10

Strategies of Causal Assessment in Comparative Historical Analysis

James Mahoney

Scholars who write about comparative historical methods sometimes make it appear that the research tradition has a single basic approach for identifying patterns of causation. Yet, in fact, comparative historical analysts employ a wide range of strategies of causal assessment in their substantive research. These strategies encompass both methodologies for juxtaposing cases with one another and methodologies for analyzing processes that take place within individual cases. And they include both techniques of causal assessment designed to identify the necessary or sufficient causes of an outcome and tools for locating causal factors that covary with outcomes in linear patterns. Rather than narrowly limiting themselves to any one approach, then, comparative historical researchers are eclectic in their use of methods.

In this essay, I attempt to analyze systematically these different strategies of causal analysis. My main objectives are to specify the concrete procedures entailed in the strategies, discuss their underlying assumptions about causality, and assess their comparative strengths and weaknesses. Along the way, I engage the long-standing debate about small-N versus large-N research.[1] I devote particular attention to the ways in which different comparative historical methods are or are not compatible with the assumptions that guide

[1] See Lijphart (1971, 1975), Smelser (1976), Lieberson (1991, 1994, 1998), Collier (1993), Goldthorpe (1997), Goldstone (1997), Rueschemeyer and Stephens (1997), Ragin (1997), Munck (1998).

Table 10.1. *Strategies of Causal Inference in Small-N Analysis*

		Level of Measurement		
		Nominal	Ordinal	Interval
Level of Aggregation	Aggregated	Nominal Strategy	Ordinal Strategy	Not Typically Used
	Disaggregated	Within-Case Analysis		

causal inference in conventional statistical methodologies. My hope is that this discussion will help clear up some of the misunderstandings that have developed between advocates of small-N and large-N research and refocus attention on the real points of contention between the traditions.

The essay identifies three basic strategies of causal analysis that are often combined in particular pieces of comparative historical research: nominal, ordinal, and within-case strategies. Each of these strategies may entail a number of more specific methodologies that differ in important respects. Nevertheless, for the purposes of classification, the strategies can be defined along two dimensions: level of measurement and level of aggregation (see Table 10.1). Both ordinal and nominal strategies make comparisons across highly aggregated units (e.g., nation-states), but the two differ in level of measurement. A nominal strategy makes comparisons using nominal measurement; an ordinal strategy employs ordinal measurement. By contrast, a within-case strategy differs from these two alternatives along the dimension of level of aggregation. Whereas nominal and ordinal strategies involve comparisons across cases, a within-case strategy entails a shift toward disaggregation and a focus on comparisons within particular cases. A within-case strategy does not contrast with ordinal and nominal strategies in terms of level of measurement. In fact, as we shall see, a within-case strategy can be used in conjunction with multiple levels of measurement, including nominal, ordinal, and interval measurement.

Important contrasts in the logic of causal inference underlie nominal and ordinal strategies. A nominal strategy implicitly or explicitly assumes a nonlinear understanding of causation built around the ideas of necessary and sufficient conditions. This understanding of causation is quite different from that employed by most large-N researchers, who often analyze linear causation and who rarely test for necessary and sufficient conditions. By contrast, an ordinal strategy is more compatible with the linear and

correlational assumptions that guide much large-N research. At the same time, however, some comparative historical researchers argue that ordinal analysis (and perhaps even most statistical research) is itself best understood in terms of necessary and sufficient conditions. Finally, different variants of a within-case strategy can be used in conjunction with different understandings of causation. In some instances, within-case analysis may replicate the procedures underlying nominal or ordinal strategies; in other cases, this kind of analysis may involve the statistical evaluation of a large number of observations.

Nominal Comparison and the Question of Determinism

Nominal (or categorical) comparison entails the use of categories that are mutually exclusive (cases cannot be classified in terms of more than one category) and collectively exhaustive (one of the categories applies to each case).[2] Examples of these categories in comparative historical research include various regime classifications (e.g., democratic, authoritarian, and totalitarian regimes), typologies of different states (e.g., conservative, liberal, and social-democratic welfare states), and countless dichotomous variables (e.g., revolution versus nonrevolution). Because vivid labels can be attached to nominal categories, comparative historical analysts often summarize their arguments with this kind of comparison. In addition, as a first step in research, they may employ nominal comparison to measure independent and dependent variables and to develop an initial sense of whether potential explanations and outcomes are roughly matched.

My concern in this section, however, is with the use of nominal comparison to evaluate systematically the validity of alternative explanations. In the discussion that follows, I consider how both deterministic and probabilistic nominal methods are used for this task.

What Is a Deterministic Explanation?

Although comparative historical analysts are often criticized for presenting "deterministic" explanations, the concept of determinism itself is often not clearly defined. For some scholars, determinism entails focusing on

---

[2] With respect to different levels of measurement, nominal categorization is sometimes considered unsophisticated because it does not involve rank ordering cases, much less quantifying the degree to which particular cases differ from one another. Yet, for conceptualizing certain kinds of phenomena, nominal categories are highly appropriate (Sartori 1987, pp. 182–5; see also Collier and Adcock 1999).

structural variables at the expense of agency (see Giddens 1979). For example, analysts who view Theda Skocpol's *States and Social Revolutions* (1979) as suffering from determinism often are really concerned with its emphasis on structure relative to agency (e.g., Colburn 1994; Selbin 1993; see also Skocpol 1994). In this essay, I am not concerned with this kind of determinism, although I do believe many comparative historical researchers effectively combine agency and structure in their studies (Mahoney and Snyder 1999).

For other scholars, determinism represents a particular understanding of the nature of the world that can be contrasted with probabilism. In this view, a fully deterministic approach assumes that all occurrences in the world are the product of nonrandom prior occurrences; given the right variables, methods, and measurement devices, the world is completely predictable. By contrast, a fully probabilistic approach assumes that all occurrences in the world are at least in part the product of inherently stochastic processes that cannot even in principle be identified. As a result, full prediction is impossible under all circumstances. These two approaches pose a sharp dichotomy, and scholars who employ deterministic and probabilistic explanations in social science research need not fully embrace one side or the other. For example, scholars who adopt deterministic explanations may assume that much of the world conforms to the probabilistic understanding and that only selected cases can be modeled deterministically. Likewise, scholars who employ probabilistic explanations may assume that the world really is deterministic, but that our inability to identify and correctly measure all relevant variables forces us to act as if it were probabilistic (see Goldthorpe 2000, pp. 137–8; King, Keohane, and Verba 1994, pp. 59–60).

In actual research, a deterministic explanation assumes the existence of causes that exhibit, in at least certain respects, invariant relationships with outcomes within a specified domain of cases. For example, in most statistical research, a deterministic explanation assumes that values on an explanatory variable fully predict values on an outcome variable without error within a specified population. As a result, at least for every case within that population, one can predict the exact causal effect on an outcome variable of a given change on an explanatory variable. The inability of statistical researchers to formulate successful deterministic models has meant that nearly all research in this field adopts probabilistic explanations that assume such prediction is impossible.

In comparative historical studies, a deterministic explanation usually involves modeling conditions (or combinations of conditions) that are

necessary or sufficient for the occurrence of an outcome within a specified population of cases. This kind of analysis also makes certain predictive assumptions. For example, with a *necessary* cause, the absence of this cause is always associated with the absence of the outcome in question, at least within the relevant population of cases. However, if a necessary cause is present, the outcome could be either present or absent. With a *sufficient* cause, the presence of this cause is always associated with the occurrence of the outcome in question (again, within the relevant population). Only when a given factor is a necessary *and* sufficient cause will the outcome *always* be present when the cause is present and *always* absent when the cause is absent.

Comparative historical researchers often implicitly understand a probabilistic explanation as one in which explanatory variables (or combinations of variables) are treated as to some specified degree necessary or sufficient for the occurrence of an outcome. For example, a probabilistic explanation may treat explanatory variables (or combinations of variables) as *almost always* necessary or *almost always* sufficient conditions. With this kind of explanation, the analyst may assume that an inherently stochastic process explains why a condition is not always necessary or always sufficient for an outcome of interest. Alternatively, the analyst may assume that the world really is fully deterministic and that measurement error accounts for the probabilistic relationship. In either case, however, the researcher is not committed to a deterministic explanation.

Deterministic Methods

Some nominal techniques are used deterministically by comparative historical researchers, including, most notably, J. S. Mill's method of agreement[3] and method of difference.[4] Although these Millian methods do not permit the analysis of multiple explanatory factors or interaction effects, they provide a sound logical basis for *eliminating* potential necessary and

[3] With the method of agreement, the analyst attempts to establish that cases that share a common outcome also share common hypothesized causal factors, despite varying in other significant ways (see Skocpol and Somers 1980).

[4] The analyst using the method of difference contrasts cases in which an outcome under investigation and hypothesized causal factors are present to other cases in which both the outcome and the hypothesized causes are absent, even though the cases may be similar in many other respects (see Skocpol and Somers 1980).

◦

sufficient causes.[5] Specifically, the method of agreement can be used to eliminate potential *necessary* causes, whereas the method of difference can be used to eliminate potential *sufficient* causes. Thus, with the method of agreement, the outcome of interest is present in all cases. Consequently, it is logically impossible for any hypothesized cause that is not shared by the cases to be individually necessary for the outcome's occurrence, since some cases possess the outcome but not the cause. By contrast, with the method of difference, the outcome is present in some cases and not present in others. Hence, any hypothesized cause that is shared by all the cases cannot by itself be sufficient for the outcome, since not all cases with the hypothesized cause experience the outcome of interest.

