

Making Comparative Analysis Count

Charles C. Ragin
Department of Sociology
University of Arizona
Tucson, Arizona 85721 USA

August 2003

Preliminary Draft

Please do not quote or cite without author's permission.

Introduction

This essay takes as its starting point the assumption that general social scientific knowledge is both possible and desirable. I call this an *assumption* because there are some social scientists who do not believe general knowledge is possible, and there are others who believe that it is not desirable, regardless of whether it is possible. (Of course, these positions depend in part on the definition of *general knowledge*.) Most social scientists today would agree that general knowledge is possible in the study of individual-level social phenomena, but not all would agree that it is possible in the study of historically situated macrolevel phenomena (e.g., Lieberman 1991; 1994). This essay, however, assumes that it is possible and asks: What kind of general knowledge of macrolevel phenomena should we seek?

Often, when social scientists hear the phrase *general knowledge* they immediately start thinking in terms of relationships between abstract concepts represented in terms of variables. They have been trained to equate general knowledge with discourse about relationships between variables. For example, a social scientist might observe that the most economically advanced countries are also stable democracies and from this observation posit that there is a general relationship between *development* and *democracy*. Thus, he or she might state, in general knowledge terms, that "economic development furthers democratic stability, as seen in the correlation between the variables democracy and development." I argue that general knowledge can come in other forms and that it

is not dependent on a discourse grounded in correlations between variables. Indeed, to limit *general knowledge* to that which is based on relationships between variables skews the debate on methods in favor of variable-oriented approaches. Instead, general knowledge can and should be built up from knowledge of cases. Knowledge of multiple cases, in turn, can be explored and represented in many different ways, not simply in terms of variables and their correlations. I sketch this alternate approach, which I call *configurational*, in this essay.

My argument begins with a discussion of the *N of cases continuum* in macrolevel social research. At one end of this continuum is the qualitative case study; at the other end is the large-N quantitative study. Social research tends to cluster at the two ends of this continuum, the very-small-N end and the large-N end. I sketch the main features of these two approaches as a series of contrasts and then examine middle paths between them. I discuss in detail the strengths of *configurational comparative research* as a middle path. This middle path is especially attractive because of its consistency with the goal of building general knowledge from case-based knowledge.

The N of Cases Continuum

The examination of almost any research area in the social sciences today reveals that there is a sharp divide separating those who do qualitative case-study research and those who do large-N, variable-oriented research. In comparative sociology and comparative politics, for example, a frequency distribution showing the number of studies with different size Ns reveals a clear U-shaped pattern, illustrated in Figure 1 (see Bollen, Entwistle, and Alderson 1993; Ragin 1989; Sigelman and Gadbois 1983). At the small-N end of the horizontal axis there are many studies, just as there are at the large-N end of this axis. But in the middle the relative number of studies is very low. There are few comparativists who study 10 or 20 countries, but many who study 1 or 2 (case-study research) or more than 50 (i.e., enough to permit the use of conventional quantitative methods).

The research styles at the two ends of the N of cases continuum offer sharply contrasting ways of studying social phenomena. The case-study researcher examines many aspects of each case and attempts to construct a representation of the case from the interconnections among its aspects. In effect, the case-study researcher's goal is to show how the different aspects of the case mutually constitute the whole. The variable-oriented researcher, by contrast, studies one or a small number of dependent variables across a very large number of cases and attempts to identify a parsimonious set of causal variables that

explains as much variation as possible in the dependent variable. This researcher constructs a general model based on patterns observed across many cases (e.g., using correlations among variables). Thus, small-N research tends to be case-oriented, qualitative, and intensive, while large-N research tends to be variable-oriented, quantitative, and extensive.

This division within the social sciences reproduces the gulf between the humanities, where the key issue is the possibility and adequacy of representation, and the physical sciences, where the problem of scientific inference is paramount. Social science lies in the middle of the continuum that runs from the humanities to the physical sciences and reproduces the polarity between them. Representation and inference are both important in social science, but social scientists differ in how they weight these two goals. At its core, this broad division among social scientists reflects their disagreement regarding the limits of general knowledge. Those who study the complexity and specificity of social phenomena tend to be skeptical of the possibility of general social scientific knowledge, while those who study patterns across many cases do so with an explicit eye toward constructing such knowledge.

A number of factors reinforce both this division and, more generally, the tendency for social scientists to gravitate toward the ends of the N-of-cases continuum:

Practical: It is very difficult to conduct an in-depth study of a large number of cases. Imagine trying to keep up with the day-to-day politics of 80 countries, to know each country's politics well. Imagine a survey researcher trying to keep track of the lives of hundreds of individuals. After a dozen or so cases, the interviewees would start to blur together. To treat each case as a singular entity and understand it on its own terms is very difficult and time consuming. Thus, social scientists who do in-depth research tend to restrict themselves to small Ns. Those who conduct quantitative analyses of large Ns usually must forfeit the opportunity to gain in-depth knowledge of the cases they study.

