critically evaluate their contribution to causal analysis.”

– Jack S. Levy, *Rutgers University*

Social scientists too often overlook the centrality of case studies to social research. Not only is it the most common form of social research, there are important aspects of case study research in almost all other forms. Yet the case study as a research form remains somewhat of a mystery, in part because of the diversity of purposes case studies may serve. *Case Study Research* unlocks many of the key mysteries. This cogently argued work is essential reading for all social scientists interested in understanding the importance of case studies in social research.”

– Charles Ragin, *University of Arizona*

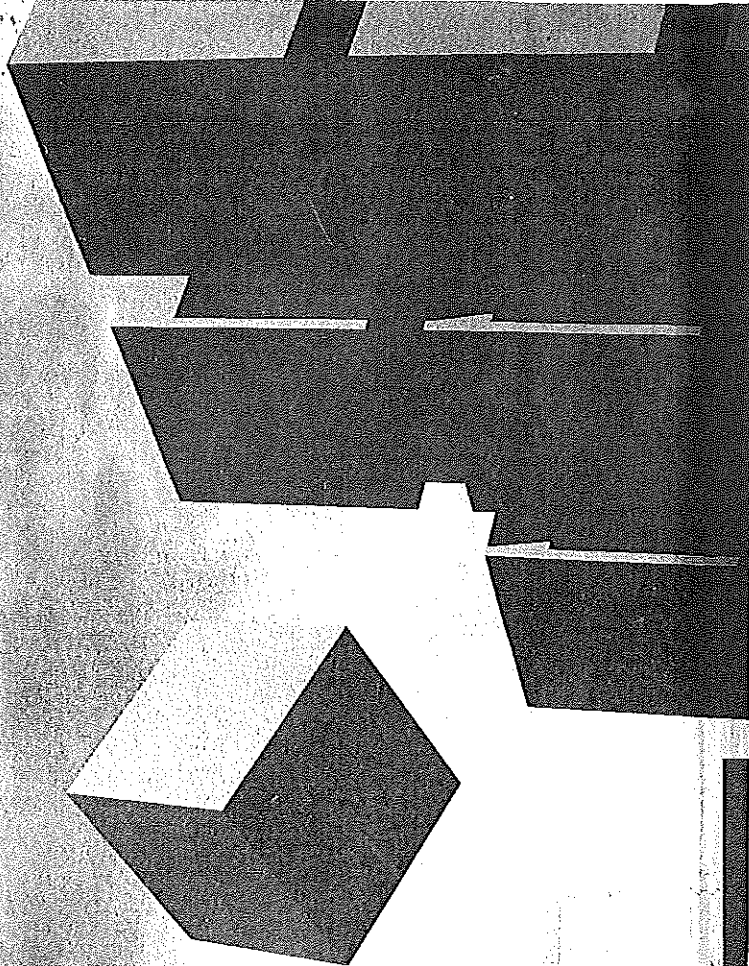
Gerring's book provides an outstanding analysis of the promise of case study research along with the tools and techniques to realize that promise. Engagingly written and elegantly argued, it makes a lasting contribution to social science methods. It should be read by substantive scholars of all methodological stripes.”

– Eric Schickler, *University of California, Berkeley*

Case Study Research

PRINCIPLES AND PRACTICES

John Gerring



Gerring

Case Study Research

CAMBRIDGE
UNIVERSITY PRESS
www.cambridge.org

ISBN 0-521-67656-8



John Gerring is currently associate professor of political science at Boston University.

LIB
301
.0722

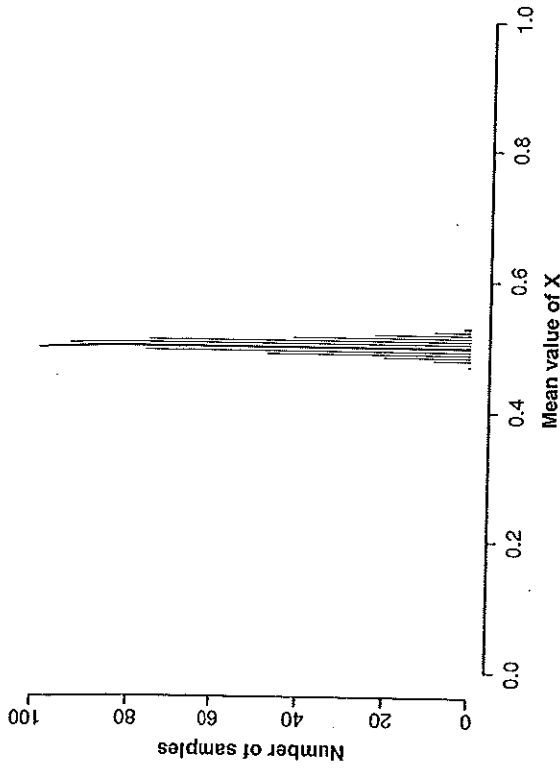


FIGURE 5.1. Sample means of large-sample draws. A histogram showing the mean values of one variable in 500 samples of 1,000 cases each. Population mean = 0.5.

However, in case study research the sample is small (by definition), and this makes randomization problematic. Consider what would happen if the sample size were changed from 1,000 cases to only 5 cases. The results are shown in Figure 5.2. On average, these small-N random samples produce the right answer, so the procedure culminates in results that are unbiased. However, many of the sample means are rather far from the population mean, and some are quite far indeed. Hence, even though this case-selection technique produces representative samples *on average*, any given sample may be wildly unrepresentative. In statistical terms, the problem is that small sample sizes tend to produce estimates with a great deal of variance – sometimes referred to as a problem of precision. For this reason, random sampling is unreliable in small-N research. (Note that in this chapter “N” refers to cases, not observations.) Moreover, there is no guarantee that a few cases, chosen randomly, will provide leverage into the research question that animates an investigation. The sample might be representative, but uninformative.

If random sampling is inappropriate as a selection method in case study research, how, then, is one to choose a sample comprised of one or several cases? Keep in mind that the goals of case selection

Case study analysis focuses on a small number of cases that are expected to provide insight into a causal relationship across a larger population of cases. This presents the researcher with a formidable problem of case selection. Which cases should be chosen?

In large-sample research, case selection is usually handled by some version of randomization. If a sample consists of a large enough number of independent random draws, the selected cases are likely to be fairly representative of the overall population on any given variable. Furthermore, if cases in the population are distributed homogeneously across the ranges of the key variables, then it is probable that some cases will be included from each important segment of those ranges, thus providing sufficient leverage for causal analysis. (For situations in which cases with theoretically relevant values of the variables are rare, a stratified sample that oversamples some subset of the population may be employed.)

A demonstration of the fact that random sampling is likely to produce a representative sample is shown in Figure 5.1, a histogram of the mean values of 500 random samples, each consisting of 1,000 cases. For each case, one variable has been measured: a continuous variable that falls somewhere between zero and one. In the population, the mean value of this variable is 0.5. How representative are the random samples? One good way of judging this is to compare the means of each of the 500 random samples to the population mean. As can be seen in the figure, all of the sample means are very close to the population mean. So random sampling was a success, and each of the 500 samples turns out to be fairly representative of the population.

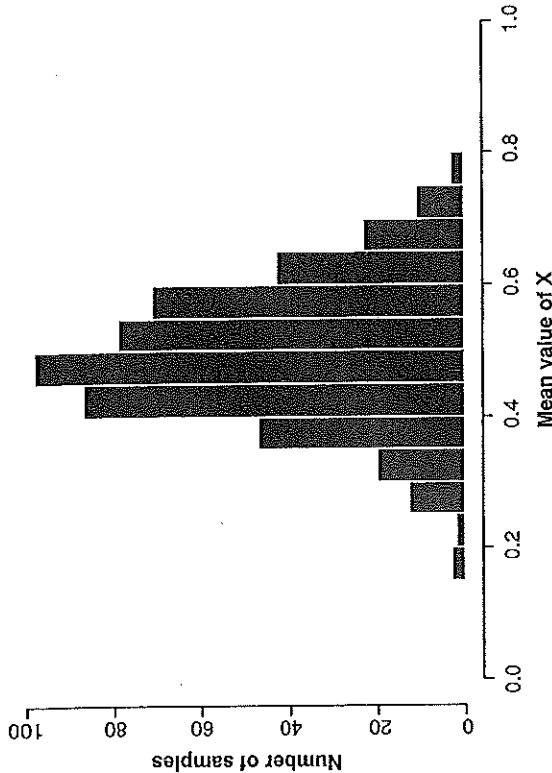


FIGURE 5.2. Sample means of small-sample draws. A histogram showing the mean values of one variable in 500 samples of 5 cases each. Population mean = 0.5.

remain the same regardless of the size of the chosen sample. Large-N cross-case analysis and case study analysis both aim to identify cases that reproduce the relevant causal features of a larger universe (representativeness) and provide variation along the dimensions of theoretical interest (causal leverage). In case study research, however, these goals must be met through purposive (nonrandom) selection procedures. These may be enumerated according to nine techniques, from which we derive nine case study types: *typical, diverse, extreme, deviant, influential, crucial, pathway, most-similar, and most-different*.

Table 5.1 summarizes each type, including its general definition, a technique for identifying 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.

The main point of this chapter is to show how case-selection procedures rest, at least implicitly, upon an analysis of a larger population of potential cases. 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

TABLE 5.1. Techniques of case-selection

1. **Typical**
 - *Definition:* Cases (one 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 (two or more) illuminate the full range of variation on X_1 , X_2 , or X_3/N .
 - *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), or (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 (one or more) exemplify extreme or unusual values on 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 to a larger sample of cases.
4. **Deviant**
 - *Definition:* Cases (one or more) deviate from some cross-case relationship.
 - *Cross-case technique:* A high-residual case (outlier).
 - *Uses:* Hypothesis generating (to develop new explanations of 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 (one 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 (one or more) are most- or least-likely to exhibit a given outcome.
 - *Cross-case technique:* Qualitative assessment of relative crucialness.

(continued)

TABLE 5.1 (continued)

- *Uses:* Hypothesis testing (confirmatory or disconfirmatory).
 - *Representativeness:* Assessable by reference to prior expectations about the case and the population.
7. Pathway
- *Definition:* Cases (one or more) where X_1 , and not X_2 , is likely to have caused a positive outcome ($Y = 1$).
 - *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.
8. Most-similar
- *Definition:* Cases (two or more) are similar on specified variables other than X_1 and/or Y .
 - *Cross-case technique:* Matching.
 - *Uses:* Hypothesis generating or hypothesis testing.
 - *Representativeness:* May be tested by examining residuals for the chosen cases.
9. Most-different
- *Definition:* Cases (two or more) are different on specified variables other than X_1 and Y .
 - *Cross-case technique:* The inverse of the most-similar method of large-N case selection (see above).
 - *Uses:* Hypothesis generating or hypothesis testing (eliminating deterministic causes).
 - *Representativeness:* May be tested by examining residuals for the chosen cases.

situated within that population. This is the origin of the terminology just listed – typical, diverse, extreme, and so on. It follows that case-selection procedures in case study research may build upon prior cross-case analysis and depend, at the very least, upon certain assumptions about a broader population. This, in turn, reinforces a central perspective of the book: case study analysis does not exist, and is impossible to conceptualize, in isolation from cross-case analysis.