With these deterministic methods, a single deviation from a hypothesized pattern of necessary or sufficient causation is enough to eliminate a given factor as a potential cause. As a result, the methods provide a powerful basis for systematically eliminating rival causal hypotheses, even when only a small number of cases are selected. For instance, in her explanation of contrasts in the national electoral systems of three post-Soviet Central Asian countries, Pauline Jones Luong (2002) uses the method of difference to eliminate plausible causal factors that do not vary across the countries. She shows that, given their similar levels of socioeconomic development, one cannot argue that level of development was by itself sufficient for an inclusive and populist electoral system in Kyrgyzstan, a restrictive and centralist electoral system in Uzbekistan, and a hybrid electoral system in Kazakhstan. Likewise, in *Making Race and Nation: A Comparison of the United States, South Africa, and Brazil*, Anthony W. Marx (1998) uses the elimination procedures of the method of agreement to assess explanations of the major similarities in racial domination in the United States and South Africa. He eliminates factors that do not vary systematically between the countries, such as the presence of an African or African-descendant majority population, because they are not by themselves necessary conditions in accounting for the major

---

[5] It bears emphasis that causal factors eliminated using these methods might, in conjunction with other factors, represent part of a larger combination that is itself necessary or sufficient for an outcome of interest. The methods of agreement and difference do not provide a basis for analyzing such combinations of variables, except insofar as the combinations are treated as a single factor. For example, Skocpol (1979) uses the methods in this fashion (see Mahoney 1999, p. 1158). It is also worth noting that the "most similar systems design" and the "most different systems design" have the same logical structure as the method of difference and the method of agreement, respectively, except that Przeworski and Teune's (1970) original formulation of these designs was based on the premise that the scholar is combining two levels of analysis.

similarity of interest. At the same time, Marx uses the method of difference to eliminate possible explanations of the contrast between repressive racial domination in these countries, on the one hand, and significant racial tolerance in Brazil, on the other. For instance, the presence of early labor coercion and colonial discrimination are eliminated because these factors do not vary across the three countries.

Against the backdrop of having eliminated initially plausible explanations using the methods of agreement and difference, comparative historical researchers frequently hold up their favored argument as the one that can survive the deterministic tests. When this takes place, the analytic logic of comparative historical analysis can be strikingly convincing. For example, Loung's explanation, which highlights the perceptions of central and regional leaders in Central Asia concerning their relative power vis-à-vis one another, seems remarkably persuasive in light of the failure of many other plausible explanations. Also compelling is Marx's explanation of major similarities between the United States and South Africa, which emphasizes the existence of significant divisions among whites following the Boer War and the Civil War. Ultimately, in these two countries, white unity and nationalist loyalty to the state were forged through the construction of systems of racial domination that systematically excluded blacks. By contrast, in Brazil, the absence of a similar intrawhite cleavage obviated the need for similar policies of racial domination to ensure white unification.

Boolean algebra is distinct from other nominal methods because it allows the analyst to treat several different *combinations* of variables as the causes of an outcome (Ragin 1987). In particular, this methodology provides a logical basis for identifying combinations of causal factors that are *sufficient* for the occurrence of an outcome.[6] Analysts who use this method may identify several different combinations of factors, each of which is sufficient for an outcome. In this way, Boolean procedures enable researchers to recognize that multiple causal combinations may produce the same outcome, what Ragin (1987) calls "multiple conjunctural causation."

A nice illustration of Boolean analysis is Wickham-Crowley's (1992) work on the origins of peasant support for guerrilla movements in Latin America. Wickham-Crowley focuses on twenty cases, and he looks at

---

[6] The Boolean minimization procedure through which causal combinations are reduced eliminates potential *necessary* causes among cases that share the same outcome. Final combinations of causal factors are understood to be *sufficient* for the occurrence of the outcome. In addition, by looking at final combinations of explanatory variables, it is possible to state whether any single factor is a necessary or sufficient cause.

four different explanatory variables: agrarian structure (A), agrarian disruption (B), rebellious cultures (C), and peasant linkages (D). He first pools cases in which the outcome of peasant support for guerrillas is present and identifies all combinations of scores on explanatory variables that are associated with this outcome. Next, he reduces the number of these combinations by assuming that if two combinations of explanatory variables differ in their scores for a single variable, then that variable can be eliminated from the combination. The implicit rationale behind this Boolean reduction procedure is that the variable is not *necessary* for the combination to have a causal effect since it is both present and absent in combinations associated with the outcome. For example, all four of Wickham-Crowley's explanatory variables are present in one combination with peasant support for guerrillas (expressed as ABCD), while the explanatory variables A, B, and C, but not variable D, are present in another combination with peasant support for guerrillas (expressed as ABCd). Hence, Wickham-Crowley assumes that variable D is irrelevant to the combination and eliminates it, reducing the causal expression to ABC.

Through this Boolean reduction procedure, Wickham-Crowley narrows the range of explanations down to four possible combinations of variables under which peasants have supported guerrillas: ABD, AC, CD, and abD. Although Wickham-Crowley does not identify any *individual* causal factor that is necessary or sufficient to produce strong support for guerrillas, the final four expressions are each understood to represent a *combination of factors* that are *sufficient* for peasant support. For example, the combination of a rebellious culture (C) and peasant linkages (D) is understood to be sufficient for strong support for guerrillas in modern Latin America.

Probabilistic Methods

Other methods that rely on nominal comparison are probabilistic in the sense that they relax standards to permit causes that are "usually" or "almost always" necessary or sufficient. Comparative historical researchers who draw on these methods readily acknowledge that inherent randomness or measurement error may make it difficult to locate patterns of association that are fully necessary or fully sufficient for an outcome. At the same time, however, these methods make it somewhat more difficult to eliminate rival explanations. For example, a single deviation often does not provide a basis for eliminating an explanation when causation is assumed to be probabilistic. As a rule, comparative historical researchers need to select more than just a

few cases to use probabilistic nominal methods as a basis for systematically eliminating rival explanations.

A nice example of how a probabilistic nominal approach can be effectively used in practice is Thomas Ertman's (1997) magisterial work *Birth of the Leviathan: Building States and Regimes in Medieval and Early Modern Europe.* Ertman's goal is to explain why particular European countries developed one of four major types of early modern regime-states, each defined in terms of constitutionalism versus absolutism and patrimonialism versus bureaucracy. Ertman first systematically eliminates rival explanations derived from the theories of state development formulated by Otto Hintze, Charles Tilly, Perry Anderson, and Michael Mann. He does so by showing how, within a population of fourteen cases, these theories lead to predictions about regime-state outcomes that are contradicted by at least four cases. Based on the failings of these theories, Ertman offers his own explanation, which emphasizes three dichotomously measured variables – administrative versus participatory government, pre-1450 versus post-1450 geopolitical competition, and existence versus nonexistence of powerful representative associations. Together these variables are understood to represent an explanation that "can account for most of the variation" in early modern state building in Europe (Ertman 1997, p. 6). In particular, the explanation is consistent with twelve of the fourteen cases, failing to explain state outcomes only in Sweden and Denmark, where "powerful contingent events conspired to confound expected paths of development" (p. 33). Because of these two exception cases, Ertman cannot treat his explanatory variables as *always* sufficient for the occurrence of outcomes. Instead, he must adopt a probabilistic approach in which his explanation is *usually* sufficient.

The decision of analysts to employ a probabilistic approach raises important issues regarding the benchmark for determining degrees of sufficiency or necessity. In his most recent methodological work, Ragin (2000) offers precise terminology and benchmark proportions for discussing causes with different degrees of necessity and sufficiency. For instance, a cause that is usually necessary or usually sufficient must meet a .65 benchmark (i.e., the cause must be necessary or sufficient at least 65 percent of the time), while a cause that is almost always necessary or sufficient must meet a .80 benchmark. These benchmarks can be applied to all of the nominal techniques discussed previously, including Boolean approaches that focus on combinations of factors.

In the case of Ertman, we saw how he found that twelve out of fourteen cases are consistent with his explanation, which yields a percentage of

.86 and is thus higher than the benchmark for an almost always sufficient cause. However, given the relatively small number of cases, he can achieve statistical significance at a .10 level only by treating his explanation as usually sufficient for the regime-state outcomes.[7] By contrast, the explanations eliminated by Ertman can explain at best ten out of the fourteen cases (.71), thus falling below the benchmark of almost always sufficient. Moreover, these alternatives cannot meet any reasonable level of statistical significance even if they are evaluated at the lower benchmark of usually sufficient. In short, Ertman has a logical and statistical basis for rejecting alternative explanations by scholars such as Hintze, Tilly, Anderson, and Mann while preserving his own explanation as identifying a set of factors that are usually sufficient for producing regime-state outcomes in early modern Europe.