Social: Social scientists tend to self-identify as qualitative (few cases, many variables) or quantitative (many cases, relatively few variables) and also tend to sort themselves according to their familiarity with and tolerance for quantitative methods of data analysis. Graduate training is organized along this divide, reproducing it among students, and "invisible colleges" (Crane 1972) have formed based on these two methodological orientations. Thus, the social organization of disciplines tends to reinforce the U-shaped pattern illustrated in Figure 1.

Methodological: Social scientists have a limited repertoire of methods. Case-study methods are geared toward the acquisition of in-depth knowledge of cases, with an eye toward understanding how all the different pieces fit together *within* each case. Generally, the "how" of each case is something that researchers construct in their heads, and it is very difficult to keep mental track of configurations of similarities and differences across more than a handful of cases. Quantitative methods, by contrast, work well only when there are many cases--the more the better. The success of a research project often hinges on having a large N, which, in turn, facilitates statistical significance. It is very difficult to achieve statistical significance with an N of 20; it is much easier with an N of 200; and it is almost automatic with an N of 2,000. When researchers attempt in-depth analysis of an intermediate-sized N (say, an N of 10 or 20), they often are confronted with what seems like unmanageable complexity.

Epistemological: Social scientists value both kinds of knowledge--general knowledge *and* culturally or historically specific knowledge. Case-oriented methods are best suited for questions that are about culturally or historically specific phenomena, and variable-oriented methods are used routinely to identify broad, theoretically relevant patterns. From this perspective, the U-shaped distribution illustrated in Figure 1 may appear to represent a justifiable and convenient division of labor, with case-study researchers seeking in-depth knowledge and quantitative researchers seeking knowledge of general patterns (Eckstein 1975).

Given the prevalence of the U-shaped distribution of number of studies by number of cases and the reinforcing factors just described, it is not surprising that the gulf between these two worlds of social research is one of the most enduring features of social science. Those who do case-oriented work can offer answers that are intensively correct, embracing a small number cases in a detailed and integrative way. Those who do variable-oriented studies can provide answers that are extensively correct, embracing many observations. The case-oriented researcher justifies findings by showing their compatibility with other aspects of the case or cases in question. The variable-oriented researcher justifies findings by showing their generality.

Major Contrasts Between the Two Dominant Types of Research

Table 1 sketches a series of contrasts between case-study research and variable-oriented research. The two are contrasted in terms of proximate goals,

conception of cases and populations, N of cases, role of theory, conception of outcomes, understanding of causation, and relative emphasis on within case versus cross-case analysis. I address each of these contrasts in turn:

1. Proximate goals: Variable-oriented social research focuses on the goal of documenting general patterns characterizing a large population of observations. If researchers can demonstrate a relationship between one or more independent variables and a dependent variable, then they can better predict cases' values on the dependent variable given knowledge of their scores on the predictor variables. Typically, the study of general patterns is conducted with a sample of observations drawn from a large population. The researcher draws inferences about general patterns in the larger population based on his or her analysis of the sample.

In case-study research the most common goal is the representation of one or a very small number of cases, selected because they are substantively or theoretically important in some way (Eckstein 1975). For example, a researcher might use the case-study method to study a social movement in an in-depth manner. Suppose this movement was considered to be especially successful in mobilizing its members for collective action. To find out how they did it, the researcher would conduct in-depth study of the movement in question. At the completion of this case-study, the case-study researcher would not only be able to construct a useful representation of the movement in question, but also would be able to identify the next "best" case for investigation, one that would allow further development of the ideas advanced in the initial study.

2. Constitution of cases and populations: Variable-oriented researchers rarely devote much intellectual energy to the problem of constituting cases and populations. The ideal-typic case is the individual survey respondent, found in a *given* population, which in turn is demarcated by geographic, temporal, and demographic boundaries. The key problematic is how to derive a representative sample from the very large population of observations that is presumed to be at the researcher's ready disposal. When dealing with macrolevel units (e.g., organizations, countries, etc.), variable-oriented researchers attempt to force these units into the survey format, viewing their cases as generic observations drawn from an empirically given population.

Case-study researchers, by contrast, treat cases as singular, whole entities purposefully defined and selected, not as homogeneous observations drawn at random from a pool of equally plausible selections. Once a case is selected, the

key problematic is often to determine to which larger category the case belongs. In other words, the researcher seeks to answer the question, "What is my case a case of?" The answers to this question may be multiple and may involve comparing the case in question to a variety of related cases. Sometimes, the in-depth study of a single case leads to the development of an entirely new conceptual category, which in turn, may have important implications for existing theoretical schemes and typologies.

3. N of cases: An important lesson in every course in quantitative social research is that "more cases is better." More is better in three main ways. First, researchers must meet a threshold number of cases in order even to apply quantitative methods, usually cited as an N of 30 to 50. Second, the smaller the N, the more the data must satisfy the assumptions of statistical methods, for example, the assumption that variables are normally distributed or the assumption that subgroup variances are roughly equal. However, small Ns almost guarantee that such assumptions will not be met in most social research, especially when the cases are macrolevel. Third, as already mentioned, the greater the number of cases, the easier it is to produce statistically significant results. This bias toward large Ns dovetails with the implicit assumption that cases are empirically given, not constructed by the researcher, and that they are naturally abundant. The only problem, in this light, is whether the researcher is willing and able to gather data on as many cases as possible, preferable hundreds if not thousands.