To be sure, the sort of cross-case analysis that might be possible in a given research context rests on how large the population of potential cases is, on how much information one has about these cases, on what sort of general model might be constructed, and with what degree of confidence that model might be applied. In order for most quantitative (statistical) case-selection techniques to be fruitful, several caveats must be satisfied. First, the inference must pertain to more than several cases; otherwise,

statistical analysis is usually 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 considerations of statistical research (e.g., identification, specification, robustness) must be carefully considered and, wherever possible, investigated. I shall not dilate further on these familiar issues except to warn the researcher against the unthinking use of statistical techniques.¹

When these requirements are not met, the researcher must employ a qualitative approach to case selection. Thus, the point of this chapter is not to insist upon quantitative techniques of case selection in case study research. My purpose, rather, is to elucidate general principles that might guide the process of case selection in case study research, whether the technique is quantitative or qualitative. Some of these principles are already widely known and widely practiced. Others are less common, or less well understood. Most of these methods are viable – indeed, are virtually identical – in qualitative and quantitative contexts. Hence, the statistical sections of this chapter usually simply reformulate the logic of qualitative case-selection procedures as they might be applied to large populations where the foregoing caveats apply.

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; I employ these two terms synonymously.² (The antonym, *deviance*, is discussed in a later section.)

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 selected a single city “to be

¹ 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).

² The latter term is often employed in the psychological literature (e.g., Hersen and Barlow 1976: 24).

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 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.³

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). Although the Lynds did not employ a statistical model to evaluate potential cases, it is easy to see how they might have done so, at least along the first five criteria. (The final criteria would be difficult to operationalize in a large sample, since it involves "peculiarities" of any sort.)

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

³ Lynd and Lynd (1929/1956), quoted in Yin (2004: 29–30).

involves, first of all, identifying the relevant variables. It involves, secondly, the selection of a case that has "typical" values relative to the overall causal model; it is well explained.

Note that cases with atypical 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.

Note also that because the typical case embodies a typical value on some set of variables, 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.

Cross-Case Technique

How might one identify a typical case from a large population of potential cases? If the causal relationship involves only a single independent variable and if the relationship is quite strong, it may be possible to identify typical cases simply by eyeballing the evidence. A strong positive association between X_1 and Y means that a case with similar (high, low, or middling) values on X_1 and Y is probably a typical case. However, there are few bivariate causal relationships in social science. Usually, more than one causal factor must be evaluated, even if the additional variables serve only as controls. Moreover, without some overall assessment of the cross-case evidence it may be difficult to say whether the general relationship is positive or negative, strong or weak. Thus, in any large- N sample (i.e., whenever the number of potential cases is great) it is advisable to perform a formal cross-case analysis in order to identify "typical" cases.

Suppose that an arbitrary case in the population, denoted as case i , has a known score on each of several relevant variables. For the sake of economy of language, let the variables involved in the relationship be labeled y_i and $x_{1,i}, \dots, x_{K,i}$, where y_i is the score of case i on one variable and each of the $x_{k,i}$'s is the score of case i on one of the K other variables under consideration. Thus, the relationship involves a total of $K + 1$ variables. K can be any integer greater than or equal to 1.

With these symbols, the established relationships among the variables can be expressed mathematically. The idea is to find a function, $f()$, such that the average score of y for cases with some specific set of scores on $x_1 \dots x_k$ is equal to $f(x_1, \dots, x_k)$. Thus, the function $f()$ should be chosen to capture the key ideas about the relationship of interest. A familiar example may make this discussion clearer.

Often, researchers choose an additive (linear) function to play the role of $f()$. Using traditional statistical notation, in which the average score of y_i across infinite repetitions of case i is denoted by its expectation, $E(y_i)$, a linear function represents a relationship in which:

$$E(y_i) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i} \quad (5.1)$$

Each of the β_k 's in this equation represents an unknown constant. Regression analysis allows researchers to use known information about the y and $x_1 \dots x_k$ variables for a set of cases to estimate these unknown constants. Estimates of β_k will be denoted here as b_k .

Using this terminology, we can now develop a formula for the degree to which a particular case is typical in light of a given relationship. A case is "typical" in the terms of small-N methodology to the extent that its score on the y variable is close to the average score on that variable for a case with the same scores on the $x_1 \dots x_k$ variables, as given by equation 5.1. That is,

$$\begin{aligned} \text{Typicality}(i) &= -\text{abs}[y_i - E(y_i | x_{1,i}, \dots, x_{k,i})] \\ &= -\text{abs}[y_i - b_0 + b_1 x_{1,i} + \dots + b_k x_{k,i}] \end{aligned} \quad (5.2)$$

According to this discussion, the typicality of a case with respect to a particular relationship is simply -1 times the absolute value of that case's error term (its residual) in regression analysis. This measure of typicality ranges, in theory, from negative infinity to zero. When a case falls close to the regression line, its typicality will be just below zero. When a case falls far from the regression line, its typicality will be far below zero. Typical cases have small residuals.

In a large-N sample, there will often be many cases with high (i.e., near-zero) typicality scores. In such situations, researchers may elect not to focus on the cases with the highest estimated typicality, for such estimates may not be accurate enough to distinguish among several almost-identical cases. Instead, researchers may choose to randomly select from the set of cases with very high typicality, or to choose from among these cases according to additional criteria, such as those to be discussed here, or by

reason of practicality (cost, convenience, etc.). However, scholars should try to avoid selecting from among the set of typical cases in a way that is correlated with relevant omitted variables; such selection procedures complicate the task of causal inference.

Consider the (presumably causal) relationship between economic development and level of democracy.⁴ Democracy is understood here as a continuous concept along a twenty-one-point scale, from -10 (most autocratic) to $+10$ (most democratic).⁵ Economic development is measured in standard fashion by per capita GDP.⁶ Figure 5.3 displays this relationship in the form of a bivariate scatterplot. The classical result is strikingly illustrated: wealthy countries are almost exclusively democratic. (For heuristic purposes, certain simplifying assumptions are adopted. I shall assume, for example, that this measure of democracy is continuous and unbounded.⁷ I shall assume, more importantly, that the true relationship between economic development and democracy is log-linear, positive, and causally asymmetric, with economic development treated as exogenous and democracy as endogenous.⁸)

Given this general relationship, how might a set of "typical" cases be selected? Recall that the Y variable is simply the democracy score, and there is only one independent variable: logged per capita GDP. Hence, the simplest relevant model is:

$$E(\text{Polity}_i) = \beta_0 + \beta_1 \text{GDP}_i \quad (5.3)$$

For our purposes, the most important feature of this model is the residuals for each case. Figure 5.4 shows a histogram of these residuals. Obviously, a fairly large number of cases have quite low residuals and therefore might be considered typical. A higher proportion of cases fall far below the regression line than far above it, suggesting that the model may be

⁴ Lipset (1959). Whether economic development has only the effect of maintaining democratic regimes (Przeworski et al. 2000) or also of causing regime transitions (Boix and Stokes 2003) is not relevant to the present discussion, where I assume a simple linear relationship between wealth and democracy.

⁵ This scoring derives from the Polity2 variable in the Polity IV dataset (Marshall and Jaggers 2005).

⁶ Data are drawn from the Penn World Tables dataset (Summers and Heston 1991).

⁷ But see Treier and Jackman (2003).

⁸ But see Gerring et al. (2005) and Przeworski et al. (2000).

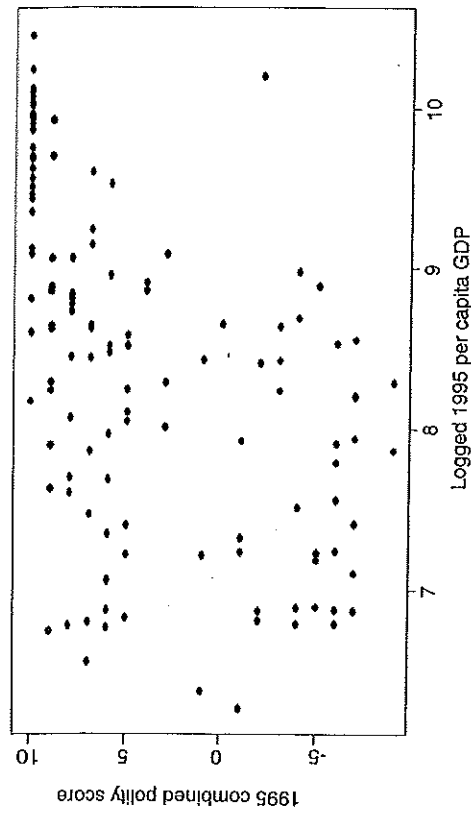


FIGURE 5.3. The presumed relationship between economic development and democracy. A scatterplot showing level of democracy (on the vertical axis) and level of wealth (on the horizontal axis) of all available countries in 1995. $N = 131$.

incomplete. Hopefully, within-case analysis will be able to shed light on the reasons for the asymmetry.⁹

Because of the large number of cases with quite small residuals, the researcher will have a range of options for selecting typical cases. Indeed, in this example, twenty-six cases have a typicality score between 0 and -1 . Any or all of these might reasonably be selected as typical cases with respect to the model described in equation 5.3.

Conclusion

Typicality responds to the first desideratum of case selection, that the chosen case be representative of a population of cases (as defined by the primary inference). Even so, it is important to remind ourselves that a single-minded pursuit of representativeness does not ensure that this desideratum will be achieved. Indeed, the issue of case representativeness is not an issue that can ever be definitively settled in a case study format. 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.

Note that the measure of typicality introduced here, the size of a case's residual, can be misleading if the statistical model is misspecified. And

⁹ In this example, the asymmetry is probably due to the failure of the model to take into account the restricted range of the dependent variable, as discussed earlier.

