Selecting Cases for Intensive Analysis: A Diversity of Goals and Methods

John Gerring¹ and Lee Cojocaru¹

Abstract
This study revisits the task of case selection in case study research, proposing a new typology of strategies that is explicit, disaggregated, and relatively comprehensive. A secondary goal is to explore the prospects for case selection by algorithm, aka ex ante, automatic, quantitative, systematic, or model-based case selection. We lay out a suggested protocol and then discuss its viability. Our conclusion is that it is a valuable tool in certain circumstances, but should probably not determine the final choice of cases unless the chosen sample is medium-sized. Our third goal is to discuss the viability of medium-n samples for case study research, an approach closely linked to algorithmic case selection and occasionally practiced by case study researchers. We argue that medium-n samples occupy an unstable methodological position, lacking the advantages of efficiency promised by traditional, small-n case studies but also lacking the advantages of representativeness promised by large-n samples.

Keywords
case study, case selection, qualitative methods, methodology, social science

¹ Department of Political Science, Boston University, Boston, MA, USA

Corresponding Author:
John Gerring, Department of Political Science, Boston University, 232 Bay State Rd, Boston, MA 02215, USA.
Email: jgerring@bu.edu
Given the central role of case selection in case study research, it is essential that we come to terms with the available options. To this end, a number of case selection options have been proposed over the years. Nearly two centuries ago, Mill ([1843] 1872) devised the method of difference (aka most-similar method) and method of agreement (aka most-different method), along with several others that have not gained traction. More recently, Lijphart (1971:691) proposes six case study types: a-theoretical, interpretative, hypothesis-generating, theory-confirming, theory-infirming, and deviant. Eckstein (1975) identifies five species: configurative-idiographic, disciplined-configurative, heuristic, plausibility probes, and crucial-case. Skocpol and Somers (1980) identify three logics of comparative history: macrocausal analysis, parallel demonstration of theory, and contrast of contexts. Gerring (2007a) and Seawright and Gerring (2008) identify nine techniques: typical, diverse, extreme, deviant, influential, crucial, pathway, most-similar, and most-different. Levy (2008) identifies five case study research designs: comparable, most-likely, least-likely, deviant, and process tracing. Rohlfing (2012:chapter 3) identifies five case types—typical, diverse, most-likely, least-likely, and deviant—which are applied differently according to the purpose of the case study. Blatter and Haverland (2012:24-26) identify three explanatory approaches—covariational, process tracing, and congruence analysis—each of which offers a variety of case selection strategies.

The typological exercise is clearly irresistible and, by all accounts, useful. Most case study work employs one of these methods, and the foregoing studies are widely cited. Nonetheless, several problems persist among extant typologies. First, many typologies are not very explicit about how rules for case selection ought to be employed, that is, which features of a prospective case should determine its selection. Second, extant typologies often conflate disparate strategies of case selection. (We show, e.g., that selecting cases for causal analysis is quite different from selecting cases for descriptive analysis and that commonly used strategies such as the “most-similar” method are more variegated than is generally acknowledged.) Third, extant typologies are incomplete. In this study, we propose a new typology that is explicit, disaggregated, and relatively comprehensive—thus rectifying, or at least ameliorating, some of the limitations of extant typologies.

Our second goal is to explore the prospects for case selection by algorithm, aka ex ante, automatic, quantitative, systematic, or model-based case selection. This general approach has been advocated in several recent studies (e.g., Dafoe and Kelsey 2014; Fearon and Laitin 2008; Lieberman 2005; Pinfari 2012; Seawright and Gerring 2008). However, the pros and cons of algorithmic case selection have never been fully explored, and the option is
generally ignored by case study researchers—perhaps because it is not clear what it entails or how it might be profitably employed. We lay out a suggested protocol for algorithmic case selection and then discuss its viability. Our conclusion is that it is a valuable tool in certain circumstances but should probably not determine the final choice of cases, unless the chosen sample is medium-sized.

Our third goal is to discuss the viability of medium-n samples for case study research, an approach closely linked to algorithmic case selection and one explicitly advocated by Fearon and Laitin (2008) and occasionally practiced by case study researchers. We argue that medium-n samples occupy an unstable methodological position, lacking the advantages of efficiency promised by traditional, small-n case studies, but also lacking the advantages of representativeness promised by large-n samples. While it may be useful in some circumstances, we do not foresee widespread adoption.

The first section lays out the typology. The second section introduces important clarifications and caveats pertaining to the typology. The third section discusses the prospects, and limitations, of algorithmic case selection. The fourth section comments on the viability of medium-n samples in case study research. The final section emphasizes the importance of transparency in case study research and especially in case selection.

**A Case Selection Typology**

A case study, for present purposes, is an intensive study of a single case or a small number of cases that promises to shed light on a larger population of cases. The evidence for a case study is presumed to be observational insofar as the researcher cannot directly manipulate the causal treatment. Case selection refers to the method by which case(s) are chosen for an intensive investigation.

The organizing feature of the proposed typology, summarized in Table 1, is the goal that a case study is intended to serve. Case studies serve a wide variety of functions, and these functions rightly structure the case selection process. The most fundamental issue is whether the case study aims for descriptive or causal inference. If the aim is causal, case studies may be further subdivided according to their specific function—exploratory, estimating, or diagnostic. For each general aim or specific function, there are several viable approaches to case selection indicated by bullet points in column 1.

Column 2 specifies the number of cases (n) in the case study. It will be seen that case studies enlist a minimum of one or two cases, with no clearly defined ceiling.
In distinguishing different case selection techniques, it is vital to clarify what the researcher knows, ex ante, prior to conducting the case study. Some aspects of a case are relevant for case selection while other aspects are irrelevant, and perhaps even counterproductive insofar as they may undermine the researcher’s goals in conducting the case study. For example, if the goal of a case study is to estimate causal effects (for a population), it would be pointless to choose cases based on their values for the outcome, $Y$. Such a procedure, known as “cherry-picking,” would lead to an obviously biased estimate.

Column 3 clarifies which dimensions (factors) of a case are relevant for case selection. This includes descriptive features ($D$), the presumed cause of theoretical interest ($X$), background factors of no theoretical interest ($Z$), and the outcome ($Y$). If selection is on $Y$, this means that the researcher knows the value of the outcome for cases that are under consideration for intensive study, but not the values of $X$ or $Z$. If selection is on $Z$ and $Y$, the researcher is assumed to know the values for these dimensions of the case, but not the value for $X$. And so on.