A major new probabilistic technique is found in Ragin's (2000) discussion of fuzzy-set methods.[8] Fuzzy-set measurement defies easy classification along standard scales of measurement (i.e., nominal, ordinal, interval, and ratio measurement), though it parallels the nominal concern with set membership. To use fuzzy sets, the analyst must score cases from 0 to 1 based on their degree of membership in a category. Some cases will be "full" instances of a category and thus receive a score of 1, even if these cases have different scores on an interval scale. For example, with respect to the category of "wealthy country," both the United States and England are full members, and thus both receive the same score of 1, despite the fact that these countries have different levels of wealth on an interval scale.[9] Other countries receive scores based on the extent to which they overlap with the category "wealthy country." For example, a country that is "mostly in" the category will receive a score only slightly less than 1 (e.g., .83), while a country that is "mostly out of" the category will receive a score closer to 0 (e.g., .17).

---

[7] I have applied the significance test identified by Ragin (2000, p. 112) with some liberty here, given that Ertman actually offers an explanation for four different outcomes. In this sense, he does not have a total of fourteen cases for each of his outcomes. However, since he uses the same set of explanatory variables to account for each of these outcomes, there is a logical basis for using the binomial test as if Ertman's variables successfully explained twelve of fourteen cases on the same outcome. The conclusion I reach here addresses a concern with Ertman's work I raised but did not resolve in an earlier article (see Mahoney 1999, p. 1175).

[8] I discuss fuzzy-set methods in greater depth in Mahoney (2001).

[9] Although the decision to ignore variation at the extreme ends of continuous variables is a source of bias in conventional statistical research, this problem does not arise in the analysis of necessary and sufficient conditions. In fact, error might arise if the additional variation were included.

The decision to assign fuzzy-set membership scores must be grounded in the analyst's substantive knowledge of actual cases.

Fuzzy-set measurement is highly appropriate for the analysis of necessary and sufficient conditions, including under probabilistic assumptions where different degrees of necessary or sufficient causation are considered. To employ the technique, the analyst must apply the fuzzy-set measures just described to all potential causal factors and outcomes and then assess the relationship between their values. With a *necessary* cause, fuzzy membership scores on the *outcome* will be less than or equal to fuzzy membership scores on the *cause*. By contrast, with a *sufficient* cause, fuzzy membership scores on the *cause* will be less than or equal to fuzzy membership scores on the *outcome*. To incorporate considerations of probabilistic causation, the researcher might argue that if no case's score on the outcome (or cause) exceeds its score on the cause (or outcome) by more than a small portion of a fuzzy membership unit, then the pattern is still consistent with the interpretation of causal necessity (or sufficiency). Likewise, the probabilistic benchmarks and significance tests mentioned earlier can be applied when using fuzzy measures of variables. Although the procedures involved become especially complicated when *combinations* of variables are considered using probabilistic criteria, a free software package that performs the operations is already available (Drass and Ragin 1999).

Although Ragin offers interesting examples of how these methods might be applied to substantive questions such as the onset of International Monetary Fund protest and the establishment of generous welfare states, the techniques are still too new (at the time of this writing) to have been used by other researchers. Some comparative historical researchers may resist the formal constraints of this methodology, preferring instead the flexibility of alternative strategies of causal inference. Nevertheless, several comparative historical researchers already have sought to apply rudimentary versions of fuzzy-set logic through analyses that combine nominal and ordinal comparison (discussed later). These works suggest that at least some comparative historical researchers will welcome the formal application of Ragin's latest innovation.

Evaluating Criticisms of Methods That Use Nominal Comparison

The sharpest criticisms of comparative historical methods have focused specifically on nominal techniques. Most of these concerns have been raised by scholars working in the tradition of statistical analysis who argue that

comparative historical researchers violate standard rules for conducting valid research. In this section, I evaluate the five most important of these criticisms, suggesting that they are often unfounded concerns based on the misapplication of conventional statistical reasoning to the study of necessary and sufficient causation.

Criticism 1: Necessary and Sufficient Conditions Are Not a Useful Way to Think About Causation   Even if logical methods exist for identifying necessary or sufficient causation, some analysts contend that it is still not a productive way to think about causation. A common concern is that many necessary or sufficient causes are not analytically helpful. For example, regarding specifically the *necessary* causes of an outcome, there are potentially an infinite number of such causes for any outcome, most of which are unimportant or trivial (e.g., the existence of human beings is a necessary cause of a social revolution). Likewise, many *sufficient* causes of an outcome are obvious or tautological (e.g., war is a sufficient cause of large-scale death).

However, examples of trivial necessary conditions and tautological sufficient conditions found in the literature are hypothetical ones offered by critics who seek to dismiss the study of these conditions (see Braumoeller and Goertz 2000). The literature offers few or no examples of real researchers who use nominal methods to analyze trivial necessary conditions or tautological sufficient conditions. In fact, critics of the study of necessary and sufficient conditions seem unaware that *empirical* criteria – as opposed to normative or political criteria – exist for differentiating trivial necessary causes from nontrivial ones and tautological sufficient causes from nontautological ones. *Trivial necessary causes* are those in which the cause is present in all cases, irrespective of the value of the dependent variable (Braumoeller and Goertz 2000; Dion 1998). For example, the existence of human beings is trivially necessary for social revolutions, because this condition is present in all cases of revolution and nonrevolution alike. Braumoeller and Goertz (2000) show through an evaluation of more than a dozen published studies that actual social researchers who make claims about necessary conditions do not refer to trivial ones. Dion (1998) reaches the same conclusion in his evaluation of the use of necessary conditions in comparative politics.

With *tautological sufficient conditions*, the analyst identifies a set of factors that are contained within the very definition of the outcome being considered. When this happens, there is no temporal separation between the cause and outcome (or the outcome may actually occur *before* the cause). For instance, the claim that war is a sufficient cause of death is tautological

because there is no definitional distinction or temporal separation between the occurrence of war and death (or the outcome of death may actually occur before full-scale war breaks out). Again, comparative historical researchers who make claims about sufficient causes are not referring to these kinds of tautological conditions. For example, it is not tautological to assert that a high level of domestic financing for military modernization in the sixteenth and seventeenth centuries was sufficient for an autocratic regime outcome in early modern Europe (Downing 1992), or that large group size and growth are together sufficient for ethnic political mobilization (Ragin 2000, p. 138), or that the combination of state structures conducive to breakdown in the face of international pressures and peasant agrarian structures that facilitate revolt are sufficient for social revolutions in agrarian bureaucratic societies (Skocpol 1979).

A somewhat different concern involves the fact that many causal factors are not necessary or sufficient conditions, but rather follow a linear pattern of causation like that assumed in many probabilistic regression models. As a consequence, the analyst using nominal methods might erroneously dismiss these kinds of linear causal factors as unimportant. Lieberson's (1991) well-known example of drunk driving and automobile accidents illustrates this criticism for deterministic versions of the methods of agreement and difference. Thus, with the method of agreement, an analyst who examines three cases of automobile accidents will eliminate drunk driving as a cause if it is present in only two of the three cases. Similarly, with the method of difference, the analyst will eliminate drunk driving as a cause of automobile accidents if it is present in cases of both accidents and nonaccidents. It is essential to recognize that this example does not call into question the ability of nominal methods to evaluate necessary and sufficient causation. These methods correctly show that drunk driving by itself is neither a necessary nor a sufficient condition for an automobile accident (i.e., some automobile accidents occur in the absence of drunk driving, and not all instances of drunk driving produce automobile accidents). Rather, the example suggests the problems that can arise if one thinks about causation in terms of necessary and sufficient conditions when more conventional linear causation likely is at work.

In response to Lieberson, one might point out that actual researchers who sought to use nominal methods to analyze the effect of drunk driving on automobile accidents would doubtless explore whether drunk driving *in combination with* other variables is a sufficient (or usually sufficient) cause of auto accidents *in a specific population of cases*. Moreover, they would arrive

at final conclusions using other techniques of causal assessment, including within-case evaluation. Nevertheless, it is worth noting that nominal methods generally will have difficulty assessing the net effect of any single factor of special interest – such as drunk driving on automobile accidents or cigarette smoking on lung cancer – that follows a roughly linear pattern of causation similar to that assumed in conventional statistical research. By the same token, of course, linear regression models will have difficulty identifying necessary and sufficient conditions, whether individually or in combination. As a general rule, analysts must rely on theory and common sense to decide whether a given factor should be tested as a necessary or sufficient condition or as a linear cause that increases or decreases values on an outcome.

Criticism 2: Achieving Statistical Significance Is Impossible with These Methods　This criticism grows out of the belief that comparative historical studies suffer from a "degrees of freedom" problem equivalent to that which arises in quantitative analysis.[10] In fact, however, the criticism is an example of an inappropriate application of conventional statistical assumptions to studies designed to analyze necessary and sufficient conditions. It is true that comparative historical studies that examine a very small number of cases (e.g., fewer than five) cannot achieve statistical significance when nominal methods are used in isolation. However, studies with a very small N do not necessarily use only nominal methods, but instead may rely extensively on other strategies of causal assessment. Moreover, in contrast to conventional quantitative research, a relatively small number of cases is often enough to achieve statistical confidence when assessing necessary and sufficient conditions. Using Bayesian assumptions, for example, Dion (1998) shows that only five cases will be enough to yield 95 percent confidence about necessary causes. Using a simple binomial probability test, Ragin (2000, pp. 113–15) shows that if one works with usually necessary or usually sufficient causes, seven consistent cases are enough to meet this level of significance. Braumoeller and Goertz (2000) offer many examples of case-oriented studies that pass such significance tests. In short, scholars need to rethink how appropriate the idea of a degrees of freedom problem

[10] As Rueschemeyer (this volume) notes, the degrees of freedom problem is probably the most common criticism of small-N analysis. For different perspectives, see Campbell (1975), Lijphart (1971, 1975), Nichols (1986), Collier (1993), Lieberson (1991), King et al. (1994), and Goldthorpe (1997).

really is for comparative historical analysts who adopt an understanding of causality built around necessary and sufficient causes. When this understanding of causality is employed, a relatively small number of cases will often meet the confidence demands of standard statistical analyses.