By contrast, case-study research is very often defined by its focus on phenomena that are of interest because they are rare--that is, precisely because the N of cases is small. Typically, these phenomena are large-scale and historically delimited, not generic in any sense. The population of cases relevant to an investigation may be limited by the historical record to only a few, maybe only one. The key contrast with variable-oriented research derives from the simple fact that many phenomena of interest to social scientists and their audiences are historically or culturally significant. To argue that social scientists should study only cases that are generic and abundant or that can be studied only in isolation from their historical and cultural contexts would severely limit both the range and value of social science. One of the key lessons in qualitative research is that having fewer cases is often better. Otherwise, in-depth knowledge is sacrificed.

4. Role of theory: Conventional presentations of social science methodology place great emphasis on theory testing. In fact, its theory-testing orientation is often presented as what makes social science scientific. Researchers are advised to follow the scientific method and develop their hypotheses in

isolation from the analysis of empirical evidence. It is assumed that existing theory is sufficiently well-formulated to permit the specification of testable hypotheses and that social scientific knowledge advances primarily through the rejection of theoretical ideas that consistently fail to find empirical support.

It is without question that theory plays a central role in social research and that in fact almost all social research is heavily dependent on theory in some way. However, it is usually not possible to apply the theory-testing paradigm in case-study research. Case-study research has a very strong inductive component. The immediate objective of most case-study research is to explain the "how" of historically or culturally significant phenomena, for example: How do guerilla movements form? Theory plays an important orienting function by providing important leads and guiding concepts for empirical research, but existing theory is rarely well-formulated enough to provide explicit hypotheses. The primary theoretical objective of case-study research is not theory testing, per se, but concept formation, elaboration, and refinement.

5. Conception of outcomes: One of the most fundamental notions in textbook presentations of social research is the idea of the variable--a trait or aspect that varies from one case to the next--and the associated idea of looking for patterns in how variables are correlated across cases. For example: Do richer countries experience less political turmoil than poorer countries? If so, then social scientists might want to claim that variation in political turmoil across countries (the dependent or outcome variable) is explained in part by variation in country wealth (the independent or causal variable). Implicit in these notions about variables is the principle that the phenomena that social scientists wish to explain must vary across the cases they study; otherwise, there is nothing to explain. Thus, each case can be characterized in terms of the degree to which it displays the attribute in question (e.g., political turmoil).

In case-study research, by contrast, cases are selected for study because of the qualitatively distinct outcomes they exhibit. The outcome is often a qualitative change in the case under investigation, an historically emergent phenomenon or pattern that constitutes a break of some sort with what existed before. In short, rather than studying an attribute or aspect that simply "varies" from one case to the next, as in variable-oriented research, case-study researchers focus on historically emergent, qualitative phenomena. Thus, in case-study research the key concern is not to account for differences in levels of the outcome (or "dependent variable") across cases, but to account for qualitative change in each case. When case-study researchers study more than one case, they are often interested in identifying

commonalities across comparable qualitative outcomes, not variation across cases in level.

6. Understanding of causation: The main goal of most analyses of social data, according to textbook presentations of the logic of social research, is to assess the relative importance of independent variables as causes of variation in a dependent variable. For example, a researcher might want to know which has the greater impact on the longevity of democratic institutions, their design or their perceived legitimacy. In this view, causal variables compete with each other to explain variation in an outcome variable. A good contender in this competition is an independent variable that is strongly correlated with the dependent variable but has only weak correlations with its competing independent variables.

Case-study researchers, by contrast, usually look at causes in terms of their combination: How did relevant causes combine to produce the outcome in question? Rather than viewing causes as competitors, case-study researchers view them as raw ingredients that combine to produce the qualitative outcomes they study. John Stuart Mill called this type of causation "chemical" because the effect of any specific causal condition depends on the presence and absence of other conditions. After constituting and selecting one or more instances of an outcome, the case-study researcher examines each case in depth, using theoretical concepts, substantive knowledge, and interests as guides, in order to answer the question of "how" the outcome came about in each case.

7. Within versus cross-case analysis: In variable-oriented social research, the analysis of cross-case patterns is the primary means for linking aspects of cases. For example: is there a connection between development and democracy? The variable-oriented method for answering this question is to compute the correlation, across many cases, between these development and democracy, as aspects that vary from one case to the next. This correlation gauges the strength of the connection between these two aspects. If the correlation is very weak, then the conclusion may be that there is no substantial connection.

Computing a correlation across cases, however, is very different from examining one case at a time to determine whether development has contributed to democracy and if so, how. This alternate approach to the analysis of the connections between case aspects, which focuses on *how* aspects are connected in each case, is central to case-oriented comparative research. In this approach, the key issues are (1) establishing whether or not there is a connection between aspects in each case, and (2) assessing the nature of the mechanisms that

instantiate and animate the connections that are found. In case-study research, connections between aspects are usually made *within* each case, not *across* cases, and the mechanisms animating a given connection may differ from one case to the next.