In distinguishing different case selection techniques, it is vital to clarify what the researcher knows, ex ante, prior to conducting the case study. Some aspects of a case are relevant for case selection while other aspects are irrelevant, and perhaps even counterproductive insofar as they may undermine the researcher’s goals in conducting the case study. For example, if the goal of a case study is to estimate causal effects (for a population), it would be pointless to choose cases based on their values for the outcome, $Y$. Such a procedure, known as “cherry-picking,” would lead to an obviously biased estimate.
Column 4 specifies the criteria used to select a case(s) from a universe of possible cases. Note that all case selection strategies rest upon a case’s relationship to a larger population of cases. For example, the deviant case is that case(s) in the studied population (or the observed sample) exhibiting the most deviant features. In addition, two general case selection strategies—most-different and most-similar—require the researcher to consider the status of the chosen cases relative to each other. Note also that virtually all case selection strategies may be executed in an informal (“qualitative”) fashion or by employing a (“quantitative”) algorithm, an issue taken up at length below.

In what follows, we outline each case selection technique, offer one or few examples, and suggest ways in which the case might be chosen algorithmically from a large population of potential cases.

**Descriptive**

Some case studies are concerned primarily with describing a phenomenon. These sorts of case studies generally employ a typical or diverse approach to case selection.

**Typical.** A typical case is intended to represent the central tendency of a distribution, which is of course not the same as the entire distribution. (To say that a case is typical, therefore, does not mean that it is representative in the way that a larger sample might be representative of a population.) For example, the Lynd’s (1929) study of “Middletown” focuses on a city (Muncie, IN) that was thought to be typical of mid-sized cities across the United States. Ladurie (1978) focuses on a village in France (“Montaillou”) that is thought to be typical of other villages in the late Middle Ages. And so on. To identify typical case(s) from a large population of potential cases, researchers may choose case(s) that lie close to the mean, mode, or median of a distribution.

**Diverse.** A descriptive case study might also focus on several cases that, together, are intended to capture the diversity of a subject. For example, Almond and Verba ([1963] 1989) choose to study political cultures in the United States, Germany, Mexico, Italy, and the United Kingdom with the idea that these countries, together, represent the diversity of political cultures in the world—which they conclude may be summarized in three ideal types: participant, subject, and parochial. To identify a small basket of diverse cases from a large population of potential cases, the researcher may choose cases with different scores (e.g., high, medium, and low) on parameters of interest,
look at the intersection of different parameters (broken into discrete categories, if the variables are interval level), or utilize techniques of factor analysis or cluster analysis.

Causal

A causal case study is organized around a central argument about a change in $X$ that generates a change in $Y$. While $X$ usually refers to a single factor, it may also include several related factors or even all the causes of $Y$ (a causes-of-effects analysis). When used for causal inference, case studies may be categorized as exploratory, estimating, or diagnostic. Before delving into details, several fundamental points bear emphasis.

First, most case studies do not attempt to estimate a precise causal effect and an accompanying confidence interval, as would be expected from a large-$n$ cross-case research. Samples of one or several are not well suited to estimate population parameters. The notion of causal inference, as the term is used here, encompasses any statement about the impact of $X$ on $Y$, whether precise (e.g., “An increase in one unit of $X$ generates a two-unit increase in $Y$”) or imprecise (e.g., “An increase in $X$ causes an increase in $Y$”).

Second, a good case (or set of cases) for purposes of causal analysis is generally one that exemplifies quasi-experimental properties, that is, it replicates the virtues of a true experiment even while lacking a manipulated treatment (Gerring and McDermott 2007). Specifically, for a given case (observed through time) or for several cases (compared to each other), variation in $X$ should not be correlated with other factors that are also causes of $Y$, which might serve as confounders, generating a spurious (noncausal) relationship between $X$ and $Y$.

Third, case selection criteria may be understood cross-sectionally or longitudinally. For example, the test of “deviance” might be a case’s status at a particular point in time, or its change in status over an observed period of time. Generally, cases exhibiting change on key parameters of interest are more informative than cases that remain static. Thus, wherever possible, researchers should administer case selection strategies using information about how cases perform through time, in addition to how they compare to other cases at a particular point in time.

Causal: Exploratory

Many case studies are exploratory, or hypothesis-generating, insofar as they aim to identify a possible cause of an outcome. The outcome, $Y$, is
established, and usually framed as a research question. What accounts for variation in Y? Or, if Y is a discrete event, Why does Y occur? The researcher may also have an idea about background conditions, Z, that influence Y but are not of theoretical interest. The purpose of the study, in any case, is to identify X, regarded as a possible or probable cause of Y. Specific exploratory techniques may be summarized as outcome, index, deviant, most-different, most-similar, or diverse.

**Outcome.** An outcome case maximizes variation on the outcome of interest. This may be achieved by a case that exhibits extreme values on Y (or ΔY). Note that studies of welfare state development often focus on the world’s largest welfare states, located in Northern Europe. Studies of war often focus on one of the two world wars. Studies of genocide often focus on the Holocaust. And so on. A second version of an outcome case applies, when an outcome is conceptualized in a binary fashion and one value (generally understood as the “positive” value) is especially rare. Case studies of war generally focus on wars—peace being the more common condition. Case studies of democratization generally focus on regime transitions—continuity being the more common condition. And so on. A third approach to achieving variation in the outcome is to choose cases lying at both tails of the distribution, that is, polar cases. Here, comparisons can be made directly across the chosen cases. For example, a study of war may focus on cases exhibiting peace and war (e.g., Fearon and Laitin 2008, 2014, 2015). All of these designs maximize variation on Y, preferably variation through time (ΔY). Our nomenclature (“outcome case”) thus subsumes cases that might be described as extreme, rare, or polar. To identify an outcome case(s), one may employ an algorithm, selecting a case(s) that lies farthest from the mean, median, or mode of the distribution of Y (or ΔY), or cases that lie on both tails of the distribution of Y (or ΔY).

**Index.** An index case is the first instance of a phenomenon. In epidemiology (where we appropriate the term), it refers to the first patient to contract a disease. Identifying the index case is useful for understanding the origin and spread of disease and for that reason a good deal of effort goes into the search for an index case, aka patient zero. In other social science disciplines, the focus on index cases is also motivated by the desire to understand the origins of a phenomenon. The presumption is that the index case plays a causal role—as a prime mover or as an illustrative or standard-setting case, establishing a practice to emulate or avoid. The primacy of “firsts” is imprinted in natural language and in the social science lexicon, wherever new phenomena
are named after the person, place, or event of their first occurrence. Additionally, the first instance of a phenomenon is uninfluenced by other instances of that phenomenon. It occurs endogenously, avoiding a common problem of nonindependence across units (aka diffusion or Galton’s problem) that often interfere with causal identification. To choose one or several pioneer case(s) from a large population of potential cases, researchers may look for the first instance(s) of a change in $Y$.

**Deviant.** A deviant case deviates from an expected causal pattern, as suggested by theories or common sense, registering a surprising result. The study of deviance may also be framed as the study of anomalies (C. Elman and Elman 2002; Lakatos 1978). The goal of a deviant case study is to explain the oddball case and, in addition, to explain other similarly deviant cases, providing a generalizable hypothesis about the phenomenon of interest. The resulting explanation may involve a new causal factor, an interaction (aka contextual) effect between two or more causal factors, or a revision to the scope conditions of a theory. For example, to shed light on the causes of mortality decline in the developing world, Caldwell (1986) focuses on three cases that, relative to their level of economic development, experienced low mortality in the present era—Costa Rica, Sri Lanka, and Kerala (a state in India). On the basis of these cases (which are contrasted schematically with other developing countries), Caldwell concludes that female autonomy and local health care play a critical role in fostering human development. In a regression model, deviance can be measured by the residual for a given case.