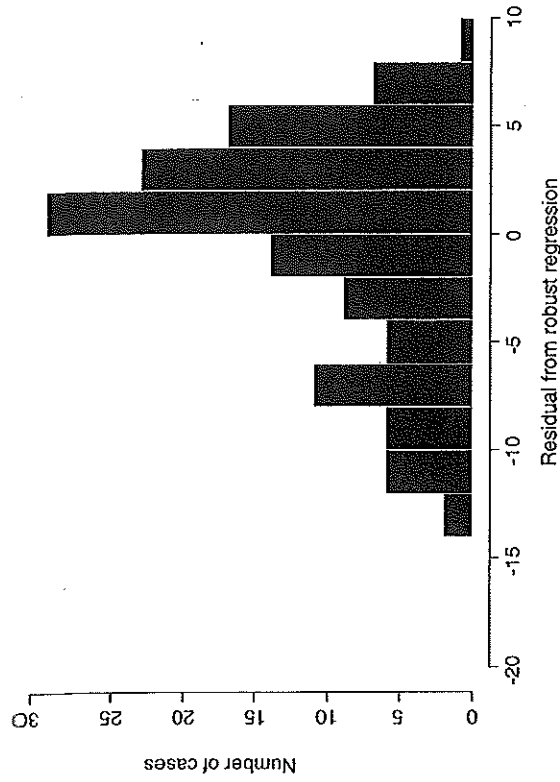


FIGURE 5.4. Potential typical cases. A histogram of the residuals from a robust regression of logged per capita GDP on level of democracy.

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 that in other cases; or unusual causal mechanisms might be at work. Most important, an analysis of residuals does not address problems of sample bias. If the large- N sample is not representative of the relevant population then any analysis based on the former is apt to be flawed. Typicality does not ensure representativeness. For these reasons, 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 general 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.

Diverse Case

A second case-selection strategy has as its primary objective the achievement of maximum variance along relevant dimensions. I refer to this as a

a theory that specifies independent variables, delineates them into nominal, ordinal, or interval categories, and provides not only hypotheses on how these variables operate singly, but contingent generalizations on how and under what conditions they behave in specified conjunctions or configurations to produce effects on specified dependent variables. We call specified conjunctions or configurations of the variables "types." A fully specified typological theory provides hypotheses on all of the mathematically possible types relating to a phenomenon, or on the full 'property space,' to use Lazarfeld's term. Typological theories are rarely fully specified, however, because researchers are usually interested only in the types that are relatively common or that have the greatest implications for theory-building or policy-making.¹¹

George and Smoke, for example, wish to explore different types of divergence failure – by "fait accompli," by "limited probe," and by "controlled pressure." Consequently, they wish to find cases that exemplify each type of causal mechanism.¹²

Diversity may thus refer to a range of variation on X_1 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.

Cross-Case Technique

Since diversity can mean many things, its employment in a large-N setting is necessarily dependent upon how it is understood. If it is understood to pertain only to a single variable (X_1 or Y), then the task is fairly simple, as we have discussed. Univariate traits are usually easy to discover in a large-N setting through descriptive statistics or through visual inspection of the data.

Where diversity refers to particular combinations of variables, the relevant cross-case technique is some version of stratified random sampling (in a probabilistic setting)¹³ or Qualitative Comparative Analysis (in a deterministic setting).¹⁴ If the researcher suspects that a causal relationship is affected not only by combinations of factors but also by their sequencing,

¹¹ George and Bennett (2005: 235). See also Elman (2005) and Lazarsfeld and Barton (1951).

¹² More precisely, George and Smoke (1974: 534, 522–36, Chapter 18; see also discussion in Collier and Mahoney 1996: 78) set out to investigate causal pathways and discovered, in the course of their investigation of many cases, these three causal types. But for our purposes what is important is that the final sample include at least one representative of each "type."

¹³ See Cochran (1977).

¹⁴ Ragin (2000).

diverse-case method. For obvious reasons, this method requires the selection of a set of cases – at minimum, two – that are intended to represent the full range of values characterizing X_1 , Y , or some particular X_1/Y relationship.¹⁰

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, the choices are not so obvious. However, the researcher is well advised to choose both extreme values (high and low), and perhaps the mean or median as well. One may also look for break-points in the distribution that seem to correspond to categorical differences among cases. Or one may follow a theoretical hunch about which threshold values count – that is, which ones 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 dichotomous 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. Again, things become 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 – for example, 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. George and Bennett define a typological theory as

¹⁰ This method has not been given much attention by 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 later. Patton (2002: 234) employs the concept of "maximum variation (heterogeneity) sampling."

then the technique of analysis must incorporate temporal elements.¹⁵ Thus, the method of identifying causal types rests upon whatever method of identifying causal relationships is presumed to exist.

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. 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 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."¹⁶ 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).

Conclusion

Encompassing a full range of variation is likely to enhance the representativeness of the sample of cases chosen by the researcher. 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 often has stronger claims to representativeness than any other small-N sample (including the 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 identify this method of case selection, and to explain its logic to readers.

Extreme Case

The *extreme-case* method selects a case because of its extreme value on an independent or dependent variable of interest.¹⁷ Thus, studies of domestic violence may choose to focus on extreme instances of abuse.¹⁸ Studies of altruism may focus on those rare individuals who risk their lives to help others (e.g., Holocaust resisters).¹⁹ 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.²⁰ Studies of industrial policy often focus on the most successful countries (e.g., the NICs),²¹ and so forth.²²

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, that is, their ideal types. German fascism defines the concept of fascism in part because it offers 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 from 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

¹⁷ It does not make sense to apply the extreme-case method in a confirmatory/disconfirmatory analysis. If a particular causal relationship is at issue, then both X_1 and Y must be taken into account when choosing cases, as described in the various scenarios that follow. At present, therefore, we shall assume that the researcher has a general question in mind, but not a specific hypothesis.

¹⁸ Browne (1987).

¹⁹ Monroe (1996).

²⁰ Reilly (2000/2001).

²¹ Deyo (1987).

²² For further examples, see Collier and Mahoney (1996); Geddes (1990); and Tendler (1997).

¹⁵ Abbott (2001); Abbott and Forrest (1986); Abbott and Tsay (2000).

¹⁶ Swank (2002: 11). See also Esping-Andersen (1990).

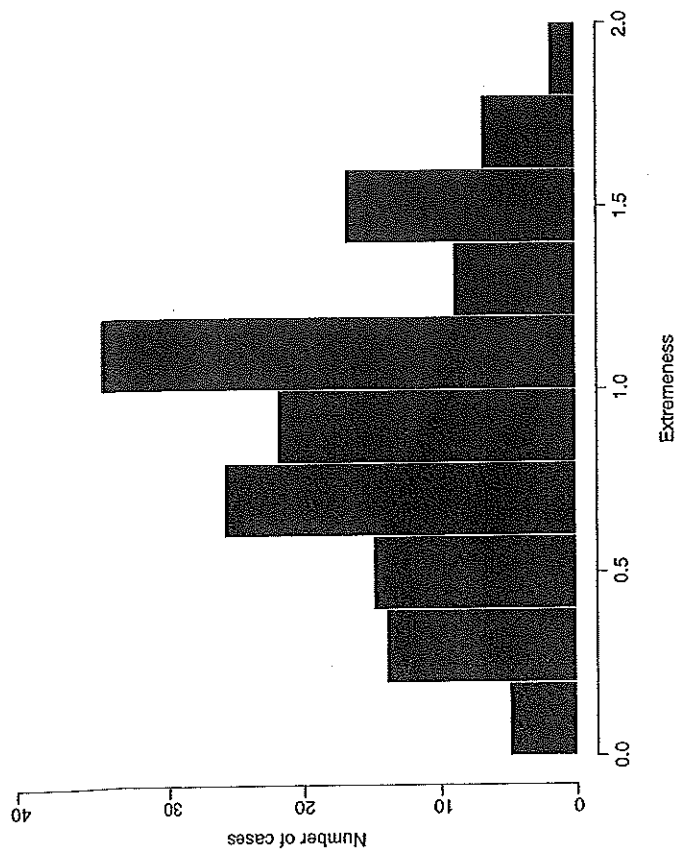


FIGURE 5.5. Potential extreme cases. A histogram of the "extremeness" of all countries on the dimension of democracy, as measured by standard deviations from the mean (absolute value).

to set an arbitrary threshold. Under assumptions of normality, cases with an extremeness score smaller than two would generally not be considered extreme. If the researcher wishes to be more conservative in classifying cases as extreme, a higher threshold may be employed. In general, the choice of threshold is left to the researcher, to be made in a way that is appropriate to the research problem at hand.

The mean of our democracy variable is 2.76, suggesting that the countries in the 1995 dataset tend to be somewhat more democratic than authoritarian (zero is defined as the break-point between democracy and autocracy). The standard deviation is 6.92, implying that there is a fair amount of scatter around the mean.

Figure 5.5 shows a histogram of the extremeness scores for all countries on level of democracy. As can easily be seen, no cases have extremeness scores greater than two. The two countries with the highest scores are Qatar and Saudi Arabia. These countries, which both have a democracy score of -10 for 1995, are probably the two best candidates for extreme case analysis.

II. Doing Case Studies

distribution. For a continuous variable, the distance from the mean may be in either direction (positive or negative). For a dichotomous variable (present/absent), I understand extreme to mean *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. All things being equal, one is concerned not only 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.²³ 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. 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 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 concerned to explain.²⁴ The reason is quite simple: there are fewer revolutionary cases than nonrevolutionary cases; thus, the variation that one wishes to explore as a clue to causal relationships is encapsulated in these cases, viewed against a backdrop of nonrevolutionary cases.

Cross-Case Technique

As stated, extreme cases lie far from the mean of a variable. Extremity (E_i), for the i th case, can be defined in terms of the sample mean (\bar{X}) and the standard deviation (s) for that variable:

$$E_i = \left| \frac{X_i - \bar{X}}{s} \right| \quad (5.4)$$

This definition of extremity is the absolute value of the standardized ("Z") score for the i th case. Cases with a large E_i qualify as extreme. Sometimes, the only criterion is a relative one. The researcher wishes to find the most extreme case(s) available. At other times, it may be helpful

²³ Traditionally, methodologists have conceptualized cases as having "positive" or "negative" values (e.g., Emigh 1997; Mahoney and Goertz 2004; Ragin 2000: 60; Ragin 2004: 126).

²⁴ Skocpol (1979).

Conclusion

The extreme-case method appears to violate the social science folk wisdom warning us not to "select on the dependent variable."²⁵ 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.²⁶ 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 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 – 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 (see the earlier discussion).

By way of conclusion, let us return to the problem of representativeness. In the context of causal analysis, representativeness refers to a case that exemplifies values on X_1 and Y that conform to a general pattern. In a cross-case model, the representativeness of an individual case is gauged by the size of its residual. The representative case is therefore a typical case (as already discussed), not a deviant case (as will be discussed). 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 a causal proposition. Once such a causal proposition has been specified, we 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). 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

²⁵ Geddes (1990); King, Keohane, and Verba (1994). See also discussions in Brady and Collier (2004); Collier and Mahoney (1996); and Rogowski (1995).

²⁶ The exception would be a circumstance in which the researcher intends to disprove a deterministic argument (Dion 1998).

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_1 , 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 has to Y , then she ought to pursue one of the other methods explored elsewhere 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.