Criticism 3: Because These Methods Select on the Dependent Variable, They Suffer from Selection Bias Drawing on insights from ordinary least squares regression, Geddes (1990) and King et al. (1994) have criticized studies that select cases based on their score on the dependent variable because this selection procedure can lead to truncation on the dependent variable and thus biased estimates of causal effects (see also Collier and Mahoney 1996). In the worst case scenario, analysts may select cases with no variation on the dependent variable, which, according to King et al. (1994, p. 129), makes it impossible to learn anything about the causes of that variable.

Notwithstanding the rules for valid research in ordinary least squares (OLS) regression analysis, selection on the dependent variable – including the selection of cases that do not vary at all on the dependent variable – is *not* a source of bias when nominal methods are used to identify necessary conditions. Selection on the dependent variable is *highly appropriate* for the study of necessary causation (see Braumoeller and Goertz 2000; Dion 1998; Most and Starr 1989; Ragin 2000). In fact, when analyzing necessary conditions, a standard design is to intentionally permit no variation on the dependent variable. The statistical concern about selection bias undermining valid research simply cannot be extended to comparative historical studies that use nominal techniques.

One might raise the more general concern that the particular sample of cases included for analysis in comparative historical work is not representative of the larger population of cases, and that therefore findings about necessary and sufficient causes in the sample analyzed are not consistent with the findings that would have emerged if a more representative sample was selected. However, comparative historical analysts are in fact among the most self-conscious researchers in terms of defining populations of cases that can be understood as homogeneous (Ragin 2000). Whereas cross-national statistical researchers commonly risk violating assumptions of causal homogeneity by arbitrarily selecting large samples of nations, comparative historical analysts assess necessary and sufficient conditions in carefully formulated populations of cases where these assumptions are more sustainable. Furthermore, to the degree that a nonarbitrary definition of

351

the population cannot be established in this kind of research, the whole idea of bias deriving from an unrepresentative sample becomes problematic (Collier and Mahoney 1996).

Criticism 4: The Deterministic Versions of These Methods Do Not Allow for Measurement Error   As suggested earlier, measurement error does not pose special problems for probabilistic nominal methods that relax assumptions about necessary and sufficient causation. However, the fully deterministic versions of these methods assume that the analyst is able to measure all variables correctly, since a change in the scoring of a variable for a single case could lead to opposite conclusions about that variable's causal relevance. In this sense, deterministic nominal methods do not permit any notion of measurement error (see Lieberson 1991).

It is not clear how important a criticism is raised by this objection, however. Comparative historical researchers are typically experts on each of their cases, and given that with this strategy they conceptualize variables as nominal – not continuous – categories, they may indeed be able to avoid measurement error for all of their variables. For example, Skocpol (1979) may well have correctly scored all major dichotomous variables for all of her cases of social revolution and nonsocial revolution. Moreover, if comparative historical researchers like Skocpol do score a particular variable incorrectly for even one case, it is likely that other case experts or comparative historical analysts will identify this error, since much debate in this kind of research entails arguments about the scoring of particular variables for specific cases.

Criticism 5: The Results Generated by These Methods Might Change Substantially if Omitted Explanatory Variables Were Included in the Model   This objection holds that variables not included in a nominal assessment might have avoided elimination if they had been included. Likewise, with techniques such as Boolean algebra and fuzzy-set analysis, variables that were eliminated might have been causally important if other relevant variables were included in the model. This criticism is correct, and it has been discussed by comparative historical methodologists at some length (e.g., Amenta and Poulsen 1994; Ragin 1987).

Yet, it is essential to recognize that the problem addressed here is one of correctly specifying a causal model, an issue that arises with *all* methods of causal assessment, including the most sophisticated statistical analyses. For example, in quantitative research, omitting an important variable can

entail misunderstanding the causal effects of the other variables that are included in the analysis (Lieberson 1985). Thus, like all other researchers using all other methodologies of causal assessment, comparative historical investigators who employ nominal methods can never know for certain whether they have correctly specified their models and thus proven the existence of causation.

To conclude this section, I would like to underline three points. First, it is sometimes useful to think about causation in terms of (probabilistic or deterministic) necessary and sufficient conditions. Second, comparative historical researchers have nominal methods that contain logical rules for locating necessary and sufficient conditions. Third, existing criticisms of comparative historical methods fail to call into question their capacity to identify successfully necessary and sufficient conditions.

Ordinal Comparison and the Question of Linear Causation

A second major strategy of causal assessment in comparative historical research – ordinal comparison across cases – relies on techniques more familiar to scholars working in the tradition of large-N statistical research. Ordinal analysis is in fact the strategy of inference that comparative historical researchers turn to when they seek to identify linear correlations across a small number of cases. Although the technique has important limitations when employed in isolation and with only a small number of cases, it can be combined in creative ways with other strategies of causal assessment.

Identifying and Interpreting Ordinal Associations

Ordinal analysis involves rank ordering cases using variables with three or more values based on the degree to which a phenomenon is present. This kind of analysis facilitates the use of J. S. Mill's method of concomitant variation, in which the analyst tries to establish causation by looking at the relationship between scores on an ordinally measured explanatory variable and scores on an ordinally measured outcome variable (DeFelice 1986; Mill 1974). For example, if values on an explanatory variable and an outcome variable are measured as high, medium, and low, then cases are compared to see if there is an association (possibly an inverse association) between the two variables. If there appears to be a strong association, the analyst may infer that the relationship is causal.

What, however, constitutes a strong association for the comparative historical researcher who uses the method of concomitant variation? Some scholars committed to a deterministic understanding of causation may argue that anything less than a perfect match between values on the explanatory variable and on the outcome variable will fail to indicate causation. However, Mill himself (1974, pp. 402–6) argued that, when using the method of concomitant variation, a perfect match between cause and outcome is not required to infer causality. Rather, associations in which the values on explanatory and outcome variables do not always match may still indicate causation. As a general rule, comparative historical researchers who follow Mill's standard should not infer causation unless they can establish statistical confidence that the relationship is not simply a product of pure chance. As we will see, this can be achieved by having a moderate number of cases and by systematically combining ordinal analysis with other strategies of causal assessment.

The discovery that an explanatory variable is related to an outcome variable in an ordinal assessment does not indicate how one should interpret the nature of the association. One possibility is to assume a linear pattern of causation like that analyzed in much conventional statistical research. Even though ordinal comparison does not strictly speaking permit the assessment of linear causation, it is not uncommon for qualitative and quantitative researchers to interpret ordinal results in this fashion. In doing so, the goal of analysis becomes estimating the proportion of an outcome that can be attributed to a particular value on an explanatory variable (see King et al. 1994, pp. 76–82).

Comparative historical researchers who use ordinal comparison to assess linear causation and estimate the causal effects of variables in this fashion face obvious disadvantages when compared to researchers who have a large number of cases at their disposal. When an ordinal strategy is used in comparative historical analysis, one can meaningfully speak of a degrees of freedom problem that inhibits the use of all but the most simple bivariate statistical methods. Furthermore, insights from statistical research about case selection, measurement error, and the construction of causal theories are often highly appropriate when applied to comparative historical works that seek to use the method of concomitant variation to identify patterns of linear causation. Yet, it must be emphasized that comparative historical researchers usually do not rely on ordinal analysis in isolation, instead combining it with nominal and within-case analysis. In this sense, one cannot mechanically extend advice from large-N analysis to

even those comparative historical studies that do rely extensively on ordinal analysis.[11]

Special problems can arise when comparative historical researchers attempt to combine an ordinal strategy with a nominal strategy in comparative historical research. In particular, the practice of first assessing a variable as a necessary or sufficient condition and then reassessing this variable at the same level of aggregation using ordinal comparison raises concerns. The problem is that necessary or sufficient causes do not typically follow a linear pattern when plotted against an outcome. Yet, when researchers use ordinal analysis to assess hypotheses, they do generally test to see whether variable scores can be matched in a pattern that resembles linear change. One would therefore not necessarily expect a necessary or sufficient condition to appear causally important when evaluated through ordinal comparison.

In light of this problem, some comparative historical researchers may reject the very idea that ordinal findings should be interpreted as modeling linear causation or representing a correlational pattern. Instead, they may argue that ordinal associations should be viewed as reflecting necessary *and* sufficient causation. For example, an explanatory factor that reveals a very strong association with an outcome variable in an ordinal analysis could be meaningfully interpreted as a (probabilistic) necessary *and* sufficient condition. In other words, the bivariate scatter plot for a variable that is usually necessary *and* sufficient typically will appear as a very strong correlation in an ordinal analysis.