**Most-similar.** Sometimes, one can identify cases that exhibit strong similarities on background conditions ($Z$) but divergent outcomes ($Y$). The generic label for this research design is most-similar (aka method of difference). For example, in a recent medical study, Rosenbaum and Silber (2001) compare patients within a hospital who have similar symptoms, some of whom survive their spell of hospitalization and others of whom die. The goal is to ascertain the causes of death by probing differences across these closely matched cases using information gleaned from a close examination of Medicare records. To identify a small number of most-similar cases from a large population of potential cases, the researcher may employ matching algorithms that minimize differences on the vector $Z$ while maximizing differences on $Y$ (Nielsen 2016).¹

**Most-different.** Most-different cases (aka the method of agreement) vary widely in background factors regarded as potential causes ($Z$), while sharing...
a common outcome ($Y$). The assumption is that background factors that differ across the cases are unlikely to be causes of $Y$ since that outcome is constant across the cases. The hope is that if a factor ($X$) can be identified that is constant across the cases it may be the cause of $Y$. For example, in searching for the cause of a disease contracted in hospitals, it makes sense to compare hospitals where the disease manifested itself that are very different from each other, for example, located in different areas or serving different populations. To identify a small number of most-different cases from a large population of potential cases, the researcher may seek to maximize differences on the vector $Z$ while minimizing differences on $Y$.

**Diverse.** A final exploratory strategy has as its objective the identification of many—or perhaps all—of the causes of an outcome (assuming causal equifinality). The chosen cases are diverse if they represent all potential factors ($Z$), including causal conjunctures, that might explain variation in $Y$. The assumption is that the true causal factors ($X$) are to be found among the putative causal factors ($Z$). For example, George and Smoke (1974:522-36, chapter 18; see also D. Collier and Mahoney 1996:78) wish to explore different types of deterrence failure—by “fait accompli,” by “limited probe,” and by “controlled pressure.” Consequently, they wish to find cases that exemplify each causal factor. Where the potential causal factor is categorical (on/off, red/black/blue, and Jewish/Protestant/Catholic), the researcher would normally choose one case from each category. For a continuous variable, one must construct cutoff points (based on theoretical understandings of the phenomenon or natural breakpoints in the data), for example, dichotomizing or trichotomizing the variable, and then choosing cases with each discrete value. If one suspects that causal factors interact, then one will look for cases that represent all possible (or actual) intersections of these variables (understood as categorical variables). Two dichotomous variables produce a matrix with four possible cells, for example. Note that where multiple categorical variables interact, the logic of diverse-case analysis rests upon a *typological* logic (C. Elman 2005; George and Bennett 2005; Lazarsfeld and Barton 1951). In choosing a small basket of diverse cases from a large population of potential cases, the researcher may draw on qualitative comparative analysis (QCA) algorithms to identify the various possible conjunctures, selecting case(s) from each configuration. Alternatively, within a regression framework, the researcher may explore various interaction effects, choosing cases that exemplify disparate interactions.
Causal: Estimating

We turn now to case studies whose goal is to test a hypothesis by estimating a causal effect. Estimating may mean a precise point estimate along with a confidence interval (as might be obtained from a time-series or synthetic matching analysis), or a less precise estimate of the “sign” of a relationship, that is, whether $X$ has a positive, negative, or no relationship to $Y$. The latter is more common, as we have observed, not only because of the small size of the sample (at the case level) but also because it is more likely to be generalizable. In either situation, case selection must rest on information about $X$ and $Z$—not $Y$. Two general approaches are viable for estimating causal effects in a small-$n$ setting—longitudinal and most similar—as outlined below.

Before entering this discussion, it may be appropriate to observe that this is not the most common use of case study research. This arises from a logistical feature of the world. In situations where a researcher has identified a specific hypothesis ($H_x$) that pertains to a broad population, it is usually possible to estimate that relationship across a large number of cases, generating a large-$n$ cross-case research design. This mode of analysis usually—although not invariably—is able to provide a stronger test of the hypothesis and is therefore the preferred research design.

Nonetheless, there are some circumstances in which it is possible, and advisable, to estimate the impact of $X$ on $Y$ using a very small sample of cases. The justification for this small-sample approach lies in the inability, or inadvisability, of extending the sample to include additional cases. That is, the longitudinal evidence provided by a single case, perhaps accompanied by one or several “control” cases, provides stronger grounds for inference than the corresponding large-$n$ design—presumably because additional cases are heterogeneous and would introduce potential confounders to the analysis.

Longitudinal. A longitudinal case study mimics a one-group experiment, where $X$ changes in an as-if random fashion while $Z$ remains constant and $Y$ is observed before and after the intervention. It might also be referred to as an interrupted time-series (Campbell [1968] 1988) or a repeated measures (or repeated observations) design. For example, in order to understand the interrelationship between monetary policy and economic fluctuations, Friedman and Schwartz (1963) look closely at U.S. history, locating four occasions in which the stock of money changed due to policy choices largely unrelated to the behavior of the economy (and hence exogenous to the research question). These four interventions consisted of “the increase in the discount rate in the first half of 1920, the increase in the discount rate in
October 1931, the increase in reserve requirements in 1936–1937, and the failure of the Federal Reserve to stem the tide of falling money in 1929–1931” (Miron 1994:19). It turns out that each was followed by a substantial change in the behavior of the stock of money, validating a central pillar of monetarist theory. To identify a longitudinal case(s) from a large population of potential cases, one may look for instances where change in $X$ is not accompanied by a change in $Z$.

**Most similar (Estimating).** Estimating a causal effect with a most similar design is similar to the longitudinal design with the notable addition of a control case—which (ideally) experiences no change on either $X$ or $Z$. That is, chosen cases exhibit different values on $X$ and similar values on $Z$. Under these circumstances, and with a variety of assumptions, realized outcomes across the cases ($Y$) allow one to estimate a causal effect. Differences and similarities may manifest themselves in a binary fashion, or they may be matters of degree. If the latter, one seeks cases that maximize variance on $X$ and minimize variance on $Z$. For example, Mondak (1995) examines two cities—Cleveland and Pittsburgh—that are similar in background conditions. One of the cities experiences a newspaper strike, exhibiting a change in the causal factor of theoretical interest. The goal of the study is to determine to what extent the absence of a newspaper affects citizen knowledge of politics, as measured through a posttest survey of political awareness. To identify a small number of most-similar cases from a large population of potential cases, the researcher may employ matching algorithms that maximize differences across cases in $X$ (or $\Delta X$) while minimizing differences on the vector $Z$ (or $\Delta Z$; Nielsen 2016).²

**Causal: Diagnostic**

Case studies, finally, may perform a diagnostic function—helping to confirm, disconfirm, or refine a causal hypothesis (garnered from the literature on a subject or from the researcher’s own ruminations) and identifying the generative agent at work in that relationship. Specific strategies may be classified as influential, pathway, or most-similar, as discussed below.