Deviant Case

The *deviant-case* method selects the case(s) that, by reference to some general understanding of a topic (either a specific theory or common sense), demonstrates a surprising value. Barbara Geddes 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."²⁷ Similarly, medical researchers are keen to investigate those rare individuals who have not succumbed, despite repeated exposure, to the AIDS virus.²⁸ 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?

Case studies in psychology and sociology are often 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

²⁷ Geddes (2003: 131). For other examples of case work from the annals of medicine, see "Clinical Reports" in *The Lancet*; "Case Studies" in *The 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).

²⁸ Buchbinder and Vittinghoff (1999); Haynes, Pantaleo, and Fauci (1996).

of a formal, mathematical representation of the hypothesis at hand, a deviant case lies as far as possible from that prediction. Referring back to the model developed in equation 5.3, we can define the extent to which a case deviates from the predicted relationship as follows:

$$\begin{aligned} \text{Deviance}(i) &= \text{abs}[y_i - E(y_i | x_{1,i}, \dots, x_{k,i})] \\ &= \text{abs}[y_i - b_0 + b_1x_{1,i} + \dots + b_kx_{k,i}] \end{aligned} \quad (5.5)$$

Deviance ranges from 0, for cases exactly on the regression line, to a theoretical limit of infinity. Researchers will usually be interested in selecting from the cases with the highest overall estimated deviance.

In our running example, a two-variable model with economic development (X_1) and democracy (Y), the most deviant cases fall below the regression line. This can be seen in Figure 5.4. In fact, all eight cases with a deviance score of more than ten have negative residuals; their scores on the outcome are lower than they "should" be, given their level of development. These eight cases are Croatia, Cuba, Indonesia, Iran, Morocco, Singapore, Syria, and Uzbekistan. Our general model of democracy does not explain these cases very well. Quite possibly, we could develop a better model if we understood what — aside from GDP per capita — might be driving the choice of regime type in these polities. This is the usual purpose for which deviant-case analysis is employed.

Conclusion

As I have noted, the deviant-case method is an exploratory form of analysis. As soon as a researcher's exploration of a particular case has identified a factor to explain that case, it is no longer (by definition) deviant. (The exception would be a circumstance in which a case's outcome is deemed to be accidental or idiosyncratic, and therefore inexplicable by any general model.) If the new explanation can be accurately measured as a single variable (or set of variables) across a larger sample of cases, then a new cross-case model is in order. In this fashion, a case study initially framed as a deviant case is likely to be transformed into some other sort of analysis.

This feature of the deviant-case study also helps to resolve doubts about its representativeness. Evidently, the representativeness of a deviant case is problematic, since the case in question is, by construction, atypical. However, this problem can be mitigated if the researcher generalizes whatever proposition is provided by the case study to other cases. In a large-N model, this is accomplished by the creation of a variable to represent the new hypothesis that the case study has identified. This may require some

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.

In all fields, the deviant case is closely linked to the investigation of theoretical anomalies. Indeed, to say "deviant" is to imply "anomalous."²⁹ Note that while extreme cases are judged relative to the mean of a single distribution (the distribution of values along a single dimension), deviant cases are judged relative to some general model of causal relations. The deviant-case method selects cases that, by reference to some general cross-case relationship, demonstrate a surprising value. They are "deviant" in that they are poorly explained by the multivariate model. The important point is that deviancy can only be assessed relative to the general (quantitative or qualitative) model employed.

This means that the relative deviancy 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 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?*

The purpose of a deviant-case analysis is usually 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 a crucial case, as will be discussed later.) Thus, the deviant-case method is only slightly more determinate than the extreme-case method. It, too, is an exploratory form of research. The researcher hopes that causal processes within the deviant case will illustrate some causal factor that is applicable to other (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.

Cross-Case Technique

In statistical terms, deviant-case selection is the opposite of typical-case selection. Where a typical case lies as close as possible to the prediction

²⁹ For a discussion of the important role of anomalies in the development of scientific theorizing, see Elman (2003) and Lakatos (1978). For examples of deviant-case research designs in the social sciences, see Amenta (1991); Coppedge (2004); Eckstein (1975); Emigh (1997); and Kendall and Wolf (1949/1955).

original coding of cases (in addition to the case under intensive study). However, so long as the underlying information for this coding is available, it should be possible to test the new hypothesis in a cross-case model. If the new variable is successful in explaining the studied case, it should no longer be deviant; or, at the very least, it will be less deviant. In statistical terms, its residual will have shrunk. It is now typical, or at least more typical, and this relieves concerns about possible unrepresentativeness.

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 and empirical 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, as well as the appropriate techniques for case identification, is different from that of the deviant-case study, I offer a new term for this method. The *influential case* is a case that appears at first glance to invalidate a theory, or at least to cast doubt upon a theory. Possibly, upon closer inspection, it does not. Indeed, it may end up confirming that theory – perhaps in some slightly altered form. In this guise, the influential case is the "case that proves the rule."

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 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.³⁰

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.³¹

The reader will have noted that influential-case analysis is similar to deviant-case analysis. Both focus on outliers, unusual cases (relative to the theory at hand). However, as we shall see, they focus on different kinds of unusual cases. Moreover, the animating goals of these two research designs are quite different. The influential-case analysis 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 between these two case-study types 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.

In their study of Roberto Michels's "iron law of oligarchy," Lipset, Trow, and Coleman choose to focus on an organization – the International

³⁰ Munck (2004: 118). See also Ertman (1997).

³¹ Ertman (1997: 316).

Typographical Union – that appears to violate the central presupposition.³² 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.³³ Thus, 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. Thus, Michels’s law is shown to be valid for most organizations, but not all. It is valid 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 respect, the use of an influential case to confirm an existing theory is quite similar to the use of a deviant case to unearth a new theory.

Cross-Case Technique

Influential cases in regression are those cases that, if counterfactually assigned a different value on the dependent variable, would most substantially change the resulting estimates. Two quantitative measures of influence are commonly applied in regression diagnostics.³⁴ The first, often referred to as the “leverage” of a case, derives from what is called the *hat matrix*.³⁵ Suppose that the scores on the independent variables for all of the cases in a regression are represented by the matrix X , which has N rows (representing each of the N cases) and $K + 1$ columns (representing the K independent variables and allowing for a constant). Further, allow Y to represent the scores on the dependent variable for all of the cases. Therefore, Y will have N rows and only one column.

Using these symbols, the formula for the hat matrix, H , is as follows:

$$H = X(X^T X)^{-1} X^T \quad (5.6)$$

In this equation, the symbol “ T ” represents a matrix transpose operation, and the symbol “ -1 ” represents a matrix inverse operation.³⁶ A measure

³² Lipset, Trow, and Coleman (1956).

³³ Lipset (1959: 70).

³⁴ Belsey, Kuh, and Welsch (2004).

³⁵ This somewhat curious name derives from the fact that, if the hat matrix is multiplied by the vector containing values of the dependent variable, the result is the vector of fitted values for each case. Typically, the vector of fitted values for the dependent variable is distinguished from the actual vector of values on the dependent variable by the use of the “ $\hat{}$ ” or “hat” symbol. Hence, the hat matrix, which produces the fitted values, can be said to put the hat on the dependent variable.

³⁶ See Greene (2002) for a brief review.

of the leverage of each case can be derived from the diagonal of the hat matrix. Specifically, the leverage of case i is given by the number in the (i,i) position in the hat matrix, or $H_{i,i}$.³⁷

For any X matrix, the diagonal entries in the hat matrix will automatically add up to $K + 1$. Hence, interpretations of the leverage scores for different cases will necessarily depend on the overall number of cases. Clearly, any case with a score near one is a case with a great deal of leverage. In most regression situations, however, no case has a score that high. A standard rule of thumb is to pay close attention to cases with a leverage score higher than $2(K + 1)/N$. Cases with a leverage score above this value are good candidates for influential-case selection.

An interesting feature of the hat matrix is that it does not depend on the values of the dependent variable. Indeed, the Y vector does not appear in equation 5.6. This means that the measure of leverage derived from the hat matrix is, in effect, a measure of *potential* influence. It tells us how much difference the case would make in the final estimate if it were to have an unusual score on the dependent variable, but it does not tell us how much difference each case actually made in the final estimate.

Analysts involved in selecting influential cases will sometimes be interested in measures of potential influence, because such measures are relevant in selecting cases when there may be some a priori uncertainty about scores on the dependent variable. Much of the information in such case studies comes from a careful, in-depth measurement of the dependent variable – which may sometimes be unknown, or only approximately known, before the case study begins. The measure of leverage derived from the hat matrix is appropriate for such situations because it does not require actual scores for the dependent variable.

A second commonly discussed measure of influence in statistics is *Cook’s distance*. This statistic is a measure of the extent to which the estimates of the β_i parameters would change if a given case were omitted from the analysis. Because regression analysis typically includes more than one β_i parameter, a measure of influence requires some method of combining the differences in each parameter to produce an overall measure of a case’s influence. The Cook’s distance statistic resolves this dilemma by

³⁷ The discussion here involves the use of the hat matrix in linear regression. Analysts may also be interested in situations that do not resemble linear regression problems, e.g., where the dependent variable is dichotomous or categorical. Sometimes, these situations can be accommodated within the framework of generalized linear models, which includes its own generalization of the hat matrix (McCullagh and Nelder 1989).

is denoted by ε_i , then the Studentized residual, r_i^2 , can be computed as follows. (All symbols in this expression are as previously defined.)

$$r_i = \frac{\varepsilon_i}{\sqrt{MSE(1 - H_{i,i})}} \quad (5.9)$$

As can be seen from an inspection of equations 5.8 and 5.9, Cook's distance for a case depends primarily on two quantities: the size of the regression residual for that case and the leverage for that case. The most influential cases are those with substantial leverage that lie significantly off the regression line.

Cook's distance for a given case provides a summary of the overall difference that the decision to include that case makes for the parameter estimates. Cases with a large Cook's distance contribute quite a lot to the inferences drawn from the 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 reasonable to select influential cases for in-depth study.

To summarize, three statistical concepts have been introduced in this section. The hat matrix provides a measure of leverage, or potential influence. 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. Cook's distance goes further, considering scores on both the independent and the dependent variables in order to tell us how much the overall regression estimates would be affected if each case were to be dropped from the analysis. This produces a measure of how much actual influence each case has on the overall regression.

Either the hat matrix or Cook's distance may serve as an acceptable measure of influence for selecting case studies, although the differences just discussed must be kept in mind. In the following examples, Cook's distance will be used as the primary measure of influence because our interest is in whether any particular cases might be influencing the coefficient estimates in our democracy-and-development regression. A third concept, the Studentized residual, was introduced as a necessary element in computing Cook's distance. (The hat matrix is, of course, also a necessary ingredient in Cook's distance.)