Middle Paths Between Case-Study and Variable-Oriented Research

Despite the fact that case-study and variable-oriented researchers utilize a common set of theoretical ideas and share many substantive interests, the contrasts between these two approaches are stark. There seems to be little hope of reconciling them. One very prominent attempt to do so was published in 1994 by Gary King, Robert Keohane and Sidney Verba, *Designing Social Inquiry*. This work generated a great deal of debate, both before and after its publication. The collaboration of these three internationally renowned political scientists, with a well-known quantitative methodologist as lead author, seemed to hold out the promise of resolving a long-standing methodological schism.

King et al.'s vision of the middle path is straightforward. Unlike some (e.g., Lieberman), they accept the scientific validity and utility of case-study research, acknowledging its many strengths. Their essential recommendation is that case-oriented scholars conduct their research in ways that enhance its compatibility with variable-oriented research. For example, they discourage use of the common case-oriented strategy of looking for shared causal conditions across multiple instances of an outcome (e.g., across several anti-colonial revolutions). From the viewpoint of variable-oriented research, this strategy is flawed because (1) neither the outcome nor the causes vary substantially across cases and (2) it commits the error of "selecting on the dependent variable," a practice that is widely discouraged in textbooks on quantitative methods. Whether or not this common case-oriented strategy is defective, as King et al. maintain, it is clearly at odds with the variable-oriented approach. Their implicit argument is that if case-study researchers would abandon this unscientific practice, then it would be easier to reconcile the results of case-study research with the findings of variable-oriented research.

Of course, this recommendation is only one among the many suggestions offered by King et al. Most of their advice follows directly from the variable-oriented column of Table 1. For example, it is clear that they believe that case-study researchers should increase their N of cases whenever possible, in order to gain "analytic leverage" (the ability to assess the relative importance of independent variables). If it is not possible to increase the N at the macrolevel

(e.g., at the level of countries), then the analysis should shift to a lower level, where cases are more abundant (e.g., at the level of subnational units such as provinces within countries). It is also clear that they assume that theory testing, not learning about cases and representing this knowledge, is the central goal of empirical social science. They argue that case-study researchers should organize their investigations from the outset according to well formulated hypotheses. It is also clear that they do not consider the constitution of cases and populations as inherently problematic, but rather tend to see them as empirically given, and so on.

Overall, King et al.'s recommendations are thoughtful and some are quite useful. However, almost all their recommendations assume the priority and preeminence of variable-oriented research. Their vision of the middle path is that it is an extension of the principles of variable-oriented research to case-oriented research. This vision is commanding, but it is also limited.

Its first shortcoming is that it assumes that valid general knowledge follows directly from proper application of variable-oriented methods. In essence, King et al. assert that social science already has a good technology for generating general knowledge, variable-oriented methods, and that the task at hand is to remake case-oriented methods so that they generate findings that are more easily reconciled with variable-oriented research. The problem with this view is that it leaves unquestioned the assumption that variable-oriented methods produce valid and useful general knowledge. Even if we accept the idea that variable-oriented methods yield valid general knowledge, we can still debate whether this type of general knowledge--statements about the relative importance of independent variables viewed in analytic isolation from one another--is useful knowledge. (I return to this issue subsequently.)

The second problem with their vision of the middle path is that it is at odds with everyday logic and experience. The most common route to general knowledge, especially of social phenomena, is through accumulated knowledge of specific instances or cases. In everyday experience we build knowledge of the general from knowledge of the particular. For example, we learn about the temperaments of our colleagues from our repeated interactions with them. Sometimes we test what we have learned, as when we predict what a colleague will say or do in an upcoming meeting, but our "tests" rest on a firm foundation of knowledge of specific instances. In this light, the middle path between case-study and variable-oriented research should consist of methods for building general knowledge from case-based knowledge, that is, from understandings of specific cases. It should not consist of methods that supplant case-based knowledge with

an altogether different form, organized around the attempt to isolate the separate effects of independent variables. If social scientists were to construct general knowledge out of accumulated case-based knowledge, then Figure 1 would *not* be U-shaped. Instead, the line should have a negative, monotonic slope, as shown in Figure 2.

Third, and finally, the middle path King et al. chart is essentially one that is a restricted or compromised version of existing case-oriented methods. They argue, in effect, that some case-oriented practices are more productive than others and that researchers should utilize only the most productive (i.e., those that offer the greatest "analytic leverage"). Thus, the bridge they build from variable-oriented research to case-study research establishes a link only to a narrow subset of existing case-oriented methods.

The alternate vision of the middle path offered in this paper seeks a path that is not a compromise between variable-oriented and case-study research, but rather one that transcends many of the limitations of the other two paths. The goal of this middle path is not simply to link the two strategies, but to forge a new path. I turn now to a sketch of this path, which I call configurational comparative research.

Configurational Comparative Research: Its Essential Features

Table 2 adds a middle column to Table 1, representing the essential features of configurational comparative research (hereafter referred to as configurational research, for the sake of brevity). Of special importance is the fact that this alternate vision of the middle path, as shown in Table 2, combines the strengths of within-case and cross-case analysis. Unlike the case-study approach, configurational research attends to cross-case patterns; unlike variable-oriented research, it also attends to the specifics of each case and attempts cross-case analysis of within-case relationships.