This discussion raises two general points of relevance to both statistical researchers and comparative historical researchers. First, conventional methods used by statistical researchers could potentially be recast as tools that seek to identify necessary and sufficient conditions. For example, additive linear models might be seen as tools designed to locate causal factors that are probabilistically sufficient but not necessary for an outcome.[12] These models compute the effects of each explanatory variable net of all other variables under the assumption that an increase on an explanatory variable leads (probabilistically) to an increase (or decrease) on an outcome variable all by itself. This notion that individual causes are capable of producing an effect by themselves within some probability is consistent with the

---

[11] In addition, some comparative historical researchers who use ordinal analysis extensively to study linear causation may select a rather large number of cases similar in size to that of much cross-national research (e.g., Rueschemeyer, Stephens, and Stephens 1992).

[12] Thanks to Charles Ragin for pointing this out to me.

understanding of a probabilistic sufficient cause (but not a probabilistic necessary cause). An additive linear equation that fully explained the variation on the outcome could be seen as having identified all of the individual variables that are sufficient for values on the dependent variable. A key priority for methodologists should be to explore the extent to which various statistical methods can be reframed in terms of necessary or sufficient conditions.

Second, comparative historical researchers who combine nominal and ordinal analysis should be explicit if they believe the explanatory variable under consideration operates as a condition that is necessary and sufficient, a condition that is necessary but not sufficient, a condition that is sufficient but not necessary, or a linear cause that increases or decreases values on an outcome. In general, logical problems will arise if the variable is treated as either a necessary but not sufficient condition or a sufficient but not necessary condition in a nominal evaluation and then is reassessed as either a necessary and sufficient condition or as a probabilistic linear cause in the ordinal analysis. The exception to this rule is scholars who disaggregate the variable when shifting from nominal to ordinal analysis (e.g., Skocpol 1979). When this takes place, the analyst can meaningfully assume that the disaggregated variables used in the ordinal assessment follow a linear pattern or a pattern of necessary and sufficient causation, whereas the aggregated nominal variable follows a pattern of necessary but not sufficient or sufficient but not necessary causation.

Examples of Ordinal Analysis in Comparative Historical Research

In some cases, comparative historical analysts may use ordinal analysis to strengthen an argument that is developed primarily through nominal comparison. In other cases, they may employ ordinal comparison as the central strategy of investigation. It is instructive to consider examples of both uses here.

Luebbert's (1987, 1991) work on interwar regimes in Europe employs ordinal analysis to strengthen a primarily nominal argument. In his overarching nominal assessment, Luebbert argues that the presence or absence of "lib-labism" (i.e., a liberal party–labor alliance) before World War I explains liberal versus nonliberal regime outcomes during the interwar period (see Table 10.2). Thus, when lib-labism was present before World War I, as in England, Switzerland, and France, a liberal regime developed in the interwar period. By contrast, when lib-labism was absent, as in Norway, Sweden, Denmark, Germany, Italy, and Spain, a nonliberal regime

Table 10.2. *Nominal Comparison in Luebbert's Analysis of
Interwar Regimes*

	Prewar Lib-Labism	Interwar Liberalism
Switzerland	Yes	Yes
England	Yes	Yes
France	Yes	Yes
Belgium	No	No
Netherlands	No	No
Denmark	No	No
Italy	No	No
Norway	No	No
Spain	No	No
Sweden	No	No
Germany	No	No

developed. Hence, the dichotomous explanatory variable of lib-labism is
perfectly matched with interwar liberalism.

Although his main explanation relies on nominal comparison, Luebbert
reconceptualizes lib-labism and interwar liberalism as variables that are
ranked across cases. Table 10.3 presents estimates of this ranking for
the main countries considered by Luebbert. If the two ordinal variables are
compared across cases, it is apparent that there is substantial, but not per-
fect, matching. Among the eleven cases, six of them (France, Belgium, the
Netherlands, Denmark, Spain, and Germany) maintain their rank order;
three cases (Switzerland, England, and Norway) move only one rank order
or less; and two cases (Italy and Sweden) move about three rank orders.
Lib-labism thus emerges as strongly, but not perfectly, associated with lib-
eralism (the Spearman rank order correlation for the data in the Table is .92;
the correlation is significant at a .01 level).

The use of ordinal comparison allows Luebbert to state certain find-
ings in a more nuanced way than is possible when nominal categories are
strictly employed. For example, he points out that among the countries
where lib-labism failed before World War I, Belgium and the Netherlands
"most closely approximated the British-French-Swiss pattern of liberal
hegemony" (Luebbert 1991, p. 56). That is, these two countries were
borderline success cases vis-à-vis the dichotomous explanatory variable
of lib-labism. In turn, Belgium and the Netherlands established interwar
regimes in which significant liberal elements were present, making them a
subtype of liberal regimes (Luebbert 1991, pp. 248, 250). It appears that

357

Table 10.3. *Ordinal Comparison in Luebbert's Analysis of Interwar Regimes*[*]

	Prewar Lib–Labism[†] (1 = least; 11 = most)	Interwar Liberalism (1 = least; 11 = most)
Switzerland	11	10
England	10	11
France	9	9
Belgium	7.5	7.5
Netherlands	7.5	7.5
Denmark	6	6
Italy	5	2
Norway	4	4.5
Spain	3	3
Sweden	2	4.5
Germany	1	1

[*] I have rank ordered cases based primarily on passages on the following pages of Luebbert's 1987 ("Social Origins") and 1991 (*Liberalism, Fascism, or Social Democracy*) works: England: (1987, p. 452), (1991, pp. 37–48, 166); Switzerland: (1991, pp. 49, 166); France: (1987, pp. 455–6), (1991, pp. 37–48, 166); Belgium and the Netherlands: (1987, p. 451), (1991, pp. 56–7, 248, 250); Denmark: (1987, p. 451), (1991, pp. 57–8, 236–7, 270–1); Sweden and Norway: (1991, pp. 57–8, 239–42, 270); Italy: (1991, pp. 57–8, 272–7); Spain: (1991, pp. 151–3, 272–7); and Germany: (1991, pp. 272–7).

[†] Luebbert measures lib-labism as the "degree of liberal hegemony" present before World War I.

Luebbert believes lib-labism is an almost always necessary and sufficient cause of interwar liberalism in Europe. In fact, his combined nominal-ordinal analysis could be seen as a rudimentary attempt to use fuzzy-set methods to assess this kind of causation.

In her work on social revolutions, Skocpol (1979) also uses ordinal comparison to supplement what is primarily a nominal argument. She does so by disaggregating nominal variables into constituent subvariables that are evaluated through ordinal comparison (see Mahoney 1999). Hence, while Skocpol's (1979, pp. 154–7) explanatory variable of "conditions for state breakdown" is treated as a dichotomous variable for the purpose of using the methods of agreement and difference, this variable is disaggregated into three constituent subvariables (international pressure, state autonomy, and agrarian backwardness) when evaluated through ordinal comparison. These constituent variables are ranked across all cases of revolution and

nonrevolution. Likewise, Skocpol's dichotomous outcome variable of social revolution is reassessed in terms of ordinal constituent processes. Social revolution is defined in part as "rapid, basic transformations of a society's state and class structures" (Skocpol 1979, p. 4), and Skocpol notes ordinal differences among cases along these dimensions. For example, social revolution unfolded most rapidly in Russia, least rapidly in China, and at an intermediate pace in France. These differences are explained in part by ordinal contrasts on key explanatory variables, including the extent of international pressure that marked the revolutionary process (Skocpol 1979, p. 172). In sum, even though Skocpol's book is famous for its use of nominal comparison through the methods of agreement and difference, ordinal analysis plays a major role in underpinning the nominal argument.

While Luebbert and Skocpol use ordinal analysis to strengthen and support their overall nominal arguments, other scholars use ordinal comparison as the principal cross-case method of investigation. One example is Collier and Collier's (1991) work on labor incorporation in eight Latin American countries. Collier and Collier identify four types of labor incorporation periods, and they seek to explain ordinal differences in the "scope of mobilization" that characterized these periods. They first eliminate certain explanatory factors that lack any consistent relation to the scope of mobilization. For example, the authors reject explanations centered on the strength of the labor movement because there is "no systematic relationship between labor movement strength and type of incorporation period" (p. 750). The authors then show how their main explanatory variable – the political strength of the oligarchy – does reveal a clear pattern with mobilization during the incorporation period. In particular, the authors show that there is "an inverse relation between the political strength of the oligarchy... and the degree to which... mobilization was pursued in the incorporation period" (p. 748). Although there is a clear inverse relationship for six of Collier and Collier's eight cases, two cases deviate from this pattern. In Peru and Argentina, the oligarchy was in many spheres quite powerful, yet these cases exhibited relatively high levels of labor mobilization during the reform period, thus seemingly violating the hypothesized inverse pattern. The authors explain these deviations based on crucial "flaws" in the strength of the Peruvian and Argentine oligarchies (pp. 748–9). Once these flaws are taken into account, the inverse relationship at work for the other six cases also makes sense for Peru and Argentina. In this way, the authors show how what initially appears to be a deviation in fact reflects the general inverse pattern once more appropriate measures are introduced.