Since all elements of a causal model—$X$, $Z$, and $Y$—are generally involved in the selection of a diagnostic case, the reader may wonder what is left for case study research to accomplish. Actually, a good deal remains on the table. Information gleaned from a diagnostic case study may be used to confirm, reject, or revise a theory or to refine a large-$n$ cross-case model, for example, by helping to respecify that model (Gordon and Smith 2004; Seawright 2015). More specifically, diagnostic case studies may assess:
• **Measurement error:** Are $X$, $Z$, and $Y$ properly measured? (Seawright 2015)

• **Scope conditions:** Is the chosen case rightly classified as part of the population? What are the appropriate scope conditions for the hypothesis? (Skocpol and Somers 1980)

• **Causal heterogeneity:** Are there background factors that mediate the $X \rightarrow Y$ relationship?

• **Confounders:** Is the actual data generating process consistent with the chosen causal model? Are there pre- or posttreatment confounders? (Dunning 2012; Seawright 2015)

• **Causal mechanisms:** What is the pathway through which $X$ affects $Y$? (see below)

Of particular interest is the latter feature—the mechanisms ($M$) connecting $X$ to $Y$. Mechanisms not only help to confirm or disconfirm a hypothesis, they also explain it, since in specifying a mechanism we also specify the generative process by which $X$ causes $Y$. Note that when there is no strong theoretical prior about the nature of the mechanism, the case analysis assumes an open-ended, inductive quest—to identify $M$. When there is a theoretical expectation, the analysis assumes a deductive format—to test the existence of a prespecified pattern thought to be indicative of $M$ (e.g., Dafoe and Kelsey 2014). This is sometimes referred to as congruence testing, pattern-matching, or implication analysis.

Diagnostic cases are probably the most complex sort of case study from a research design perspective, so this section is somewhat longer than the previous sections.

**Influential.** An influential case is one whose status has a profound effect on the probability of a hypothesis being true, $P(H_x)$. If an intensive study of that case reveals measurement error, faulty scope conditions, causal heterogeneity, confounders, or errors in causal mechanisms the hypothesis will be called into question. If, on the other hand, problems are not discovered in these areas—and a clear mechanism can be traced connecting $X$ with $Y$—the hypothesis is corroborated.

In social science settings, the most influential cases are usually those that falsify, or threaten to falsify, a hypothesis. Decisively corroborating cases are rare. But there are many instances in which a single case, or a small set of cases, disconfirms—or at least casts a pall of suspicion on—a hypothesis. The most influential case is one that, by itself, falsifies a hypothesis. This is possible if the proposition is strictly deterministic (Dion 1998). Suppose that
$X = 1$ is regarded as a necessary condition of an outcome, $Y = 1$. A falsifying case would have the attributes $X = 0$, $Y = 1$. If, on the other hand, $X = 1$ is regarded as a sufficient condition of an outcome, $Y = 1$, a falsifying case would exhibit the attributes $X = 1$, $Y = 0$. (Necessary and sufficient conditions are mirror images of each other; which terminology one uses is a matter of clarity and convenience.)

The most prominent deterministic hypothesis in social science today is probably the democratic peace—the idea that democratic dyads do not wage war on each other. As it happens, there are a number of possible exceptions to this “universal” law including the Spanish–American War, the Kargil War, the Paquisha War, Lebanese participation in the Six Day War, the French–Thai War, and the Turkish invasion of Cyprus. Each has drawn considerable attention from supporters and skeptics of the democratic peace hypothesis (Bremer 1992, 1993; M. F. Elman 1997; Gleditsch 1992; Owen 1994; Ray 1993). Thus far, this research has not yielded a knockout blow to democratic peace theory. There is a problem of measurement insofar as apparent exceptions are often hard to classify cleanly as democratic/autocratic or peaceful/belletristic. More fundamentally, many researchers take a probabilistic view of the theory—in which case, a few exceptions are not so worrisome for the theory. For present purposes, what is important is that these cases are influential. Whether the theory is interpreted as deterministic or probabilistic, these cases have greater bearing on the validity of the theory than other cases that might be chosen for intensive analysis, which explains their enduring importance. Influential cases are often outliers, or apparent outliers—cases that don’t seem to fit the theory, as seen in the case of the democratic peace. This is true for other subjects as well, including subjects that are understood by pretty much everyone as probabilistic.

While influential cases are sometimes outliers—and in this sense, mirror deviant cases (discussed above)—they may also be conforming. Indeed, they may define the relationship of interest. What makes a case influential is not its model fit but its influence on the model. Likewise, even when an influential case is deviant the purpose of these two genres is quite different. A deviant case is designed for discovery (to identify a new hypothesis), while an influential case is designed for diagnostic purposes (to assess an existing hypothesis).

Influential cases may take the form of crucial cases, if certain background conditions hold (Eckstein 1975). If the goal is to prove a hypothesis, the crucial case is known as a least-likely case. Here, the hypothesized relationship between $X$ and $Y$ holds even though background factors ($Z$) predict otherwise. With respect to democratic peace hypothesis, a least-likely case
would be a dyad composed of two democratic countries with background characteristics that seem to predispose them to war but nonetheless are at peace. If the goal is to disprove a hypothesis, the crucial case is known as a *most-likely* case. Here, the hypothesized relationship between \( X \) and \( Y \) does not hold even though background factors \( Z \) predict that it should. With respect to the democratic peace hypothesis, a most-likely case would be a dyad composed of two democratic countries with background characteristics that seem to predispose them to peace (e.g., they are rich, culturally similar, and economically codependent), who nonetheless engage in violent conflict. Distinguishing most- from least-likely cases depends upon the hypothesis, which may be formulated in different ways. For example, if one chooses to frame the outcome as “war” rather than “peace” (an arbitrary decision, in most respects), then the terms are flipped. The logic of the crucial case remains the same, in any case.

Algorithms for choosing influential cases depend on the purpose and context. To identify influential case(s) for deterministic hypotheses from a large population of potential cases, the researcher needs only to compare values for \( X \) and \( Y \), looking for those that seem to disconfirm the theory. To identify influential case(s) for probabilistic hypotheses from a large population of potential cases in a regression context, the researcher may draw on a well-developed body of influence statistics designed to identify those cases that play an influential role—understood as cases which, if removed from the sample, would have the largest impact on the total model or—more usefully—on estimated coefficients for a particular independent variable, as revealed by the DFBETA statistic (Andersen 2008; Belsey, Kuh, and Welsch 2004; Bollen and Jackman 1985).