This focus on combining within-case with cross-case analysis is reflected in the goals of configurational research (row 1 of Table 2). In configurational research the concern is not only to understand each case, but also to view cases in terms of their broad similarities and differences. Ultimately, the goal is to construct empirically grounded, theoretically relevant typologies of cases, advancing both general theoretical knowledge and understanding of the historically specific diversity of empirical cases.

Accordingly, the definition of the set of relevant cases and the differentiation among cases in terms of types (row 2) is relatively fluid at the outset of the research and becomes more fixed as the researcher learns more about cases and deepens his or her conceptualization of decisive similarities and differences. The researcher's construction of these boundaries, around all relevant cases and between sets of cases, is guided by his or her theoretical and substantive knowledge and interacts with that knowledge as it accumulates.

Research that seeks to combine within and cross-case analysis necessarily must rest on a sufficient number of cases (row 3), usually more than a handful. Cases are added to enrich the diversity of instances, which, in turn, promotes theoretical development and elaboration. However, there are limits to the enlargement of the number of cases. As the scope of an investigation broadens, the heterogeneity of cases can increase dramatically, which in turn dilutes the clarity and interpretive value of the concepts that structure the investigation. Of necessity, therefore, comparative researchers must guard against overextending the empirical scope of their investigations.

This concern for carefully delimiting the set of relevant cases (i.e., the domain) and for differentiating among cases according to their broad patterns of similarities and differences necessarily implies a style of research that has a distinctly grounded component (row 4). Existing theory provides important orienting concepts and leads but rarely offers enough specificity for the formulation of explicit hypotheses about historically situated cases. A theory this powerful and well-specified is not an impossibility. However, such a theory, of necessity, must be based on a firm foundation of in-depth knowledge of theoretically pertinent cases.

Configurational researchers are interested in large-scale historical outcomes that join and separate the fates of their cases (row 5). Sometimes they focus only on cases that experience the same outcome ("positive cases"); sometimes they contrast these positive cases with cases that seemed to be candidates for the outcome in question but nevertheless failed to exhibit it (i.e., "negative cases"); and sometimes they study divergent developments across a range of cases ("multiple outcomes"). The key is the focus on large-scale, historically emergent phenomena distinguishing sets of cases.

The explanation of historically emergent, large-scale phenomena generally calls for nuanced accounts of the conditions that combine to produce them (row 6). In this respect, configurational research is like case-study research: the

investigator builds an account of each case, focusing on intersections of actors, events, and structures. The key difference is that while building an account of each case, the configurational researchers also look across cases, to identify similarities and differences. A common finding is that the same outcome may be reached through several different paths, with a different combination of causally relevant conditions defining each path to the outcome.

A key feature of configurational research is that it attempts to balance within-case and cross-case analysis (row 7). The focus on cases as configurations facilitates this balance. In essence, each case is examined as a specific intersection of causally relevant conditions, and cases are compared with each other as configurations. By viewing cases as configurations it is possible to retain what is lost in variable-oriented research and in King et al.'s vision of the middle path--a way to represent and maintain the specificity and integrity of individual cases in the analytic shift from within-case analysis to cross-case analysis. Viewing cases in terms of configurations of causally relevant conditions, and not in terms of separable "independent variables," facilitates the construction of general knowledge from case-based knowledge.

Further Advantages of Configurational Comparative Analysis

One advantage of the configurational approach is its ability to address causal complexity. At the outset I noted that researchers tend to avoid investigations with a *moderate* number of cases (roughly 5 to 50--the middle section of the U-shaped relation depicted in Figure 1) because such studies seem to involve unmanageable complexity. When the number of cases is in this middle range, it is feasible for the researcher to learn about each case. At the same time, however, it is also very difficult for the researcher to keep track of and grasp their diversity. The difficulty of this task is one reason that many researchers opt for variable-oriented methods--if the N of cases is large enough to permit their use. However, embedded within the variable-oriented approach are routine assumptions about causation that directly contradict case-based understandings. Configurational methods solve this problem.

Consider, for example, the configurational strategies presented in Table 3. There are six strategies presented, based on the crosstabulation of whether the researcher seeks to model a single path to each outcome or allows multiple paths, and then whether the researcher is interested in (1) only the positive cases, (2) both positive and negative cases, or (3) divergent outcomes across a range of cases. All six strategies are configurational. The two simplest strategies, represented in cells

a and b, are the most common. One well-known instance of the strategy described in cell b is Theda's Skocpol's *States and Social Revolution*. She seeks to identify a single formula for social revolution across three positive cases and then argues that the formula was not satisfied in relevant negative cases. Strategies in cell c are not as common, but there are well-known exemplars, for example, Barrington Moore's *Social Origins of Dictatorship and Democracy*, which investigates three distinct outcomes: bourgeois democracy, fascism, and communism. Applications of the three strategies listed on the second row of Table 3 are still less common, but have become more numerous since the publication of *The Comparative Method*, which explains techniques for assessing multiple conjunctural causation. In essence, Table 3 answers the question: What kinds of research strategies should we use to build general knowledge from case-based knowledge?