Another work that relies extensively on ordinal comparison is Orloff's (1993) comparative historical analysis of social provision for the elderly in Britain, Canada, and the United States. Orloff not only conceptualizes her own central explanatory variables as ordinal categories, but also evaluates rival explanations using this type of assessment. For example, she (pp. 47–8) uses ordinal comparison to reject explanations of the relative timing of social provision that focus on changes in the number of aged persons present in society. She argues that the expected relationship between age distribution and the timing of pension legislation is not supported by her cases. For example, Britain witnessed only a marginal increase in the number of elderly citizens but nevertheless enacted pensions at an early date. By contrast, Canada saw a very substantial increase in the elderly population but adopted old age insurance at a relatively late date. For the United States, the elderly population exhibited intermediate to high levels, yet pensions were adopted at a very late date. Hence, there is no apparent relationship between the two variables, and Orloff eliminates size of the elderly population as a potential explanatory factor.

Within-Case Analysis

In addition to comparing cases with one another, most comparative historical analysts also compare processes drawn from within particular cases. This strategy of within-case analysis entails examining multiple features of what was originally considered only a single case to assess hypotheses developed through cross-case analysis. In making within-case evaluations, analysts will often rely on nominal and ordinal measurement. However, whereas the nominal and ordinal strategies discussed previously entail highly aggregated comparisons across cases, the procedures discussed in this section involve disaggregated comparisons within cases. Hence, a within-case strategy is distinguished from nominal and ordinal strategies in terms of level of aggregation.

Within-case analysis is a tool specifically designed to compensate for limitations associated with cross-case methods. The most general type of within-case analysis is "pattern matching," a procedure in which the analyst assesses cross-case associations in light of multiple within-case hypotheses. An important subtype of this procedure is "process tracing," a technique in which the analyst attempts to locate the causal mechanisms linking a hypothesized explanatory variable to an outcome. Finally, a third technique – "causal narrative" – combines cross-case and within-case

analysis by comparing cases in terms of highly disaggregated sequences of processes and events that lead to outcomes.

Pattern Matching

Causal patterns derived from cross-case comparison often suggest additional hypotheses about aspects of specific cases. Following a procedure that Campbell (1975) calls "pattern matching," comparative historical analysts test these additional hypotheses, evaluating whether patterns derived from cross-case analysis can be matched with observations from within specific cases.[13] Campbell (1975, p. 182) points out that pattern matching provides a powerful tool for theory falsification in small-N research: investigators routinely find that their arguments cannot be sustained when within-case hypotheses are assessed. Alternatively, if within-case observations are repeatedly consistent with a cross-case finding, researchers have stronger grounds for believing the cross-case finding is valid.

Comparative historical analysts pursue pattern matching using different levels of measurement. Both nominal and ordinal measurement can be used in conjunction with pattern matching. Researchers may also use interval measurement when assessing within-case hypotheses. Indeed, if a large number of within-case observations are measured at an interval level, researchers may employ statistical methods with the pattern-matching procedure.

A nice example of the use of statistical research for the purpose of pattern matching is Goldstone's (1991) work on revolutions during the early modern period. Goldstone's cross-case nominal argument suggests that demographic growth leads to revolutions by triggering structural crises (i.e., fiscal crises, elite–state and intraelite conflict, and mass opposition). To bolster this cross-case, small-N argument, he conceptualizes explanatory variables in terms of a large number of within-case quantitative measures and combines these measures into an overall "political stress indicator" that is evaluated statistically. These statistical evaluations are used as supplementary evidence to assess hypotheses that apply to a small number of cases, offering powerful confirmatory evidence in support of Goldstone's small-N argument.

In contrast to Goldstone's statistical analysis of within-case patterns, Luebbert (1991) uses ordinal and nominal comparison when employing

[13] See also Collier (1993), Eckstein (1975), and George (1979).

pattern matching. For example, his argument that an alliance between the socialist party and the middle peasantry (a "red–green" alliance) caused social democracy in interwar Europe has multiple within-case implications. Some of these within-case implications entail ordinal propositions – for example, the governing social coalition will lack a high level of stability; the working class will exercise a high degree of autonomy from the state; and high levels of strikes and labor activism will develop (Luebbert 1991, pp. 234–6). Other within-case implications entail nominal propositions – for example, socialists will not challenge the distribution of wealth in the countryside or try to mobilize the rural proletariat; and the middle peasantry will not provide a viable social base for the socialists (Luebbert, 1991, pp. 268–9, 272, 286–8). Luebbert's within-case analysis finds support for these hypotheses, significantly enhancing one's confidence that the cross-case argument is correct.

Regardless of the level of measurement (nominal, ordinal, interval) employed, the additional leverage offered by pattern matching helps compensate for the weaknesses of cross-case strategies. For example, one of the limitations of cross-case nominal methods is that several explanations may be supported by the data, leaving the analyst without a clear basis for deciding which explanatory factor is the most important. Pattern matching helps narrow the range of potential explanations by offering an additional means of eliminating variables. After variables are eliminated through pattern matching, analysts are often left with much more parsimonious explanations.

Pattern matching is also a key tool for those comparative historical researchers who seek to avoid the determinism of nominal methods but who lack enough cases to employ these methods probabilistically. Analysts can use pattern matching to show that a relationship is causal despite the fact that a cross-case nominal comparison reveals one or more cases in which scores on the explanatory and outcome variables deviate from a general pattern of matching. For example, if only three out of four cases reflect a general pattern of matching on nominal explanatory and outcome variables, scholars may conclude that the pattern reflects causation if they find significant within-case support. Likewise, analysts may use pattern matching to argue that a relationship is *not* causal even though scores on an explanatory variable are perfectly matched with scores on an outcome variable in a cross-case nominal assessment. They can do so by showing how the cross-case pattern is not supported when assessed against multiple within-case patterns. Hence, comparative historical analysts who choose to think about causation in probabilistic terms may use pattern matching as a basis for

retaining explanatory variables that do not withstand deterministic cross-case nominal tests while rejecting other explanatory variables even though they do withstand such tests.

Finally, pattern matching can be a valuable supplement to cross-case ordinal comparison. Pattern matching can help analysts make a better judgment about the causal status of a relationship that is ambiguous when evaluated through cross-case ordinal analysis. Likewise, pattern matching can call into question the findings of ordinal comparison, showing how an apparently causal relationship is in fact not causal when viewed in light of multiple within-case implications.

Process Tracing

An important part of causal analysis involves establishing that there is some association between explanatory variables and an outcome variable. Yet, for many comparative historical analysts, an equally important part involves identifying the causal mechanisms that link explanatory variables with the outcome variable (Blalock 1961, p. 9; Elster 1989, pp. 4–7; Goldthorpe 1997; Hedström and Swedberg 1998; Salmon 1984, ch. 5). Causal mechanisms can be defined as the processes and intervening variables through which an explanatory variable exerts a causal effect on an outcome variable (Bennett 1997). Following George and McKeown (1985), the effort to infer causality through the identification of causal mechanisms can be called "process tracing."[14]

Process tracing is often used to help the analyst who works with a small number of cases avoid mistaking a spurious correlation for a causal association. The problem of spuriousness arises when two correlated variables appear to be causally related but in fact are the product of an antecedent variable. In small-N research, cross-case comparative methods are often vulnerable to this problem. For example, when three temporally ordered variables are correlated in a sequence, small-N analysts have difficulty using cross-case methods to determine if the sequence represents a causal path or a spurious correlation. The first variable in such a sequence is often perfectly correlated with both the second and third variables. Small-N cross-case methods do not provide a strong basis for judging whether this first variable represents an antecedent cause that "explains away" the presumed

---

[14] Because causal mechanisms are within-case implications of cross-case patterns, process tracing is actually a particular application of pattern matching.

a

causal relationship between the second and third variables, or whether the first and third variables are correlated because of the presence of the second variable, in which case the idea of causal path makes sense. Process tracing can help the analyst distinguish between these two possibilities by showing whether causal mechanisms link the variables together. Thus, if hypothesized causal mechanisms can be identified between the second and third variables through process tracing, the analyst has a basis for believing that the sequence is a causal path; that is, the second variable has a real causal effect on the third variable. Alternatively, if causal mechanisms cannot be identified between the second and third variables, the analyst has grounds for believing that the sequence may be a spurious correlation; that is, the second and third variables are correlated only because of the presence of the first antecedent variable.

Comparative historical analysts frequently argue that a correlation identified through cross-case analysis is not causal because mechanisms linking the presumed explanatory variable and the outcome variable cannot be identified. For example, Skocpol's (1979, pp. 170–1) work on the origins of revolutions uses process tracing to reject causal variables – such as ideologically motivated vanguard movements – that were not eliminated through cross-case methods. Although ideologically motivated vanguard movements were present in her three cases of social revolution, she argues that they did not exert an important causal effect in bringing about revolutions. In particular, in contrast to what other scholars have hypothesized, she argues that vanguard movements were not responsible for triggering widespread revolts against landlords and state agents. Rather, according to Skocpol, vanguard movements were marginal to the central political processes that defined social revolutions, emerging on the scene only very late to take advantage of situations they did not create. Hence, she concludes that these movements were not a crucial cause of social revolutions in France, Russia, and China.

Likewise, Luebbert uses process tracing to eliminate the "Moore–Gerschenkron thesis," which holds that fascist regimes result from the presence of a labor-repressive landed elite that is able to draw substantial lower-class rural support for fascism (Luebbert 1991, pp. 308–9). Although there is a matching between the presence/absence of a repressive landed elite and the presence/absence of fascism, Luebbert suggests that the mechanisms through which this specific factor supposedly produces fascism are not supported by the historical record of the fascist cases. Thus, rural support for fascism was generally not present in areas where a landed elite predominated. Likewise, the evidence shows that the landed elites who could deliver

large numbers of votes did not usually support fascism (Luebbert 1991, pp. 308–9). In short, despite the matching, Luebbert rejects the Moore–Gerschenkron hypothesis because it is not validated by process tracing.