**Pathway.** A pathway case is one where the apparent impact of \( X \) on \( Y \) conforms to theoretical expectations and is strongest (in magnitude), while background conditions \( Z \) are held constant or exert a “conservative” bias (Gerring 2007a, 2007b; Weller and Barnes 2014). This might also be called a *conforming* or *typical* case (Lieberman 2005; Schneider and Rohlfing 2016), since it conforms to or typifies a causal relationship of interest. However, the ideal pathway case does more than simply conform to an expected pattern.

In a setting where the relationship between \( X \) and \( Y \) is highly uncertain—perhaps because it has not yet been (or cannot be) tested in a large-\( n \) cross-case format—the pathway case serves an illustrative function. By showing that the theory fits the chosen case, the case study illustrates the contents of the theory and demonstrates its plausibility. If it works here, it may apply
elsewhere. For example, in presenting the theory of path dependence, David (1985) draws on the curious case of the “QWERTY” keyboard.

In a setting where the relationship between $X$ and $Y$ is well established—perhaps as a result of cross-case analysis (the researcher’s or someone else’s)—the pathway case is usually focused specifically on causal mechanisms ($M$). An example is provided by Mansfield and Snyder’s (2005) research on regime transitions and war. The authors find a strong relationship between democratization and bellicose behavior in their large-$n$ cross-national analysis. To ascertain whether their hypothesized causal mechanisms are actually at work in generating this relationship, they look closely at 10 countries where the posited covariational pattern between $X$ and $Y$ clearly holds, that is, where democratization is followed by war.

Algorithmic approaches to pathway case selection are discussed in Gerring (2007a, 2007b) and Weller and Barnes (2014).

**Most-similar (Diagnostic).** When employed for diagnostic purposes, the most similar design consists of a pathway case (as above) plus a control case, which exhibits minimal variation in $X$ and $Z$. That is, chosen cases exhibit different values on $X$, similar values on $Z$, and different values on $Y$. In this context, one is more or less assuming that values for $X$, $Z$, and $Y$ represent values for the cases that are realized over time rather than at just one point in time. Thus, one may read $X$, $Z$, and $Y$ as $\Delta X$, $\Delta Z$, and $\Delta Y$. As with other most-similar designs, differences and similarities may manifest themselves in a binary fashion or they may be matters of degree. If the latter, one seeks cases that maximize variance on $X$ and $Y$ and minimize variance on $Z$. These research design features may be identified in a large sample of potential cases by employing matching algorithms that maximize differences across cases in $X$ and $Y$ (or $\Delta X$ and $\Delta Y$) in ways that are consistent with the hypothesis while minimizing differences on the vector $Z$ (or $\Delta Z$; Nielsen 2016).

For example, Lutfey and Freese (2005) are interested in uncovering the mechanisms at work in a persistent, and oft-noted, relationship between socioeconomic status and health. Poor people experience poor health, which is presumably—at least in some respects—a product of their poverty. The researchers compare high- and low-status individuals who suffer from diabetes, with the knowledge that the latter are more likely to succumb to the effects of the disease. This is accomplished by focusing on two endocrinology clinics, one located in an affluent neighborhood and the other in a poor neighborhood. The focus of the study is on factors inside the clinic (continuity of care, in-clinic educational resources, and bureaucratic organization), outside the clinic (financial limitations, occupational constraints, and
social support networks), and among the patients (motivation, cognitive ability) that might affect compliance with an exacting medical regime. These are regarded as prima facie causal mechanisms in the relationship between socioeconomic status and health.

**Clarifications**

The typology presented in the previous section requires some caveats and clarifications. The following points pertain to the goals, scope, configuration, and interpretation of the typology summarized in Table 1.

1. The *scope* of the typology is intended to encompass case selection practices commonly used within the social sciences (including borderline disciplines such as history and psychology). It does not extend to case studies whose selection procedures are highly idiosyncratic.

2. The goal of the typology is primarily *descriptive*—to describe how case studies are conducted. But it also contains some *prescriptive* elements. Once a researcher has defined the goal a case study is intended to serve, we argue that there are a limited set of strategies available—identified by bullet points in Table 1. Which strategy is ultimately chosen from this menu depends upon various contextual factors that we are not in a position to prejudge. Nor are we in a position to make strong statements about the internal or external validity of case studies resulting from these case selection strategies.

3. The typology encompasses cases deemed appropriate for intensive analysis, not those that provide brief points of comparison for the case(s) of primary interest, that is, *shadow* cases.

4. The typology is *lengthy* (although most other typologies are similar in length). While it would be pleasing to be able to reduce the complexity of case selection in case study research to a small number of core strategies, this sort of reductionism may cause greater confusion downstream by conflating goals or techniques that are, in important respects, distinct. For example, we argue that there are three versions of the most-similar design—exploratory (selecting on $ZY$), estimating (selecting on $XZ$), and diagnostic (selecting on $XZY$). Each serves quite different functions and, because selection is on different factors, each is likely to result in different choices.

5. The typology bears on the *initial* decision to select cases. Usually, this decision is informed by a central objective, which we have
classified as descriptive or causal and, if the latter, as exploratory, estimating, or diagnostic. At the same time, we should recognize that once a case is chosen, it is likely to be exploited for all the information it can render. This includes possible hypotheses about \( Y \) (which may be regarded as part of a total explanation or as rival explanations vis-à-vis a favored hypothesis), and for each hypothesis, the causal effect, the mechanism, the scope conditions, possible causal heterogeneity, and potential confounders. Since case studies are informative along so many dimensions, it is often difficult to discern the initial goal and technique employed for case selection—unless the author is scrupulous in revealing this information, as discussed below.

6. With an eye to fulfilling multiple goals, researchers sometimes design case selection strategies with diverse objectives. For example, in work focused on the causes of civil war, P. Collier and Sambanis (2005a, 2005b) choose cases that maximize variation along several independent variables of interest—regime type, violence, ethnic fragmentation, and resource dependence (\( \text{diverse cases} \)). Next, they select countries that fit their cross-case model (\( \text{pathway cases} \)) and countries that do not (\( \text{deviant cases} \); Sambanis 2004:6). Other examples of “composite” case selection strategies exist (e.g., Fairfield 2013, 2015; Ostrom 1990; Pinfari 2012). However, incorporating multiple criteria in the selection of cases is relatively rare, and likely to remain so. In selecting a small basket of cases, one is forced to prioritize among various goals. Note that examples of composite case selection generally employ a fairly large basket of cases, qualifying these studies as medium-\( n \) rather than small-\( n \), as discussed below.

7. Case selection sometimes occurs across several levels. For example, Fairfield (2013, 2015) first selects countries and then tax reform proposals within a specified period for each country. The latter are referred to as cases because they constitute the sort of event her theory tries to explain. However, the selection of countries—Argentina, Bolivia, and Chile—gets little attention. If cases are chosen at several levels, each level constitutes a distinct case selection event and deserves to be treated as such.