Notice that there is an implicit N of cases dimension built into Table 3, running from cell a to cell f. The strategy described in cell a requires relatively few cases; the strategy described in cell f requires a substantial number. The principle embodied in this table is that as the N of cases increases, so should the allowance for causal complexity. This principle is the direct opposite of what conventionally happens when researchers enlarge their Ns--they fall back on methods that make radical simplifying assumptions about causation. These variable-oriented methods assume that causation is linear and additive and that the researcher's goal is simply to assess the relative importance of "independent variables" viewed in analytic isolation from each other. It is not possible with the variable-oriented approach to compare and contrast the different causal mixtures and combinations behind diverse empirical outcomes.

Recall that Figure 2 shows the frequency distribution for studies with different size Ns that would be manifest if in fact social scientists were to construct general knowledge more exclusively from case-based knowledge. In effect, Table 3 explains why it would have this shape (a negative, monotonic slope). With larger Ns, there must be more allowance for complexity, as the researcher examines a greater number of outcomes and considers the multiple routes to each.

Another advantage of the configurational approach is that its understanding of causal complexity is essentially set-theoretic in nature. This advantage is multifaceted; addressing its different aspects is beyond the scope of this essay (see Ragin 2000). However, an elemental contrast with variable-oriented methods is revealing. In most variable-oriented studies, the relevant causal variables are moderately correlated with each other and with the outcome variable. Paul Mehl

describes this as an inherent feature of social phenomena and argues that almost any pair of randomly selected variables of interest to social scientists will have a significantly nonzero correlation. The analysis of the net effects of causal variables on an outcome variable in such situations rarely yields decisive results. The usual finding is that there are many significant causal variables, each with a modest effect. In short, the very thing that often drives many researchers to assess net effects--the problem of correlated independent variables--also explains why the results of these efforts are often equivocal.

The reason most variables of interest to social scientists are at least modestly correlated is that case characteristics tend to lump together, that is, characteristics of cases tend to occur in interpretable combinations. This lumpiness is an essential feature of naturally occurring social phenomena. Thus, the calculation of a net effect is somewhat illusory. In the configurational approach the lumpiness of case characteristics is seen as an appropriate starting point for understanding cases, not as something that interferes with attempts to assess causation. Furthermore, the lumpiness of cases is quite easy to represent as the intersection of sets (i.e., in set-theoretic terms). It is difficult to represent lumpiness when the focus is on competing "independent" variables.

When cases are viewed as configurations (i.e., as intersections of sets), it no longer makes sense to ask questions about the net effects of independent variables. Instead, the important questions have a set-theoretic character. Thus, instead of asking: What is the net effect of variable x on the outcome variable across all cases? the researcher may ask: In what contexts and in what kinds of cases is x linked to the outcome in question? This latter question is set-theoretic in nature. It seeks to establish the contextual or enabling conditions for a link between a causal condition (x) and the outcome. When case characteristics are lumpy, as they are in most naturally occurring social phenomena, causal connections are often contingent on context--on the presence or absence of other causal conditions. accordingly, the analytic focus should be on the intersection of causally relevant conditions, which in turn is fundamentally a set-theoretic relation.

A simple example: In variable-oriented research using individual-level data, it is very easy to show that being married has a modest positive net effect on a variety of policy relevant outcomes, for example, avoiding poverty. In this U.S., this net effect holds for males and females and for Blacks and Whites. It is the sort of finding that has an "of course" quality to it, and it is rarely questioned. The policy implication of this robust net effect is that the state should encourage marriage as a blanket social policy. But is it true that marriage helps people avoid

poverty under all conditions? This interpretation of the effect of marriage is inherent in the calculation of its "net" effect.

From a configurational viewpoint, however, it is important to ask: In what contexts and in what kinds of cases is marriage linked to desirable policy outcomes? When the same policy data used to calculate the net effect of marriage on avoiding poverty are re-analyzed using configurational methods, the results show (1) that there are many different combinations of conditions linked to avoiding poverty, (2) that not all of these combinations involve marriage, and (3) that those that do involve marriage also involve important resources such as graduation from college or high-income parents. In other words, even though marriage is prevalent among those who successfully avoid poverty, it is a mistake to view this greater prevalence as an independent or "net" effect. The policy implication of the configurational analysis is that it is illusory to promote marriage as a blanket social policy.

What Kind of General Knowledge Do We Seek?

The configurational approach is a distinct path. It is not a compromise between case-study and variable-oriented research, but instead combines many of the strengths of the two approaches. This middle path maintains the integrity of cases as configurations and thus allows the construction of general knowledge from case-based knowledge. It does not supplant case-based knowledge with statements about the independent effects of causal variables.