Figure 5.6 shows the Cook's distance scores for each of the countries in the 1995 per capita GDP and democracy dataset. Most countries have

II. Doing Case Studies

taking a weighted sum of the squared parameter differences associated with deleting a specific case. Specifically, the formula for Cook's distance is:

$$\frac{(b_{-i} - b)^T X^T X (b_{-i} - b)}{(K + 1)MSE} \quad (5.7)$$

In this formula, b represents all of the parameter estimates from the regression using the whole set of cases, and b_{-i} represents the parameter estimates from the regression that excludes the i th case. X , as above, represents the matrix of independent variables. K is the total number of independent variables (not including the constant, which is allowed for in the formula by the use of $K + 1$). Finally, MSE stands for the mean squared error, which is a measure of the amount of variation in the dependent variable not linearly associated with the independent variables.³⁸

This somewhat intimidating mathematical notation gives precise expression to the intuitive idea, discussed earlier, of measuring influence as a weighted sum of the differences that result in each parameter estimate when a single case is deleted from the sample. One disadvantage of this formula is that it requires a number of extra regressions to be run in order to compute measures of influence for each case. The overall regression must of course be computed, and then an additional regression, with one case deleted, is required for each case.

Fortunately, matrix-algebraic manipulation demonstrates that the expression for Cook's distance given in equation 5.7 is equivalent to the following, computationally much easier expression:

$$\frac{r_i^2 H_{i,i}}{(K + 1)(1 - H_{i,i})} \quad (5.8)$$

In this expression, $H_{i,i}$ refers to the measure of leverage for the i th case, taken from the diagonal of the hat matrix, as already discussed. K once again represents the number of independent variables. Finally, r_i^2 is a special, modified version of the i th case's regression residual, known as the Studentized residual, which needs to be separately computed.

The Studentized residual is designed so that the residuals for all cases will have the same variance. If the standard regression residual for case i

³⁸ Specifically, the MSE is found by summing the squared residuals from the full regression and then dividing by $N - K - 1$, where N is the number of cases and K is the number of independent variables.

cases are likely to be and, hence, the less likely a researcher is to use hat matrix and Cook's distance statistics for purposes of case selection. In these instances, it may not matter very much what values individual cases display. (It may of course matter for the purpose of investigating causal mechanisms. However, for this purpose one would not employ influential statistics to choose cases.)

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 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.”³⁹ 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.⁴⁰ Yet the idea of any single case playing a crucial (or “critical”) role is not widely accepted among most methodologists.⁴¹ (Even its progenitor seems to have had doubts.)

Unfortunately, discussion of this method has focused misleadingly on what are presumed to be largely inductive issues. Are there good crucial

³⁹ Eckstein (1975: 118).

⁴⁰ For examples of the crucial-case method, see Bennett, Leppold, and Unger (1994); Desch (2002); Goodin and Smitsman (2000); Kemp (1986); and Reilly and Phillpot (2003). For general discussion, see George and Bennett (2005); Levy (2002a); and Stinchcombe (1968: 24–8).

⁴¹ See, e.g., Sekhon (2004).

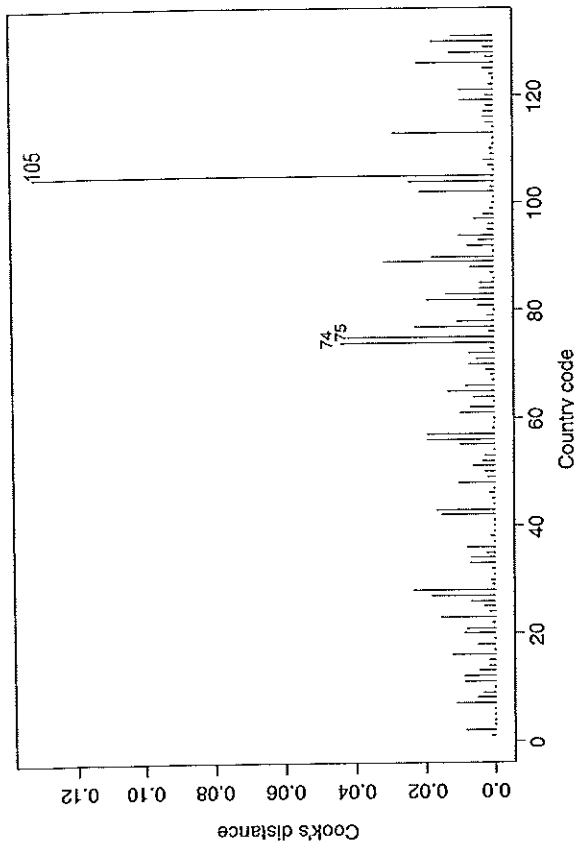


FIGURE 5.6. Potential influential cases. The Cook's distance scores for an OLS regression of democracy on logged per capita GDP. The three numbered cases have high Cook's distance scores.

quite low scores. The three most serious exceptions to this generalization are the numbered lines in the figure: Jamaica (74), Japan (75), and Nepal (105). Of these three, Nepal is clearly the most influential by a wide margin. Hence, any study of influential cases would want to start with an in-depth consideration of Nepal.

Conclusion

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 – for example, 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 dataset 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

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.⁴⁴

While these Popperian/Bayesian insights⁴⁵ 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 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, the riskiness of the theory may compensate for a paucity of empirical evidence.

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 that are lawlike in their precision, 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.⁴⁶ 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. Stephen Van Evera describes the impact of this prediction on the validation of Einstein's theory.

⁴⁴ Ibid.

⁴⁵ A third position, which purports to be neither Popperian nor Bayesian, has been articulated by Mayo (1996: Chapter 6). From this perspective, the same idea is articulated as a matter of "severe tests."

⁴⁶ See, e.g., Hempel (1942).

cases out there in the empirical world? Have social scientists done a good job in identifying them? Yet the practicability of this method rests on issues that are largely deductive in nature, as we shall see.

The Confirmatory (Least-Likely) Crucial Case

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 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 event which was incompatible with the theory – an 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.⁴²

A risky prediction is therefore one that is highly precise and determinate, and thus unlikely to be explainable by 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 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."⁴³ 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 offer the following:

⁴² Popper (1963: 36). See also Popper (1934/1968).

⁴³ Howson and Urbach (1989: 86).

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.⁴⁷

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 knock-out 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 (though, as we shall discuss, the theory may be *disconfirmed* with a single case).⁴⁸

However, if one adopts a softer version of the crucial-case method – the least-likely (most difficult) case – then possibilities abound. Lily Tsai's investigation of governance at the village level in China employs several in-depth case studies of villages that 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 that, along other dimensions, would be judged least-likely to develop good governance – that is, 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 would one explain this otherwise anomalous result? This is the strength of the least-likely case, where all other plausible explanations for an outcome have been mitigated.⁴⁹

Jack Levy refers to this, evocatively, as a “Sinatra inference”: if it can make it here, it can make it anywhere.⁵⁰ 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 (without extensive cross-case analysis). Indeed, I suspect that, *implicitly*, most case study work that focuses on a single case and is not nested within a 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 may in fact be 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.

⁴⁹ Tsai (2007). 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.

⁵⁰ Levy (2002a: 144). See also Khong (1992: 49); Sagan (1995: 49); and Shafer (1988: 14–6).

Netherlands, India, and Papua New Guinea score on *other* factors that affect democracy and social peace.

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 may 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.

Conclusion

In this section, I have argued that the degree to which crucial cases can provide decisive confirmation or disconfirmation of a theory is in large part a product of the structure of the theory to be tested. It is a deductive matter rather than an inductive matter, strictly speaking. In this respect, a "positivist" orientation toward the work of social science may lead to a greater appreciation of the case study format – not a denigration of that format, as is usually supposed. Those who, with Eckstein, embrace the notion of covering laws are likely to be attracted to the idea of cases that are crucial. By the same token, those who are impressed by the irregularity and complexity of social behavior are unlikely to be persuaded by crucial-case studies, except as a method of *disconfirming* absurdly rigid causal laws.

I have shown, relatedly, that it is almost always easier to disconfirm a theory than to confirm it with a single case. Thus, a theory that is understood to be deterministic may be disconfirmed by a case study, properly chosen. This is the most common employment of the crucial-case method in social science settings.

Note that the crucial-case method of case selection *cannot* be employed in a large-N context. This is because the method of selection would render the 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 will, by itself, confirm or disconfirm the theory in question. The case study is thenceforth irrelevant, at least as a means of confirmation or disconfirmation. It remains highly relevant as a means of

The Disconfirmatory (Most-Likeby) Crucial Case

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.⁵¹ In Chapter Three, we discussed an older theory that stipulated that political stability would occur only in countries that are relatively homogeneous, or where existing heterogeneities are mitigated by cross-cutting cleavages.⁵² Arend Lijphart's 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.⁵³

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 – that is, 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 another case, 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 when *disconfirming* a causal argument, background causal factors are irrelevant (except as they might affect the classification of the case within the population of an inference). It does not matter how the

⁵¹ Goertz and Levy (forthcoming); Goertz and Starr (2003).

⁵² Almond (1956); Bentley (1908/1967); Lipset (1960/1963); Truman (1951).

⁵³ Lijphart (1968). See also discussions 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 (2000/2001); and the discussions in Dion (1998) and Rogowski (1995).

Consider the following examples culled by Bear Braumoeller and drawn from diverse fields of political science.⁵⁵ The decision to seek an alliance is motivated by the search for either autonomy or security.⁵⁶ Conflict is prevented by either deterrence or defense.⁵⁷ Civilian intervention in military affairs is caused by either political isolation or geographical encirclement.⁵⁸ War is the product of miscalculation or loss of control.⁵⁹ Nonvoting is caused by ignorance, indifference, dissatisfaction, or inactivity.⁶⁰ Voting decisions are influenced either by high levels of information or by the use of candidate gender as a proxy for social information.⁶¹ Democratization comes about through leadership-initiated reform, a controlled opening to opposition, or the collapse of an authoritarian regime.⁶² These, and many other, social science arguments take the form of causal substitutability – multiple paths to a given outcome.

For heuristic purposes, it will be helpful to pursue one of these examples in greater detail. For consistency, I focus on the last of the exemplars – democratization. The literature, according to Braumoeller, identifies three main avenues of democratization (there may be more, but for present purposes let us assume that the universe is limited to three). 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. This covariational assumption about the relationship

⁵⁵ *Ibid.* My chosen examples are limited to those that might plausibly be modeled with dichotomous variables. For further discussion and additional examples, see Most and Starr (1984) and Clouff-Revilla and Starr (1995).

⁵⁶ Morrow (1991: 905).

⁵⁷ Schelling (1966: 78).

⁵⁸ Posen (1984: 79).

⁵⁹ Levy (1983: 86).

⁶⁰ Ragsdale and Rusk (1993: 723–4).

⁶¹ McDermott (1997).

⁶² Colomer (1991).

exploring causal mechanisms, of course. However, because this objective is quite different from that which is usually associated with the term, I enlist a new term for this technique.