Other scholars use process tracing not to eliminate causal factors but to support their own explanations. For example, Collier and Collier (1991) identify mechanisms linking different types of labor incorporation periods with different types of party systems. In their analysis of Colombia and Uruguay, Collier and Collier systematically identify the processes and events through which the incorporation pattern of "electoral mobilization by a traditional party" led to the party system outcome of "electoral stability and social conflict." These processes included a period in which the party that oversaw incorporation briefly maintained power, the gradual emergence of conservative opposition, a period of intense political polarization, a military coup, and, finally, the creation of a party system marked by stable electoral politics and social conflict. Each of these events acts as a mechanism linking labor incorporation with a particular party system outcome. The ability of the authors to show how these and other processes connected explanatory and outcome variables is crucial to the success of their argument. Indeed, although any work can potentially benefit from process tracing, it is an especially important tool for those studies such as Collier and Collier's in which explanatory and outcome variables are separated by long periods of time.

Causal Narrative

A final procedure illustrates how comparative historical researchers use cross-case comparisons of within-case chronologies as a basis for making causal inferences. With this technique of "causal narrative," to use Sewell's (1996) terminology, the analyst attempts to validate aggregated cross-case associations by "breaking apart" variables into constituent sequences of disaggregated events and comparing these disaggregated sequences across cases. The purpose of unpacking aggregated variables through narrative is not only to provide a contextualized description of cases; rather, the goal is to support a cross-case argument at a more disaggregated level.

This technique relies on historical narrative, which has received significant attention in recent methodological discussions (e.g., Abbott 1990, 1992; Aminzade 1992; Franzosi 1998; Griffin 1992, 1993; Haydu 1998; Isaac 1997; Mahoney 1999; Somers 1992; Stryker 1996). However, the procedures through which analysts decide whether a narrative account lends

support to a cross-case causal pattern have not been well specified. Griffin's (1993; see also Heise 1989) discussion of event-structure analysis is the most well developed statement on how narrative can be wedded to causal inference. Event-structure analysis provides a formal apparatus for unpacking events and reconstituting their constituent parts as a causal interpretation of historical processes. This procedure can underpin causal narrative by identifying the causally linked processes that constitute highly aggregated variables in cross-case analysis (Mahoney 1999, pp. 1165–7). With causal narrative, the analyst compares event sequences across cases to determine if the cases can reasonably be seen as following aggregated causal patterns at a more fine-grained level. In this sense, causal narrative entails the matching of event structures across cases (see Griffin and Ragin 1994, pp. 14–15; Sewell 1996, p. 262).[15] In addition, causal narrative can be used to show how two or more cases that are marked by important differences in causal processes at an aggregated level of analysis are also characterized by substantially different event structures at a disaggregated level.

A good example of the use of causal narrative to compare event structures is found in Skocpol's (1979) work on social revolutions. Many of Skocpol's key explanatory variables are actually made up of numerous causally linked processes. Likewise, the outcome of social revolution is itself composed of a series of causally connected events. These constituent processes represent an event-structure pattern that could be formally diagrammed and compared across cases (see Mahoney 1999). Although Skocpol does not carry out a formal mapping of event structures, she does implicitly compare the event structure of her cases to judge whether they follow a similar causal logic at a disaggregated level. According to Sewell (1996), Skocpol's ability to show that a similar event sequence is at work in each case of social revolution greatly contributes to the persuasiveness of her argument.

An interesting example of a work that uses causal narrative to contrast event-structure sequences is Yashar's (1997) excellent analysis of the origins of democracy in Costa Rica in 1949 and authoritarianism in Guatemala in 1954. Yashar notes that both Costa Rica and Guatemala experienced major democratic and social reform periods in the 1940s and 1950s. However, her narrative shows that these reform periods were actually composed of quite different event processes, and these different processes were highly

---

[15] Causal narrative cannot be used to assess cases that arrive at a given outcome through different causal processes. When this is true, one would expect the cases to be characterized by different – not similar – event-structure sequences.

consequential for the development of contrasting regimes. Thus, Yashar's narrative shows how particular actor choices about coalitional allies triggered differing reform efforts, reactions, and counterreactions. These differences in event sequences lend support to Yashar's overarching argument, which stresses the importance of political coalitions and alliance patterns.

Most comparative historical analysts use causal narrative as an informal technique presented through "stories" of event processes. It remains to be seen whether causal narrative can be more formally employed through event-structure diagrams that explicitly map each step and logical connection in a narrative argument. Griffin's (1993) work on event-structure analysis suggests that formally diagramming narratives can be complicated when a large number of events are considered. Yet, without such formal diagramming, the procedures through which analysts compare and contrast event-structure sequences across cases cannot be easily evaluated. The complex trade-offs between the clarity of informal narrative presentations and the rigor of explicitly diagrammed narrative accounts warrant further attention from methodologists.

Conclusion

Methodological recipes for producing successful research are neither possible nor desirable. Nevertheless, an awareness of the different tools available can help analysts improve the quality of their work and better present their findings to scholars from other traditions. Toward this end, this essay has identified and discussed three basic strategies of causal assessment used in comparative historical research: nominal comparison, ordinal comparison, and within-case analysis. By way of conclusion, I would like to bring together some of the unresolved questions and future methodological agendas suggested by the discussion of these strategies.

First, with respect to nominal methods, the study of necessary and sufficient conditions (including *probabilistic* necessary and sufficient conditions) deserves more attention from methodologists, including especially quantitative methodologists who do not ordinarily think about causation in this way. Comparative historical researchers have productively studied these conditions, and their work provides a valuable point of departure for those who are interested in learning more. It is indeed unfortunate that graduate students often lack the training to meaningfully evaluate arguments about necessary and sufficient causation. I would encourage both quantitative and qualitative methodologists to learn more about these methods, include

367

them as a basic component of their courses on research methods, and let students make up their own minds about potential applications in comparative research.

Second, the fact that many comparative historical researchers combine an ordinal strategy with a nominal strategy suggests that they may think about ordinal relationships in terms of necessary and sufficient conditions. Yet, other comparative historical researchers believe ordinal relationships reflect a linear pattern of causation similar to that studied in much statistical research. These two alternatives need to be sorted out in future work. In general, questions about the extent to which linear statistical findings can or should be translated into the language of necessary and sufficient conditions will not be resolved until methodologists give more attention to nominal methods.

Finally, more methodological work needs to be done on the relationship between cross-case and within-case analysis. Part of the problem in understanding this relationship is that comparative historical researchers often do not say enough about how they use within-case analysis in their substantive studies. For example, we still do not have many systematic discussions of within-case analysis informing the selection of categories and cases *before* formal nominal tools are applied. Moreover, once initial nominal associations are established, more needs to be said about the application of particular types of within-case analysis. With pattern matching, researchers need to consider the special issues that may arise when a given study uses both nominal and statistical methods to evaluate the same within-case observations. Likewise, when using process tracing, methodologists need to explore more seriously the extent to which it is really possible to specify causal mechanisms as empirical hypotheses with directly testable implications. Regarding causal narrative, the overall debate between using words versus formally diagramming event structures as a means of representing disaggregated patterns of within-case causation needs to be sorted out.

These agendas will likely be at the forefront of the next generation of work on comparative historical methods.

References

Abbott, Andrew. 1990. "Conceptions of Time and Events in Social Science Methods: Causal and Narrative Approaches." *Historical Methods* 23: 140–50.

1992. "From Causes to Events: Notes on Narrative Positivism." *Sociological Methods and Research* 20: 428–55.

•

Amenta, Edwin and Jane D. Poulsen. 1994. "Where to Begin: A Survey of Five Approaches to Selecting Independent Variables for Qualitative Comparative Analysis." *Sociological Methods and Research* 23: 22–53.

Aminzade, Ronald. 1992. "Historical Sociology and Time." *Sociological Methods and Research* 20: 456–80.

Bennett, Andrew. 1997. "Lost in the Translation: Big (*N*) Misinterpretations of Case Study Research." Paper Presented at the 38th annual convention of the International Studies Association, Toronto, March 18–22.

Blalock, Hubert M. 1961. *Causal Inferences in Nonexperimental Research*. Chapel Hill: University of North Carolina Press.

Braumoeller, Bear F. and Gary Goertz. 2000. "The Methodology of Necessary Conditions." *American Journal of Political Science* 44: 844–58.

Campbell, Donald T. 1975. "'Degrees of Freedom' and the Case Study." *Comparative Political Studies* 8: 178–93.

Colburn, Forest D. 1994. *The Vogue of Revolution in Poor Countries*. Princeton, NJ: Princeton University Press.

Collier, David. 1993. "The Comparative Method." Pp. 105–19 in *Political Science: The State of the Discipline II*, edited by Ada Finifter. Washington, DC: American Political Science Association.

Collier, David and Robert Adcock. 1999. "Democracy and Dichotomies: A Pragmatic Approach to Choices about Concepts." *Annual Review of Political Science* 2: 537–65.