8. The status of a case may change during the course of a researcher’s investigation, which may last for many years. For example, one might choose a single outcome case and later add a second case to form a most-similar analysis. If cases are not chosen all at once, there
is an issue of sequencing to resolve. Likewise, one might choose an outcome case and later decide that it conforms to a longitudinal design. Here, the case(s) remain the same, but the description of the case(s) changes. For example, P. Collier and Sambanis (2005a:27) note that their case selection guidelines “changed somewhat over time, as we moved away from the idea of using cases to test the theory and toward the idea of using the cases to develop theory and explore other issues such as mechanisms, sequences, measurement, and unit homogeneity.” The changing status of a case is virtually inevitable, when the researcher starts out in an exploratory mode. Exploratory methods of case selection are quite vague and are likely to morph into diagnostic designs, once a specific hypothesis has been identified.

9. This derives from an important feature of case study research: Case selection and case analysis are enmeshed. Indeed, the terms “case selection” and “research design” are virtually interchangeable. Choosing a case implies a method of analysis, but it does not entirely determine that method of analysis. Because case selection methods also describe methods of analysis, it is useful to state the role of a case, ex post. If a case chosen for one purpose ends up serving another purpose, this is an important information. And it need not cause confusion—so long as the researcher is careful to distinguish between the ex ante method of case selection and the ex post method of case analysis (e.g., Fairfield 2015:300).

10. This raises another issue, that of transparency. Researchers should be clear about how they chose their cases and about any changes in their treatment of those cases as the research progresses. As Alan Stuart (1984; quoted in Henry 1990:29) remarks, “The sample itself can never tell us whether the process that engendered it was free from bias. We must know what the process of selection was if we are not forever to be dogged by the shadow of selection bias.” Unfortunately, many researchers are not as forthright as Collier/Sambanis and Fairfield. This is especially true for older studies, where methodological issues are generally not front and center.

Several features enhance ambiguity. First, researchers sometimes mean different things when they invoke case-study terms and rarely do authors differentiate between versions of the same generic design (e.g., most-similar exploratory, estimating, and diagnostic designs). Second, it is sometimes difficult to differentiate studies whose main purpose is descriptive from those whose main purpose is causal. This
is because arguments are often rather loosely framed and may include a mixture of both elements. It is also because many case study researchers adopt a rather diffuse vision of causality. Third, it is often difficult to tell which features of the cases were known to the author prior to case selection. For example, it is often unclear when a researcher selected on \( X \) and when she selected on \( X \) and \( Y \). We return to the value of transparency in the concluding section of this study.

11. Finally, and perhaps most importantly, the typology focuses on features that differentiate case selection strategies. We should not lose sight of the fact that case selection also has some generic features—features applying broadly (although perhaps not universally).

First, the selection of cases is often influenced by the *intrinsic importance* of a case. Some cases—for example, world wars, genocides, key inventions, and revolutions—matter more than others because they have an obvious world historical significance. Others matter because they are important for a specific group of readers. If the intrinsic value of a case is the primary factor driving case selection, the resulting study may be described as *idiographic* (Eckstein 1975; Levy 2008; Lijphart 1971). Such studies appear to disavow any claims to generality and are thus not case studies by our definition. However, cases chosen for idiographic reasons may result in insights that have broader applicability. So the selection of a case, by itself, does not determine its future status as a case study. Indeed, there are plenty of examples of historical studies—usually selected for reasons of their intrinsic importance—that come to be regarded, much later, as case studies of a more general phenomenon.

Second, if a case study is designed to shed light on a causal question, the chosen cases should, ideally, be *independent* of each other and of other cases in the population. This is implicit in our understanding of a case as a relatively bounded unit. If cases are not bounded—if they affect each other with respect to the outcomes of concern—they are not providing independent evidence.

Third, if a case is to add to our knowledge of a subject, it must provide new evidence—evidence that is presumably not available—or not easily available or not in as precise or reliable form—for a larger sample. If sources are unreliable, scarce, or for one reason or another inaccessible, the case is of little value. Usually, this new evidence exists at a lower level of analysis, which we refer to as *within-case*. Within-case evidence is often the main value-added
offered by a case study relative to a cross-case analysis that has been, or might be, undertaken. Hence, the availability of within-case evidence plays a critical role in decisions about which case, or cases, to study.

Fourth, the availability of within-case evidence is partly a product of the case itself and partly a product of the researcher’s personal attributes—his or her linguistic competences, connections, and previous acquaintance with a region, time period, or topic. We assume that these logistical features are taken into account—implicitly if not explicitly—in any case selection process.

Finally, in order to be a case of something broader than itself—a “case study”—the chosen case must be representative of a larger population in whatever ways are relevant for the larger argument. Granted, one can never know for sure whether a sample is representative of a population, especially with respect to causal properties. Representativeness is a matter of probability. Even so, some cases are more likely to be representative than others. And some cases may be dismissed out of hand as self-evidently idiosyncratic.

Algorithmic Case Selection

While large-\( n \) cross-case research aspires to collect observations randomly from a known universe, case study research generally does not. Instead, researchers select on particular dimensions of the potential cases—\( D, X, Z, \) and/or \( Y \)—as summarized in Table 1. This is sometimes referred to as purposive case selection. However, there is no reason why a specification of dimensions cannot be combined with a random element, as in a stratified random sample. In doing so, the researcher may enlist an algorithm to guide the case selection process.

Algorithmic case selection follows a set of rules executed in a sequence of steps, which we envision as follows.

1. Define the research question and the population of theoretical interest.
2. Identify a sample of potential cases. Ideally, this sampling frame should be representative of the population of interest.
3. Measure relevant features of the cases—for example, \( D, X, Y, \) and/or \( Z \)—across the sample.
4. Combine diverse indicators of \( D, X, Y, \) and/or \( Z \) into indices, if necessary.
5. Construct a causal model, if required.
6. Apply an algorithm to identify the case, or cases, eligible for the investigation.
7. If more than one case satisfies the criteria, select cases randomly from within the subset of eligible cases (stratified random sampling).

Algorithms might be simple, as for the typical case, where the case(s) lying closest to the mean, median, or mode is usually regarded as sufficient (although things become more complicated, when there are many factors whose distributions across a population must be considered and weighted). They might also involve complex causal models using a variety of estimators—regression, matching, QCA, and so on.

Whatever the chosen technique, it deserves to be emphasized that all case selection strategies except some sorts of diagnostic cases may be implemented in an algorithmic fashion. Although this approach is still fairly unusual, it is now an established method of case selection and appears to be growing in importance, so it warrants close attention. We attempt to weigh the advantages, and then the disadvantages, concluding with a very tentative resolution of this issue.

**Advantages**

Four main advantages may be credited to the algorithmic approach. First, case selection is explicit and replicable. It would be difficult, by contrast, to describe and to replicate the complex judgments involved in qualitative case selection.