Pathway Case

One of the most important functions of case study research is the elucidation of causal mechanisms. This is well established (see Chapter Three). But what 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 one factor can be isolated from other potentially confounding factors. I shall call this a *pathway case* to indicate its uniquely penetrating insight into causal mechanisms.

To clarify, the pathway case exists only in circumstances where cross-case covariational patterns are well studied but 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 stand-alone pathway case. Thus, the following discussion focuses on how to select one (or a few) cases from a cross-case sample.

Cross-Case Technique with Binary Variables

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 cause a particular outcome, Y , understood as a unidirectional or asymmetric causal relationship. 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 deemed to be substitutable for one another, each factor is conceptualized (individually) as sufficient.⁵⁴ Situations of causal equifinality presume causal sufficiency on the part of each factor or set of conjoint factors. The QCA technique, for example, presumes causal sufficiency for each of the designated causal paths.

⁵⁴ Braumoeller (2003).

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 (zero or one). "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 's positive value ($Y = 1$) 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 is only one such case in Table 5.2 - H. In all other cases, the mechanism running from X_1 to Y would be difficult to discern, because the outcome to be explained does not occur ($Y = 0$), because X_1 and Y are not correlated in the usual way (violating 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 . Consequently, an in-depth examination of cases A-G is not likely to be very revealing.

Keep in mind that because we already know from our cross-case examination what the general causal relationships are, we know (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 for verifying or falsifying general propositions (which are already apparent from the cross-case evidence).⁶³

Now let us complicate matters a bit by imagining a scenario in which at least some of these substitutable causes are conjoint (a.k.a. conjunctural). That is, several combinations of factors - $X_a + X_b$ or $X_c + X_d$ - are sufficient to produce the outcome, Y . This is known in philosophical circles as an INUS condition,⁶⁴ and it is the pattern of causation assumed in most

⁶³ Of course, we should leave open the possibility that an investigation of causal mechanisms might 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).

⁶⁴ An INUS condition refers to an Insufficient but Necessary part of a condition which is itself Unnecessary but Sufficient for a particular result. Thus, when one identifies a short circuit as the "cause" of a fire, one is saying, in effect, that the fire was caused by a short

TABLE 5.2. Pathway case with dichotomous causal factors

	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	0	1	0
H	1	0	1

X_1 = the variable of theoretical interest. X_2 = a vector of controls (a score of zero indicates that all control variables have a score of zero, while a score of one indicates that all control variables have a score of one). Y = the outcome of interest. A-H = case types (the N for each case type is indeterminate). H = pathway case. Sample size = indeterminate.

Assumptions: (a) all variables can be coded dichotomously; (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 .

between X_1 and Y is presumably sustained by the cross-case evidence (if it is not, there is no justification for a pathway case study).

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 zero or one) 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 focus only on one causal factor at a time.) In this scenario, and regardless of how 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 5.2. 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 5.2) 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

QCA (Qualitative Comparative Analysis) models.⁶⁵ Here, everything that has been said so far must be adjusted so that X_1 refers to a set of causes (e.g., $X_a + X_b$) and X_2 refers to a vector of sets (e.g., $X_c + X_d$, $X_e + X_f$, $X_g + X_h$, ...). The scoring of all these variables makes matters more difficult than in the previous set of examples. However, the logical task is identical, and can be accomplished in a similar fashion, that is, in small-N datasets with truth tables and in large-N datasets with cross-tabs. Case H now refers to a conjunction of causes, but it is still the only possible pathway case.

Cross-Case Technique with Continuous Variables

Finally, we must tackle the most complicated scenario – when 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, that is, that all rival factors can be eliminated. Even so, the search for a pathway case may 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 identify the causal mechanisms that lie between X_1 and Y.

In a large-N sample, these two desiderata may be judged by a careful attention to the residuals attached to each case. Recall that the question of deviance, which we have discussed in previous sections, is a matter of degree. Cases are more or less typical/deviant relative to a general model, as judged by the size of their residuals. It is easy enough to exclude cases with very high residuals (e.g., standardized residual $> |2|$). For cases that lie closer to their predicted value, small differences in the size of residuals may not matter so much. But, *ceteris paribus*, one would prefer a case that lies closer to the regression line.

circuit in conjunction with some other background factors (e.g., oxygen) that were also necessary to that outcome. But one is not implying that a short circuit was necessary to that fire, which might have been (under different circumstances) caused by other factors. See Mackie (1965/1993).

⁶⁵ Ragin (2000).

Achieving the second desideratum requires a bit of manipulation. In order to determine which (non-outlier) cases are most strongly affected by X_1 , given all the other parameters in the model, one must compare the size of the residuals (their absolute value) for each case in a reduced-form model, $Y = \text{Constant} + X_2 + \text{Res}_{\text{reduced}}$, to 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, that shows the greatest difference between the residuals for the reduced-form model and the full model ($\Delta \text{Residual}$). Thus,

$$\text{Pathway} = |\text{Res}_{\text{reduced}} - \text{Res}_{\text{full}}|, \text{ if } |\text{Res}_{\text{reduced}}| > |\text{Res}_{\text{full}}| \quad (5.10)$$

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 toward the regression line, that is, it helps to “explain” the 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).⁶⁶ The cross-country evidence is robust. Yet, as is often the case, causal mechanisms remain rather obscure. Consider the following list of possible causal pathways, summarized by Michael Ross:

A ‘rentier effect’ ... suggests that resources rich governments use low tax rates and patronage to relieve pressures for greater accountability; a ‘repression effect’ ... argues that resource wealth retards democratization by enabling governments to boost their funding for internal security; and a ‘modernization effect’ ... holds that growth based on the export of oil and minerals fails to bring about the social and cultural changes that tend to produce democratic government.⁶⁷

Are all three causal mechanisms at work? Although Ross attempts to test these factors in a large-N cross-country setting, his answers remain rather

⁶⁶ Barro (1999), Humphreys (2005); Ross (2001).

⁶⁷ Ross (2001: 327–8).

TABLE 5.3. Possible pathway cases where variables are scalar and assumptions probabilistic

Country	Res _{reduced}	Res _{full}	ΔResidual
Iran	-.282	-.456	.175
Turkmenistan	-1.220	-1.398	.178
Mauritania	-.076	-.255	.179
Turkey	2.261	2.069	.192
Switzerland	.177	-.028	.205
Venezuela	.148	.355	-.207
Belgium	.518	.310	.208
Morocco	-.540	-.776	.236
Jordan	.382	.142	.240
Djibouti	-.451	-.696	.245
Bahrain	-1.411	-1.673	.262
Luxembourg	.559	.291	.269
Singapore	-1.593	-1.864	.271
Oman	-1.270	-.981	-.289
Gabon	-1.743	-1.418	-.325
Saudi Arabia	-1.681	-1.253	-.428
Norway	.315	1.285	-.971
United Arab Emirates	-1.256	-.081	-1.175
Kuwait	-1.007	.925	-1.932

Res_{reduced} = the standardized residual for a case obtained from the reduced model (without Oil) - equation 5.12.

Res_{full} = the standardized residual for a case obtained from the full model (with Oil) - equation 5.11.

ΔResidual = Res_{reduced} - Res_{full}. Listed in order of absolute value.

these may be summarily removed from consideration by virtue of the fact that $|\text{Res}_{\text{reduced}}| < |\text{Res}_{\text{full}}|$. Thus, we see that the inclusion of Oil increases the residual for Norway; this case is apparently better explained *without* the inclusion of the variable of theoretical interest. Needless to say, this is not a good case to explore if we wish to examine the causal mechanisms that lie between natural resource wealth and democracy. (It might, however, be a good case for model diagnostics, as discussed in the previous section on influential cases.)

Among cases where the residual declines from the reduced to the full model, several are clear-cut favorites as pathway cases. The United Arab Emirates and Kuwait have the highest ΔResidual values and also have fairly modest residuals in the full model (Res_{full}), signifying that these cases are not extreme outliers; indeed, according to the parameters of this model, the United Arab Emirates would be regarded as a typical case. The

II. Doing Case Studies

speculative.⁶⁸ Let us see how this might be handled by a pathway-case approach.

The factor of theoretical interest, oil wealth, may be operationalized as per capita oil production (barrels of oil produced, divided by the total population of a country).⁶⁹ As previously, we measure democracy with a continuous variable coded from -10 (most authoritarian) to +10 (most democratic). Additional factors in the model include GDP per capita (logged), Muslims (as percent of the population), European language (percent speaking a European language), and ethnic fractionalization (1 - likelihood of two randomly chosen individuals belonging to the same ethnic group).⁷⁰ These are regarded as background variables (X₂) that may affect a country's propensity to democratize. The full model, limited to 1995 (as in previous analyses), is as follows:

$$\text{Democracy} = -3.71 \text{ Constant} + 1.258 \text{ GDP} \quad (5.11)$$

$$+ .075 \text{ Muslim} + 1.843 \text{ European}$$

$$+ .2093 \text{ Ethnic fract} + .7662 \text{ Oil}$$

$$R2_{\text{adj}} = .450 \quad (N = 149)$$

The reduced-form model is identical except that the variable of theoretical interest, Oil, is removed.

$$\text{Democracy} = -.831 \text{ Constant} + .909 \text{ GDP} \quad (5.12)$$

$$+ .086 \text{ Muslim} + 2.242 \text{ European}$$

$$+ .3023 \text{ Ethnic fract}$$

$$R2_{\text{adj}} = .428 \quad (N = 149)$$

What does a comparison of the residuals across equations 5.11 and 5.12 reveal? Table 5.3 displays the highest ΔResidual cases. Several of

⁶⁸ Ross tests these various causal mechanisms with cross-country data, employing various proxies for these concepts in the benchmark model and observing the effect of these - presumably intermediary - effects on the main variable of interest (oil resources). This is a good example of how cross-case evidence can be mustered to shed light on causal mechanisms; one is not limited to case study formats, as discussed in Chapter Three. Still, as Ross notes (2001: 356), these tests are by no means definitive. Indeed, the coefficient on the key oil variable remains fairly constant, except in circumstances where the sample is severely constrained.

⁶⁹ Derived from Humphreys (2005).

⁷⁰ GDPpc data are from World Bank (2003). Muslims and European language are coded by the author. Ethnic fractionalization is drawn from Alesina et al. (2003).

analysis suggests, therefore, that 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 such as, 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 countries, an effect that is well modeled by our general theory and by the available cross-case evidence. And this effect should be interpretable in a case-study format – more interpretable, at any rate, than it would be in other cases.

Conclusion

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, this discussion has shown why the selection of a pathway case is 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, that is, no cases of type H (in Table 5.2). 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 test a given hypothesis independently of all others.

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 (as judged by its standardized coefficient or by F-tests comparing the reduced-form model and the full model). 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 must be underlined that the identification of a causal-pathway case does not obviate the utility of exploring other cases. However, this sort of multicase investigation moves beyond the logic of the causal pathway case, underlining a point that we shall return to in the concluding

section of the chapter: case-selection procedures often combine different logics.