Collier, David and James Mahoney. 1996. "Insights and Pitfalls: Selection Bias in Qualitative Research." *World Politics* 49: 56–91.

Collier, Ruth Berins and David Collier. 1991. *Shaping the Political Arena: Critical Junctures, the Labor Movement, and Regime Dynamics in Latin America*. Princeton, NJ: Princeton University Press.

DeFelice, E. Gene. 1986. "Causal Inference and Comparative Methods." *Comparative Political Studies* 19: 415–37.

Dion, Douglas. 1998. "Evidence and Inference in the Comparative Case Study." *Comparative Politics* 30: 127–46.

Downing, Brian M. 1992. *The Military Revolution and Political Change: Origins of Democracy and Autocracy in Early Modern Europe*. Princeton, NJ: Princeton University Press.

Drass, Kriss and Charles C. Ragin. 1999. *QC/FSA: Qualitative Comparative/Fuzzy-Set Analysis*. Evanston, IL: Institute for Policy Research, Northwestern University.

Eckstein, Harry. 1975. "Case Studies and Theory in Political Science." Pp. 79–138 in *Handbook of Political Science* 7, edited by Fred Greenstein and Nelson W. Polsby. Reading, MA: Addison-Wesley.

Elster, Jon. 1989. *Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.

Ertman, Thomas. 1997. *Birth of the Leviathan: Building States and Regimes in Medieval and Early Modern Europe*. Cambridge: Cambridge University Press.

Franzosi, Roberto. 1998. "Narrative as Data: Linguistic and Statistical Tools for the Qualitative Study of Historical Events." *International Review of Social History* 43: 81–104.

Geddes, Barbara. 1990. "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics." Pp. 131–50 in *Political Analysis*, vol. 2, edited by James A. Stimson. Ann Arbor: University of Michigan Press.

George, Alexander L. 1979. "Case Studies and Theory Development: The Method of Structured, Focused Comparison." Pp. 43–68 in *Diplomacy: New Approaches in History, Theory, and Policy*, edited by P. G. Lauren. New York: Free Press.

George, Alexander L. and Timothy J. McKeown. 1985. "Case Studies and Theories of Organizational Decision Making." *Advances in Information Processing in Organizations* 2: 21–58.

Giddens, Anthony. 1979. *Central Problems in Social Theory: Action, Structure and Contradiction in Social Analysis*. Berkeley: University of California Press.

Goldstone, Jack A. 1991. *Revolution and Rebellion in the Early Modern World*. Berkeley: University of California Press.

——— 1997. "Methodological Issues in Comparative Macrosociology." *Comparative Social Research* 16: 107–20.

Goldthorpe, John H. 1997. "Current Issues in Comparative Macrosociology: A Debate on Methodological Issues." *Comparative Social Research* 16: 1–26.

——— 2000. *On Sociology: Numbers, Narratives, and the Integration of Research and Theory*. Oxford: Oxford University Press.

Griffin, Larry J. 1992. "Temporality, Events, and Explanation in Historical Sociology: An Introduction." *Sociological Methods and Research* 20: 403–27.

——— 1993. "Narrative, Event-Structure, and Causal Interpretation in Historical Sociology." *American Journal of Sociology* 98: 1094–133.

Griffin, Larry and Charles C. Ragin. 1994. "Some Observations on Formal Methods of Qualitative Analysis." *Sociological Methods and Research* 23: 4–21.

Haydu, Jeffrey. 1998. "Making Use of the Past: Time Periods as Cases to Compare and as Sequences of Problem Solving." *American Journal of Sociology* 104: 339–71.

Hedström, Peter and Richard Swedberg, eds. 1998. *Social Mechanisms: An Analytical Approach to Social Theory*. New York: Cambridge University Press.

Heise, David. 1989. "Modeling Event Structures." *Journal of Mathematical Sociology* 14: 139–69.

Isaac, Larry W. 1997. "Transforming Localities: Reflections on Time, Causality, and Narrative in Contemporary Historical Sociology." *Historical Methods* 30: 4–12.

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

Lieberson, Stanley. 1985. *Making It Count: The Improvement of Social Research and Theory*. Berkeley: University of California Press.

——— 1991. "Small *N*'s and Big Conclusions: An Examination of the Reasoning in Comparative Studies Based on a Small Number of Cases." *Social Forces* 70: 307–20.

1994. "More on the Uneasy Case for Using Mill-*Type* Methods in Small-*N* Comparative Studies." *Social Forces* 72: 1225–37.

1998. "Causal Analysis and Comparative Research: What Can We Learn from Studies Based on a Small Number of Cases?" Pp. 129–45 in *Rational Choice Theory and Large-Scale Data Analysis*, edited by Hans-Peter Blossfeld and Gerald Prein. Boulder, CO: Westview.

Lijphart, Arend. 1971. "Comparative Politics and the Comparative Method." *American Political Science Review* 65: 682–93.

1975. "The Comparable Cases Strategy in Comparative Research." *Comparative Political Studies* 8: 158–77.

Loung, Pauline Jones. 2002. *Institutional Change and Political Continuity in Post-Soviet Central Asia: Power, Perceptions, and Pacts*. New York: Cambridge University Press.

Luebbert, Gregory M. 1987. "Social Foundations of Political Order in Interwar Europe." *World Politics* 39: 449–78.

1991. *Liberalism, Fascism, or Social Democracy: Social Classes and the Political Origins of Regimes in Interwar Europe*. New York: Oxford University Press.

Mahoney, James. 1999. "Nominal, Ordinal, and Narrative Appraisal in Macrocausal Analysis." *American Journal of Sociology* 104: 1154–96.

2001. "Beyond Correlational Analysis: Recent Innovations in Theory and Method." *Sociological Forum* 16: 575–93.

Mahoney, James and Richard Snyder. 1999. "Rethinking Agency and Structure in the Study of Regime Change." *Studies in Comparative International Development* 34: 3–32.

Marx, Anthony W. 1998. *Making Race and Nation: A Comparison of the United States, South Africa, and Brazil*. Cambridge: Cambridge University Press.

Mill, John Stuart. [1843] 1974. *A System of Logic*. Toronto: University of Toronto Press.

Most, Benjamin and Harvey Starr. 1989. *Inquiry, Logic, and International Politics*. Columbia: University of South Carolina Press.

Munck, Gerardo L. 1998. "Canons of Research Design in Qualitative Analysis." *Studies in Comparative International Development* 33: 18–45.

Nichols, Elizabeth. 1986. "Skocpol on Revolution: Comparative Analysis vs. Historical Conjuncture." *Comparative Social Research* 9: 163–86.

Orloff, Ann Shola. 1993. *The Politics of Pensions: A Comparative Analysis of Britain, Canada, and the United States, 1880–1940*. Madison: University of Wisconsin Press.

Przeworski, Adam and Henry Teune. 1970. *The Logic of Comparative Social Inquiry*. New York: Wiley.

Ragin, Charles C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.

1997. "Turning the Tables: How Case-Oriented Research Challenges Variable-Oriented Research." *Comparative Social Research* 16: 27–42.

2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.

371

Rueschemeyer, Dietrich and John D. Stephens. 1997. "Comparing Historical Sequences – A Powerful Tool for Causal Analysis." *Comparative Social Research* 17: 55–72.

Rueschemeyer, Dietrich, Evelyne Huber Stephens, and John D. Stephens. 1992. *Capitalist Development and Democracy.* Chicago: University of Chicago Press.

Salmon, Wesley. 1984. *Scientific Explanation and the Causal Structure of the World.* Princeton, NJ: Princeton University Press.

Sartori, Giovanni. 1987. *Theory of Democracy Revisited.* Chatham, NJ: Chatham House.

Selbin, Eric. 1993. *Modern Latin American Revolutions.* Boulder, CO: Westview.

Sewell, William H., Jr. 1996. "Three Temporalities: Toward an Eventful Sociology." Pp. 245–80 in *The Historic Turn in the Human Sciences,* edited by Terrence J. McDonald. Ann Arbor: University of Michigan Press.

Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China.* Cambridge: Cambridge University Press.

——— 1994. "Reflections on Recent Scholarship about Social Revolutions and How to Study Them." Pp. 301–44 in *Social Revolutions in the Modern World,* edited by Theda Skocpol. Cambridge: Cambridge University Press.

Skocpol, Theda and Margaret Somers. 1980. "The Uses of Comparative History in Macrosocial Inquiry." *Comparative Studies in Society and History* 22: 174–97.

Smelser, Neil. 1976. *Comparative Methods in the Social Sciences.* Englewood Cliffs, NJ: Prentice-Hall.

Somers, Margaret R. 1992. "Narrativity, Narrative Identity, and Social Action: Rethinking English Working-Class Formation." *Social Science History* 16: 591–630.

Stryker, Robin. 1996. "Beyond History versus Theory: Strategic Narrative and Sociological Explanation." *Sociological Methods and Research* 24: 304–52.

Wickham-Crowley, Timothy. 1992. *Guerrillas and Revolution in Latin America: A Comparative Study of Insurgents and Regimes Since 1956.* Princeton, NJ: Princeton University Press.

Yashar, Deborah J. 1997. *Demanding Democracy: Reform and Reaction in Costa Rica and Guatemala, 1870s–1950s.* Stanford, CA: Stanford University Press.