Second, the algorithm assists researchers in identifying a best case(s) wherever case selection criteria are complicated and/or the number of potential cases is large—perhaps numbering in the hundreds, thousands, or even millions. So long as a criterion can be reduced to a formula, and so long as it can be measured across the potential cases, the algorithmic approach is seaworthy.

Third, claims to sample representativeness are easier to define and to defend. That is, it is clearer what the chosen case(s) might be a case of (what population one is sampling from). Of course, that population might not be coterminous with the population of the hypothesis that results from the research. For example, if one chooses a case(s) to maximize variation in \( Y \), one would not claim that the resulting sample of outcome case(s) is representative of the population of cases for which \( X \rightarrow Y \) (once \( X \) has been identified). But one might claim that the sample is representative of a population for which \( Y \) is “high” or “low.”

Fourth, the algorithmic approach allows for a clear separation between theory generation and theory testing. Only in this fashion can the problem of cherry-picking (choosing cases that fit the researcher’s theory or preconceptions) be avoided.
For all these reasons, an algorithmic approach to case selection is worth considering.

**Disadvantages**

There are also reasons to be dubious of an “automatic” algorithmic procedure for selecting cases.

First, the protocol for algorithmic case selection outlined above is demanding. One must have a clearly defined research question and population of theoretical interest. One must possess data sufficient to measure relevant parameters for a large sample of cases, and the data must be relatively reliable. If a causal model is required, one must be able to construct a model that is plausible. If any of these requirements is not satisfied, one may be better off with an informal method of case selection.

Second, concerns about separating theory formation and theory testing are irrelevant if the goal of the case study is exploratory, that is, to discover a new cause of Y. Exploratory work always entails an interplay between theory and evidence (Rueschemeyer 2009), a function that the case study format is well suited to facilitate. Algorithmic case selection might compromise a key function of case-study research—if the researcher’s goal is exploratory.

Third, algorithmic case selection makes sense when choosing a medium-sized sample of cases. However, most case studies focus on only one or several cases. In a sample, these small problems arise with a purely algorithmic selection procedure. Specifically, sampling variance is high, so any given sample is likely to fall far from the population parameter of theoretical interest (Seawright and Gerring 2008). Moreover, detailed knowledge of a case may be sufficient to call into question the representativeness of a case. If, after close examination, a case seems unrepresentative, it may make more sense to jettison that case—based on qualitative considerations—rather than to persist with a case that is obviously flawed.

Most important, some criteria, almost by definition, are hard to measure, ex ante, across a large sample of cases. Being unmeasured, they cannot be conditioned in an algorithmic case selection method procedure. Foremost among these “unmeasurables” are omnibus criteria (discussed above) such as the independence of a case and the availability of within-case evidence. Note also that for cases that serve a role in causal inference, one must consider the degree to which the chosen case(s) exhibits experimental qualities (Gerring and McDermott 2007). A case with many potential confounders is less useful than a case that is relatively “clean,” for example, where X changes without corresponding changes in Z. While case selection
algorithms often attempt to control for background characteristics, confounders are notoriously devious and therefore not always easy to measure—especially in a large-\(n\) cross-case framework.

Consider that if quasi-experimental cases could be identified by algorithms, we could find good natural experiments simply by running selection algorithms across thousands of potential sites of investigation. In practice, researchers must invest an enormous amount of time looking into the fine details of a site before it can be ascertained whether—or in what respects—it satisfies the methodological criteria of a natural experiment, that is, as-if random assignment and the absence of posttreatment confounders (Dunning 2012). What is true of sites (suitable for large-\(n\) natural experiments) is also true for cases (suitable for small-\(n\) natural experiments): There are no mechanical tools for finding them.

\textbf{A Tentative Resolution}

Algorithmic case selection makes good sense if one is choosing a medium-\(n\) sample of cases, for all the reasons it makes sense in large samples. Here, traditional sampling theory applies. Indeed, it is not clear how one would select several dozen cases in an informal (qualitative) fashion. When a sample is expanded beyond a dozen one is more or less presuming that automatic selection criteria can be applied. The most outspoken proponents of model-based case selection, Fearon and Laitin (2008, 2014, 2015), incorporate 25 country cases in their “random narratives” project, each of which is intensively studied. With a sample of this size, threats from stochastic error are minimized, and sufficient leverage on factors of theoretical interest is assured.

For traditional case study work, where cases number from one to several, case selection algorithms are an excellent, and probably underutilized, tool for situations in which the requirements of algorithmic case selection (discussed above) are met. Even here, however, the algorithmic choice should probably not be followed slavishly in the final selection of cases, for all the reasons discussed above. A case that looks good from the perspective of an algorithm may not look so appealing when one becomes familiar with the intricacies of the setting.

Having considered the pros and cons of algorithmic case selection, we offer a final thought on why this approach to case selection is elusive, and likely to remain so. Let us suppose, optimistically, that some dimensions of our subject—\(D, X, Z,\) and/or \(Y\)—are measurable for a large sample of potential cases. On this basis, one could implement an algorithmic selection of cases, following the protocol laid out above. So far, so good.

It seems likely that if some dimensions of a subject are measurable across a large sample of cases, others might also be measurable. If one can
measure some aspects of \( D \), why not others? If one can measure \( X \) and \( Z \), why not \( Y \)? If one can measure \( Z \) and \( Y \), why not \( X \)? If one can measure \( X \), \( Z \), and \( Y \), why not \( M \)?

If the missing variable can be measured, it seems plausible to test it in a large-\( n \) cross-case model. In this scenario, the main justifications for a case study would be for exploratory purposes (to identify a new hypothesis) or for hypothesis testing—if the small-sample analysis offers greater leverage, perhaps because the large sample is subject to irremediable confounders. Even so, in order to prove that the case study has inferential advantages over the large-sample setting, one would need to examine the former in great detail, as discussed above—in which circumstance knowledge of how the case scores on the missing variable would be virtually impossible to avoid, compromising a purely algorithmic case selection protocol.

Algorithmic case selection requires knowing something about the cases in a population so that an informed choice can be made about which case, or cases, to devote special attention to. The paradox is that the more one knows about the population the greater confidence one has in the algorithmic case selection procedure—but, at the same time, the less informative an intensive study of a particular case is likely to be.

Consider the typical case, where one attempts to choose a case that is descriptively representative of a population. Let us assume that there are a number of factors of theoretical interest and that the researcher’s goal is to describe those features based on the features of the chosen case. Evidently, if one could measure all of those features in the population, there would be no point in conducting an in-depth study of a single case. The employment of a case study design presumes that some things of theoretical interest cannot be measured across the population and therefore remain mysterious. It is these missing elements that the case study is intended to shed light on. The justification of case studies arises from situations, where cross-case evidence is unavailing or uncertain. But the fact that many questions remain should also raise doubts about the algorithm one is using to select a case. The less we know about the population, the more we can learn from a case study. At the same time, our ignorance about the population may make it difficult to employ algorithmic case selection procedures with any degree of assurance.