General knowledge can come in many different forms. Variable-oriented methods, especially when focused on the net effects of independent variables, offer *one way* to represent general knowledge, but this way of representing knowledge is limited. Further, I have argued that the estimation of net effects may be somewhat illusory, especially given the tendency for naturally occurring social phenomena to occur in interpretable configurations of case characteristics. The estimation of net effects also may foster policy recommendations that are too crude, or at least insufficiently nuanced, leading to blanket policy recommendations.

When general knowledge is stated exclusively in terms of relationships between variables, cases disappear. To say that there is a relationship between the two variables or that one variable has a significant net effect on another says nothing about any case or any set of cases. The value of the configurational approach is that not only does it allow the construction of general knowledge from

case-based knowledge, but it also permits the reconnection of general knowledge to specific cases. When general knowledge is stated in terms of intersections of conditions and configurations of case characteristics, its relevance to specific cases and categories of cases is made explicit.

References will be furnished in a subsequent draft.

Table 1: Contrasts Between Case-Study and Variable-Oriented Research

	Case-Study Research	Variable-Oriented Research
1. Proximate goals	Case study researchers focus on the problem of making sense of a very small number of cases, usually one and rarely more than three, selected because they are substantively or theoretically important in some way. The key concern is the representation of the case.	Variable-oriented research seeks to document general cross-case relationships between variables characterizing a large population of generic observations. The key focus is on the relative conformity of cross-case relationships with theoretically based models.
2. Populations	The case-study researcher's answer to "What is my case a case of?" may change throughout the course of the investigation, as the investigator learns more about the phenomenon in question and refines his or her guiding concepts and analytic schemes. The fact that a single case can be defined in multiple ways is usually seen as a strength, making the case "rich."	In variable-oriented research, cases and populations are typically seen as given. The ideal-typic case (or "observation") is the survey respondent. Macrolevel cases such as countries are treated in the same generic manner. The key issue is how to derive a representative sample from the abundant supply of "given" observations.
3. N of cases	Case-study research is often defined by its focus on phenomena that are of interest because they are rare-- that is, often an <u>N</u> of only one. Empirical depth is more important than breadth; therefore, enlarging the <u>N</u> is typically viewed as hazardous. Comparability of cases is never assumed and usually viewed as limited at best.	Variable-oriented researchers are encouraged to enlarge their number of cases whenever possible; more is always better. With more cases, researchers can make more precise estimates of the strength of the connections among variables. The individuality of each case is relegated to the error vector, giving the researcher a distilled representation of what is general across cases.
4. Role of	Case-study researchers use in-depth	In variable-oriented research, it is

Theory	study of cases to advance theory. Thus, they often choose cases that are anomalous in some way from the viewpoint of current theory. A case study is successful even if it succeeds in showing only that existing theory is inadequate. Thus, case selection is critically important.	often presumed that researchers have well-defined theories and well-formulated hypotheses at their disposal from the very outset of their research. Theory testing is the centerpiece of social research. The ideal variable-oriented investigation adjudicates between competing theories.
5. Conception of outcomes	Case-study researchers often select cases specifically because of their uncommon or anomalous outcomes. The usual goal is to resolve the anomaly in a theoretically progressive way, based on in-depth knowledge of the selected case(s). Often there is no sharp separation of causal conditions and outcomes, for an outcome may seem inherent in the constitution of the case.	Variable-oriented researchers are advised to direct their attention to "dependent variables" that display a healthy range of variation across a systematic sample of cases drawn from a large population. Usually, the more fine-grained this variation, the better. Outcomes that do not vary across cases cannot be studied because there is no variation to explain.
6. Understanding of causation	Case-study researchers examine causation holistically, in terms of a convergence of structures, actors, and events. They are also centrally concerned with sequences and timing of events, with an eye toward turning points and path dependence.	Variable-oriented researchers assess the relative importance of competing independent variables in order to test theory. The key focus is on the relative importance of causal variables across cases, not on how they come together or combine in any single case. A single causal model is derived that applies equally to all cases.
7. Within versus cross-case analysis	Case-study research is focused almost entirely on within-case patterns. Researchers examine parts of the case as mutually constitutive of each other and the whole they form together. Case-study researchers often ask: 'What kind of whole has parts like this?' as they explore connections among case aspects.	Variable-oriented researchers give priority to cross-case patterns. Aspects of cases are viewed primarily in terms of how they vary and co-vary across cases. How aspects of cases connect within each case is more or less ignored. The idiosyncracies of cases cancel each other out, as deviations from general patterns are assigned to the error vector of probabilistic models.