Despite the technical nature of this discussion, it should be noted that when researchers refer to a particular case as an “example” of a broader phenomenon, they are often referring to a pathway case. This sort of case illustrates the causal relationship of interest in a particularly vivid manner, and therefore may be regarded as a common trope among case study researchers.

Most-Similar Case

The most-similar method, unlike the previous methods, employs a minimum of two cases.⁷¹ 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 5.4 (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 the putative causes.

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 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 5.4 (B). If she discovers such a case, it is regarded as providing confirmatory evidence for the proposition, as well as fodder for an exploration of causal mechanisms.

The point is that the purpose of a most-similar research design, and hence its basic set-up, may change as a researcher moves from an exploratory to a confirmatory mode of analysis. However, regardless of

⁷¹ 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: Chapter 9); Lijphart (1971, 1975); Meckstroth (1975); Przeworski and Teune (1970); and Skocpol and Somers (1980).

TABLE 5.4. Most-similar analysis with two case types

(A) Hypothesis-generating (Y -centered):

	X_1	X_2	Y
Case A	?	0	1
Case B	?	0	0

(B) Hypothesis-testing (X_1/X_2 -centered):

	X_1	X_2	Y
Case A	?	0	?
Case B	0	0	?

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

where one begins, the results, when published, look like a hypothesis-testing research design. Question marks have been removed: (A) becomes (B) in Table 5.4. Consequently, the notion of a "most-similar" analysis is usually understood as a tool for understanding a specific X_1/Y relationship.

As an example, let us consider Leon Epstein's classic study of party cohesion, which focuses on two similar countries, the United States and Canada. Canada has highly disciplined parties whose members vote together on the floor of the House of 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 a fairly loose party structure with strong regional bases and a weak center. These are the "control" variables (X_2). 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 differentiating cause (X_1).⁷²

Several caveats apply to any most-similar analysis (in addition to the usual set of assumptions applying to all case study analysis). First,

⁷² For further examples of the most-similar method, see Brenner (1976); Hamilton (1977); Lipset (1968); Miguel (2004); Moulder (1977); and Posner (2004).

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. This is relatively unproblematic if the actual scores on this dimension are quite different (on X_1 and Y) or virtually identical (on X_2). Unfortunately, the empirical universe does not always oblige the requirements of Millean-style analysis, and in these instances the logic of most-similar comparison becomes questionable.

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 to stronger parties, this element of nonidentity does not challenge Epstein's conclusions. Indeed, it sets up a most-difficult research scenario, as discussed earlier. At the same time, Epstein's description of Canadian and American parties as "loose" might be questioned. Arguably, American parties, dominated in the latter twentieth century by direct primaries (open to all who declare themselves a member of a party and, in some states, even to those who are members of the opposing party), are considerably more diffuse than Canadian parties. The problem of coding continuous variables in a dichotomous manner is threatening to any most-similar analysis.

In one respect, however, 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. This can be a huge advantage over large- N cross-case methods, where each case must be assigned a specific score on all relevant control variables — often a highly questionable procedure, and one that must impose strong assumptions

about the shape of the underlying causal relationship (usually presumed to be linear).

Cross-Case Technique

The most useful statistical tool for identifying cases for in-depth analysis in a most-similar setting is some variety of "matching" strategy.⁷³ Statistical estimates of causal effects based on matching techniques have been a major topic in quantitative methodology over the last twenty-five years, first in statistics⁷⁴ and subsequently in econometrics⁷⁵ and political science.⁷⁶

Matching techniques are based on an extension of experimental logic. In a randomized experiment, elaborate statistical models are unnecessary for causal inference because, for a large enough selection of cases, the treatment group and the control group have a high probability of being similar in their background characteristics (X_2). Hence, a simple difference-of-means test is often sufficient to analyze the effects of a treatment variable (X_1) across groups.

In observational studies where the hypothesized causal factor (X_1) is dichotomous, the situation is superficially the same. For purposes of discussion, we shall refer to cases with a "high" score on X_1 as members of the treatment group, and to cases with "low" scores as members of the control group. Thus are observational studies translated into the lexicon of experimental analysis.

However, in observational studies it is unusual to find cases that differ on X_1 but not on various background characteristics (X_2) that might affect the outcome of interest. For example, countries that are strongly democratic (or strongly authoritarian) are likely to be similar in more than one respect. This greatly complicates the analysis of X_1 's independent effect on the outcome.

The traditional approach to this problem is to introduce a variable for each potential confounder in a regression model of causal relationships. But this standard-issue technique requires a strong set of assumptions about the behavior of the various factors introduced into the model. Matching techniques have been developed as an explicit alternative to

⁷³ For good introductions, see Ho et al. (2004); Morgan and Harding (2005); Rosenbaum (2004); and Rosenbaum and Silber (2001). For a discussion of matching procedures in Stata, see Abadie et al. (2001).

⁷⁴ Rosenbaum and Rubin (1985); Rosenbaum (2004).

⁷⁵ Hahn (1998).

⁷⁶ Ho et al. (2004); Imai (2005).

the control-variable approach. This alternative begins by identifying a set of variables (other than the dependent variable or the main independent variable) on which the cases are to be matched. Then, for each case in the treatment group, the researcher identifies as many cases as possible from the control group with the exact same scores on the matching variables (the covariates). Finally, the researcher looks at the difference on the dependent variable between the cases in the treatment group and the matching cases in the control group. If the set of matching variables is broad enough to include all confounders, the average difference between the treatment-group and the matching control-group cases should provide a good estimate of the causal effect. Even in a situation in which the set of matching variables includes some, but not all, confounders, matching may produce better causal inferences than regression models because cases that match on a set of explicitly selected variables are also more likely to be similar on unmeasured confounders.⁷⁷

Unfortunately, the relatively simple matching procedure just described, known as *exact matching*, is often impossible. This procedure typically fails for continuous variables such as wealth, age, and distance, since there may be no two cases with the same score on a continuous variable. For example, there is no undemocratic country with the exact same per capita GDP as the United States. Note that the larger the number of covariates, the lower the likelihood of finding exact matches.

In situations where exact matching is infeasible, researchers may instead employ *approximate matching*, where cases from the control group that are close enough to matching cases from the treatment group are accepted as matches. Major weaknesses of this approach include the fact that the definition of "close enough" is inevitably arbitrary, as well as the fact that, for large sets of matching variables, few treatment cases are likely to have even approximate matches.

To deal with situations in which exact matching is impossible, methodologists have developed an alternative procedure known as *propensity-score matching*. This approach suggests a somewhat different definition of similarity than the previous two. Rather than focusing on sharing scores on the matching variables, propensity-score matching focuses on sharing a similar estimated probability of having been in the treatment group,

⁷⁷ However, matching is clearly inferior to a well-designed and well-executed randomized experiment. The benefits of matching extend only so far as equivalence on the variables explicitly included and any unmeasured variables that fortuitously happen to be similar across the cases. By contrast, proper randomization handles all unmeasured variables.

In order to select most-similar cases for the study of the relationship between wealth and democracy, one must arrive at a statistical model of the causes of a country's wealth. Obviously, such a proposition is complex. Since this is an illustrative example, we shall be satisfied with a cartoon model that includes only a few independent variables. A country's wealth will be assumed to be a function of the origin of its legal system (measured by dummy variables for English legal heritage, French legal heritage, socialist legal heritage, German legal heritage, and Scandinavian legal heritage) and its geographic endowments (measured by the distance of each country's capital city from the equator).

The first step in selecting most-similar cases is to run a nonparametric regression with these independent variables and logged per capita GDP (the independent variable of theoretical interest) as the dependent variable. The fitted values from this regression serve as propensity scores, and cases with similar propensity scores are interpreted as matching. The propensity score for our focus case, Costa Rica, is 7.63. Examining the propensity-score data, one sees that Benin has a propensity score of 7.58 – quite similar to Costa Rica's. At the same time, Benin's per capita GDP of \$1,163 is substantially different from Costa Rica's per capita GDP of \$5,486, as are their democracy scores in 1995 (Benin is much less democratic than Costa Rica). Hence, Costa Rica and Benin may be viewed as most-similar cases for testing the relationship between wealth and democracy, as illustrated in Table 5.5. An in-depth analysis of these two cases may shed light on the causal pathways between economic development and democracy. Indeed, these two cases are probably more informative than other two-case comparisons precisely because the case-selection procedure has identified countries whose *other* attributes are roughly equal in their propensity to democracy/authoritarianism. This means that the differences on the variable of theoretical interest (GDP per capita) and the outcome (democracy) can be given a causal interpretation – an interpretation that would probably not be suggested by a qualitative assessment of these two countries (which are quite different in culture, region, and historical experience).

It is important to keep in mind that the quality of the “match” depends entirely on the quality of the statistical model used to generate the propensity scores. A superficial model like the one used here may produce rather superficial matches. Yet, in a large-N context – where dozens, if not thousands of cases vie for inclusion – a formal approach to case selection offers significant advantages. At the very least, one's assumptions are rendered transparent.

II. Doing Case Studies

conditional on the matching variables. In other words, when looking for a match for a specific case in the treatment group, researchers look for cases in the control group that – before the score on the independent variable is known – would have been as likely to be in the treatment group as actually chosen cases. This is accomplished by a two-stage analysis, the first stage of which approaches the key independent variable, X_1 , as a dependent variable and the matching variables as independent variables. (This is similar in spirit to *selection models*, where a two-stage approach to causal inference is adopted.) Once this model has been estimated, the coefficient estimates are disregarded. Instead, the second stage of the analysis employs the fitted values for each case, which tell us the probability of that case being assigned to the treatment group, conditional on its scores on the matching variables. These fitted values are referred to as propensity scores. The final step in the process is to choose matches for each case in the treatment group. This is accomplished by selecting cases from the control group with similar propensity scores.

The end result 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, returning to the qualitative terminology. Rosenbaum and Silber summarize the results of recent medical studies:

Unlike model-based adjustments, where patients 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.⁷⁸

Matching, conclude the authors, “facilitates, rather than inhibits, thick description.”⁷⁹

Indeed, 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 paired cases in the social sciences. Suppose that, in order to study the relationship between wealth and democracy, the researcher wishes to select a case that is as similar as possible to Costa Rica in background variables, while being as different as possible on per capita GDP, the variable of theoretical interest, and the outcome of interest, democracy.

⁷⁸ Rosenbaum and Silber (2001: 223).

⁷⁹ *Ibid.*

TABLE 5.5. Paired cases resulting from matching procedure

Cases	GDP per capita score (X_1)	Propensity score (X_2)	Democracy (Y)
Benin	\$1,163	7.58	6
Costa Rica	\$5,486	7.63	10

Conclusion

The most-similar method is one of the oldest recognized techniques of qualitative analysis, harking back to J. S. Mill's classic study, *System of Logic* (first published in 1834). By contrast, matching statistics are a relatively new technique in the arsenal of the social sciences, and have rarely been employed for the purpose of selecting cases for in-depth analysis. Yet, as suggested in the foregoing discussion, there may be a fruitful interchange between the two approaches. 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.