**Medium-\( n \) Samples**

In light of our previous discussion, the advantages of a somewhat larger sample—medium-\( n \), comprising 10 or more cases—may seem apparent. A medium-\( n \) sample can be chosen algorithmically, achieving all of the
advantages of algorithmic selection without sacrificing leverage on the problem at hand. In light of these advantages, it may behoove case study researchers to follow Fearon and Laitin’s lead. Perhaps the problems of case study research can be mitigated by the simple expedient of increasing the \( n \). If case studies are good, arguably, more cases are better.

If all things were equal, medium-\( n \) samples would surely be preferred to small-\( n \) samples. However, trade-offs are encountered, when a larger number of cases are incorporated into a study. Studying additional cases intensively requires additional time and resources—to access informants or research sites, to learn new languages, to process the information and write up the results, and so on. Sometimes, these tasks can be divided up among a team of researchers; this, of course, requires greater man power than the traditional case study (usually undertaken by an individual researcher, perhaps with the help of a research assistant) and raises coordination problems. Yet, if the only obstacles to medium-\( n \) case study research were logistical, one would expect it to occur more frequently. After all, teamwork is increasingly common in the social sciences, and many projects benefit from large budgets and sizable staffs.

A more fundamental issue arises when one considers what to do with the results of 25 case studies. (We are assuming that the population of interest is much larger—numbering in the hundreds, thousands, or millions.) To integrate results from 25 cases, some sort of data reduction is required, that is, the cases will need to be reduced to numbers. This might be a simple count of the number of cases that seem to show a particular causal mechanism at work, or the number of cases that validate or invalidate the hypothesis of interest, or some new causal factor that is suggested by the cases. It might be a causal model, run on the sample of 25. The point is that once a sample reaches into the dozens, it is no longer possible to analyze the cases in a purely informal manner.

If the end result of a medium-\( n \) case study is a cross-case statistical analysis, we must inquire: Does the medium-\( n \) case study offer any advantages over a traditional large-\( n \) cross-case study? The latter might involve hand-coding cases or collecting data from extant sources, that is, a standardized survey that alleviates problems of cross-case equivalence and vastly enhances the efficiency of the case analysis relative to the laborious task open-ended, case-based research.

An example of standardized coding across cases is found in a recent study by Haggard and Kaufman (2012) who examine over 100 regime transitions in order to determine the role of distributional conflict in these events. The case profiles are housed in a lengthy online document. The published study presents the data derived from this extensive analysis, condensed into tabular
formats. The role of the in-depth qualitative investigation is thus to arrive at a binary coding of each case—as “distributive” or “nondistributive.” It is an ingenious study, and evidence of extraordinary labor on the part of a coordinated research team. However, it is hardly a case study in the sense in which we have defined the term. Indeed, it seems no different from any data collection project in which the authors conduct careful, nose-to-the-grindstone coding and preserve their notes in a codebook.

The point, then, is that a large sample of cases can be integrated through careful case-based knowledge, if data collection is directed by a systematic survey protocol. The process may be centralized (a few researchers do all the coding, as in the Haggard and Kaufman projects) or decentralized (coding is conducted by experts on different subjects or different parts of the world, as in the Varieties of Democracy project; Coppedge et al. 2015). In either context, the resulting study has stronger claims to representativeness and greater protection against stochastic error than the corresponding medium-$n$ analysis. It would, in our view, be superior to medium-$n$ analysis in all respects that medium-$n$ analysis is superior to small-$n$ analysis.

This foreshadows our conclusion. Medium-$n$ case studies are much more expensive (in terms of time and resources) than small-$n$ case studies, and the end result is generally inferior to large-$n$ analysis. As such, medium-$n$ case-based analysis is a hybrid form of research that seems to have few areas in which it enjoys a comparative advantage relative to the small-$n$ and large-$n$ alternatives.

A Plea for Transparency

In this study, we have laid out a typology that elucidates a diverse array of case selection methods (Table 1); we have commented on the viability of algorithmic case selection methods; and we have addressed the option of medium-$n$ samples in case study research. We hope that this contributes to a more self-conscious approach to case selection, one that considers all available options and employs algorithmic formulas where practicable.

By way of closing, we want to stress the importance of transparency. While transparency is widely acknowledged as a vital aspect of research—both qualitative and quantitative (Lupia and Elman 2014)—its relevance for case selection may be underappreciated. Note that in order to judge the nature of the evidence presented in a case study, it is essential that readers know how cases were selected. Without knowledge of what researchers knew, and when they knew it, one cannot distinguish, for example, among the three varieties of most-similar design (selecting on $ZY$, $XZ$, and $XZY$). If algorithms were applied, it is important to know in what order they were
implemented and whether algorithms were the sole method of case selection. If several cases were chosen for intensive analysis, it is important to know whether they were all chosen at the same time and by the same criteria. If the method of selection is different from the method of analysis, this is vital information. Likewise, if cross-case analysis was utilized ("multimethod" research), it is important to know whether this preceded or followed the case study portion of the study or occurred in tandem.

To achieve full transparency, all the procedures followed in a study must be scrupulously laid out in the order they were performed. We imagine that this might take the form of a short appendix, summarizing details contained in a longer research diary or laboratory notebook. In this fashion, processes that usually remain in the shadows would be brought into the open, allowing consumers of case studies to better understand the nature of the data and the probable strength of the findings, and facilitating replication.

This would also facilitate the development of case study research methodology, as readers could learn from prior studies—what worked and what did not. As things stand, we have little grounds to judge whether one method of case selection has been more common or more availing than another, as it is often difficult to ascertain the reasons why a case was initially chosen. Ex post explanations do not always accurately describe a researcher’s initial motives. Until this is clarified and a body of work is established from which we can generalize about the relative utility of different methods, it will be difficult to reach firm conclusions.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Nielsen’s software is currently being developed to allow for matching on $ZY$, $XZ$, or $XZY$—all three varieties of most-similar case selection.
2. An extension of the most-similar method—called “synthetic control” (Abadie, Diamond, and Hainmueller 2015)—relies explicitly on a matching framework.
3. Not everyone views the democratic peace as deterministic, but some do (Brown, Lynn-Jones, and Miller 1996). For a list of additional hypotheses that take a deterministic form, see Goertz and Starr (2003).
References


Author Biographies

John Gerring (PhD, University of California at Berkeley, 1993) is professor of Political Science at Boston University, where he teaches and conducts research on methodology and comparative politics. He is co-editor of Strategies for Social Inquiry, a book series at Cambridge University Press, and serves as co-PI of Varieties of Democracy (V-Dem) and the Global Leadership Project (GLP).

Lee Cojocaru obtained his PhD degree in Political Science from Boston University. Currently, he is a researcher at Boston University and for the Varieties of Democracy (V-Dem) project.