To be sure, the purpose of a case study is somewhat different in situations where a large-N cross-case analysis has already been conducted. Here, the general causal relationship is usually clear. We know from our cross-case study that GDP per capita is strongly associated with democracy; there is a strong presumption of causality. Of course, the case study analysis may give us reasons to doubt. Perhaps the causal pathways from economic development to regime type are difficult to identify. Perhaps the presumed causal pathways, as identified by previous research or theoretical hunch, are simply not in evidence. Even so, the usual purpose of a case study analysis in this setting is to corroborate an initial cross-case finding.

By contrast, if there is no prior cross-case investigation – at least none of a formal nature – the case study performs a somewhat different role. Here, we will be more interested in the covariational patterns that are discovered between X_1 and Y . Thus, Epstein's study of American and Canadian political parties is notable for its principal finding: that the underlying cause of party cohesion is to be found in the structure of

the executive (parliamentary/presidential). Indeed, Epstein spends relatively little time in this article discussing possible causal mechanisms; his principal focus is on “scoring” the relevant variables, as discussed. By the same token, if Epstein had already conducted a large-N cross-case analysis prior to his case study, and if this cross-case analysis had revealed a strong pattern between executive type and party cohesion, his two-case analysis of the United States and Canada (cases that we presume would have very similar propensity scores) would now serve a rather different purpose. Evidently, the function of the most-similar case study shifts subtly but importantly when the case-selection procedure is, itself, a mode of analysis, offering strong prima facie evidence of a causal relationship.

As with other methods of case selection, the most-similar method is prone to problems of non-representativeness. If this technique is employed in a qualitative fashion (without a systematic cross-case selection strategy), potential biases in the chosen cases 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 a 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.

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.⁸⁰

⁸⁰ The most-different method is sometimes referred to as the “method of agreement,” following its inventor, J. S. Mill (1843/1872). See also DeFeice (1986); Gerring (2001: 212–14); Liphart (1971, 1975); Meckstroth (1975); Przeworski and Teune (1970); and 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); and Yashar (2005: 23). However, most of these studies are described as *combining* most-similar and most-different methods.

organizations, the persistence of friendship networks, and the disappointment with post-communism.⁸³ Simply put, Howard 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."⁸⁴

Results obtained from the analysis of East Germany and Russia are presumed to apply in other post-communist polities (e.g., Lithuania, Poland, Bulgaria, Albania). Indeed, by choosing a heterogeneous sample, Howard solves potential problems of representativeness in his restricted sample. However, this sample is *not* representative across the entire population of the inference, which is intended to cover all countries, not just communist ones. (To argue that communism impedes the development of civil society is to imply that noncommunism stimulates the development of civil society. The chosen sample is truncated [censored] on the dependent variable).

Equally problematic is the lack of variation on key causal factors of interest – communism and its putative causal pathways. For this reason, it is generally 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 have never experienced communist rule. If so, they are not properly regarded as causal.

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, based on the foregoing methodological dilemmas, is that the most-different research design provides only minimal insight 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. In my opinion, these methods do most of the analytic work. East Germany may be regarded

⁸³ *Ibid.*, 122.

⁸⁴ *Ibid.*, 145.

TABLE 5.6. Most-different analysis with two cases

	X_1	X_{2a}	X_{2b}	X_{2c}	X_{2d}	Y
Case A	1	1	0	1	0	1
Case 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.

The simplest form of this two-case comparison is illustrated in Table 5.6. 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 recent work, which explores the enduring impact of communism on civil society.⁸¹ 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, 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 case selection procedure meets the requirements of the most-different research design: variance is found on all (or most) dimensions aside from the key factor of interest (communism) and the outcome (civic engagement).⁸²

What leverage is brought to the analysis by 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 claims, help to explain the laggard status of civil society in post-communist polities: "the mistrust of communist

⁸¹ Howard (2003). In the following discussion I treat the terms "social capital," "civil society," and "civic engagement" interchangeably.

⁸² Howard (2003: 6–9).

as a causal-pathway case (as discussed earlier). 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.⁸⁵ It is plausible to attribute this lack of change to its communist past, as Howard does. 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. In short, Howard's conclusions are justifiable, but not on the basis of most-different analysis.

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 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.⁸⁶ It is rarely, if ever, a stand-alone method.⁸⁷

Conclusion

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 formed the basis of Chapter Three.)

⁸⁵ *Ibid.*, 8.

⁸⁶ See, e.g., Collier and Collier (1991/2002); Karl (1997); Moore (1966); Skocpol (1979); and 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 non-petro-states (e.g., Chapter 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.

⁸⁷ 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).

Let us begin with a methodological obstacle that is faced by both Mill-lean styles of analysis – the necessity of dichotomizing every variable in the analysis. Recall that, as with most-similar analysis, differences across cases must 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 Mill-lean-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, though Russia is considerably lower. Howard assumes that this divergence is minimal enough to be understood as a difference of degree rather than of kind, 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 useful for *eliminating necessary causes*. Causal factors that do not appear across the chosen cases – e.g., X_{2a-d} in Table 5.6 – 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 shared element across cases – X_1 in Table 5.6. 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 5.6, 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 on the elimination of possible causes. In this setting, the most-different technique is useful, but only if assumptions of “causal uniqueness” hold. By this 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_1 . X_1 is necessary, and in some situations (given certain background conditions) sufficient, to cause Y .⁸⁸

Consider the following hypothetical example. Suppose that a new disease, about which little is known, has appeared in Country A. There are

⁸⁸ Another way of stating this is to say that X is a “nontrivial necessary condition” of Y .

Conclusion

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 other than 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.⁸⁹ Lisa Martin 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.”⁹⁰ Barbara Geddes charges that many analyses of industrial policy have focused exclusively on the most successful cases – primarily the East Asian NICs – leading to biased inferences.⁹¹ Anna Breman and Carolyn Shelton 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).⁹²

⁸⁹ Achen and Snidal (1989); Collier and Mahoney (1996); Geddes (1990); King, Keohane and Verba (1994); Rohlfing (2004); Sekhon (2004). Some case study researchers appear to denigrate the importance of case representativeness. George and Bennett (2005: 30) write emphatically, “Case researchers do *not* aspire to select cases that are directly ‘representative’ of diverse populations and they usually do not and should not make claims that their findings are applicable to such populations except in contingent ways.” However, it becomes clear that what the authors are inveighing against is not the goal of representativeness *per se* but rather the problem of a case study researcher who claims an inappropriately broad extension for her findings. “To the extent that there is a representativeness problem or a selection bias problem in a particular case study, it is often better described as the problem of ‘overgeneralizing’ findings to types or subclasses of cases unlike those actually studied” (*ibid.*, 32).

⁹⁰ Martin (1992: 5).

⁹¹ Geddes (1990).

⁹² Breman and Shelton (2001). See also Gerring, Thacker, and Moreno (2005).

II. Doing Case Studies

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 Millian 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 (highly dissimilar) 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 may be a reasonable operating assumption when one is dealing with certain natural phenomena like diseases. Even so, there are 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 5.6). This assumption rarely holds in social scientific settings, for 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 supplement, a way of incorporating diversity in the subsample of cases that provide the unusual outcome of interest. If the unusual outcome is revolution, 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 earlier.

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 rather superficial 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 skepticism about case study work that one commonly encounters 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 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 of the same subject), problems of representativeness are less worrisome. This is the virtue of cross-level work, a.k.a. "triangulation."

Ambiguities

Before concluding, I wish to draw attention to two ambiguities in case-selection strategies for 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 all 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

II. Doing Case Studies

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.⁹³ Case study work is particularly prone to problems of investigator bias because so much rides on the researcher's selection of one case (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 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 in the latter instance, 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-homogenous 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

⁹³ 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."

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 case legitimately serves several functions at once (without further effort on the researcher's part), there is little cost to a multipronged 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 particularly acute when a researcher begins in an exploratory mode and then proceeds to hypothesis testing (that is, she develops a specific X_1/Y proposition), or when 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 that the correlation between these two stages of research is particularly strong.

Research, after all, is about discovery, not simply the verification or falsification of existing 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. I discussed this trade-off in Chapter Three as it pertains to the cross-case/case study dilemma. It also applies to various methods of case study analysis, as presented in this chapter. 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_1 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 be so well constructed.

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 consultation of experts. A final approach is to simply jettison, or deemphasize, that portion of the 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. The point to remember is that revision of one's cross-case research design is *normal* and to be expected. Not all twists and turns on the meandering trail of truth can be anticipated.

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, the problem (sometimes known as Galton's problem or a problem of diffusion) 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 easy techniques that might be utilized to correct for such biases.⁹⁴

I have also disregarded *pragmatic/logistical* issues that might affect case selection. Evidently, case selection is often influenced by a researcher's

⁹⁴ For further discussion of this and other factors impinging upon case selection, see Gerring (2001: 178–81).

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.”⁹⁵

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, as I have elsewhere in the book.

One final caveat must be issued. While it is traditional to make a distinction between the tasks of case selection and case analysis, a close look at these processes reveals 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 (Chapter Six) and its potential for process tracing (Chapter Seven), subjects to which we now turn.

⁹⁵ Flyvbjerg (2004: 427).

Internal Validity

An Experimental Template

Let us suppose that one has chosen one’s case (or cases) according to one of the techniques (or some combination of techniques) described in the previous chapter. And let us further suppose that one has refined one’s research-question into a specific (X_1/Y) hypothesis. One then faces a problem of internal validity: How does one construct a research design that might illuminate the causal relationship of interest?

The fundamental problem of causal inference is that one cannot rerun history to see what effects X_1 actually had on Y in a particular case. At an ontological level, this problem is unsolvable. There are no time machines. However, there are various ways of reducing uncertainty so that causal inference is possible, and indeed quite plausible. The argument of this chapter is that the various methods of doing so are all quasi-experimental in nature. This is because the true experiment is the closest approximation we have at our disposal to a time machine. Through this technique, and others modeled on it, one can imagine what it would be like to go back in time, alter a “treatment,” and observe its true causal effect.

This chapter thus calls into question some of the usual assumptions applied to case study research. Most case study researchers perceive only a distant and tenuous connection between their work and the laboratory experiment, with a manipulated treatment and randomized control. They are inclined to the view that experimental and observational work inhabit different worlds, perhaps even employ different logics of inquiry. Granted, case study researchers working with observational data occasionally refer to their work as “quasi-experiments,” “natural experiments,” “thought experiments,” “crucial experiments,” or “counterfactual thought experiments.” However, these designations are often loose and ambiguous.