Table 2: Configurational Comparative Research

	Case-Study Research	Configurational Comparative Research	Variable-Oriented Research
1. Proximate goals	Case study researchers focus on the problem of making sense of a very small number of cases, usually one and rarely more than three, selected because they are substantively or theoretically important in some way. The key concern is the representation of the case.	Comparative researchers study substantively or theoretically defined categories of cases (usually five to 50 or more), with the goal making sense of both individual cases and clusters of similar cases in the light of knowledge of cross-case patterns, and vice versa.	Variable-oriented research seeks to document general cross-case relationships between variables characterizing a large population of generic observations. The key focus is on the relative conformity of cross-case relationships with theoretically based models.
2. Populations	The case-study researcher's answer to "What is my case a case of?" may change throughout the course of the investigation, as the investigator learns more about the phenomenon in question and refines his or her guiding concepts and analytic schemes. The fact that a single case can be defined in multiple ways is usually seen as a strength, making the case "rich."	In comparative research, the investigator constructs a carefully delimited set of cases, using theoretical and substantive knowledge as guides. The boundary around this set is initially flexible and becomes more fixed as the research proceeds, through the interaction of ideas and evidence. Concept formation and empirical categorization go hand-in-hand.	In variable-oriented research, cases and populations are typically seen as given. The ideal-typic case (or "observation") is the survey respondent. Macrolevel cases such as countries are treated in the same generic manner. The key issue is how to derive a representative sample from the abundant supply of "given" observations.
3. N of cases	Case-study research is often defined by its focus on phenomena	Comparative researchers often make strategic comparisons and thus need	Variable-oriented researchers are encouraged to enlarge

	<p>that are of interest because they are rare—that is, often an <u>N</u> of only one. Empirical depth is more important than breadth; therefore, enlarging the <u>N</u> is typically viewed as hazardous. Comparability of cases is never assumed and usually viewed as limited at best.</p>	<p>diverse cases. At the same time, they need to maintain case homogeneity because their cases should all be instances of or candidates for the same outcomes. Thus, comparative researchers must balance conflicting pressures when delimiting the set of relevant cases.</p>	<p>their number of cases whenever possible; more is always better. With more cases, researchers can make more precise estimates of the strength of the connections among variables. The individuality of each case is relegated to the error vector, giving the researcher a distilled representation of what is general across cases.</p>
4. Role of Theory	<p>Case-study researchers use in-depth study of cases to advance theory. Thus, they often choose cases that are anomalous in some way from the viewpoint of current theory. A case study is successful even if it succeeds in showing only that existing theory is inadequate. Thus, case selection is critically important.</p>	<p>Existing theory is rarely well-formulated enough to provide explicit hypotheses in comparative research. The primary theoretical objective of comparative research is not theory testing, but concept formation, elaboration, and refinement, and also theory development. Sharpening the definition of the set of relevant cases is often an important theoretical advance in itself.</p>	<p>In variable-oriented research, it is often presumed that researchers have well-defined theories and well-formulated hypotheses at their disposal from the very outset of their research. Theory testing is the centerpiece of social research. The ideal variable-oriented investigation adjudicates between competing theories.</p>
5. Conception of outcomes	<p>Case-study researchers often select cases specifically because of their uncommon or anomalous outcomes. The usual goal is to resolve the anomaly in a theoretically</p>	<p>Comparative researchers often begin by intentionally selecting cases that do not differ greatly from each other with respect to the outcome that is being</p>	<p>Variable-oriented researchers are advised to direct their attention to "dependent variables" that display a healthy range of variation across a systematic</p>

	<p>progressive way, based on in-depth knowledge of the selected case(s). Often there is no sharp separation of causal conditions and outcomes, for an outcome may seem inherent in the constitution of the case.</p>	<p>investigated; they are all "positive cases." The constitution and analysis of the positive cases is usually a prerequisite for the specification of relevant negative cases--if they can be reasonably identified.</p>	<p>sample of cases drawn from a large population. Usually, the more fine-grained this variation, the better. Outcomes that do not vary across cases cannot be studied because there is no variation to explain.</p>
<p>6. Understanding of causation</p>	<p>Case-study researchers examine causation holistically, in terms of a convergence of structures, actors, and events. They are also centrally concerned with sequences and timing of events, with an eye toward turning points and path dependence.</p>	<p>Comparative researchers usually look at causation in terms of multiple pathways. Positive cases often can be classified according to the general path each traveled to reach the outcome. Each path, in turn, can be seen as involving a different combination of relevant causal conditions.</p>	<p>Variable-oriented researchers assess the relative importance of competing independent variables in order to test theory. The key focus is on the relative importance of causal variables across cases, not on how they come together or combine in any single case. A single causal model is derived that applies equally to all cases.</p>
<p>7. Within versus cross-case analysis</p>	<p>Case-study research is focused almost entirely on within-case patterns. Researchers examine parts of the case as mutually constitutive of each other and the whole they form together. Case-study researchers often ask: 'What kind of whole has parts like this?' as they explore connections among case aspects.</p>	<p>Comparative researchers focus on configurations of causally relevant characteristics of cases, with the goal of determining how relevant aspects fit together. They use cross-case analysis to strengthen and deepen within-case analysis, and vice versa. To the extent possible, comparative researchers try to balance cross-case and</p>	<p>Variable-oriented researchers give priority to cross-case patterns. Aspects of cases are viewed primarily in terms of how they vary and co-vary across cases. How aspects of cases connect within each case is more or less ignored. The idiosyncracies of cases cancel each other out, as deviations from</p>

		within-case analysis.	general patterns are assigned to the error vector of probabilistic models.
--	--	-----------------------	--

Table 3: Six Configurational Research Strategies

	No Negative Cases	Negative (or null) Cases	Multiple Outcomes
Conjunctural Causation	a. Identify a single combination of causal conditions shared by instances of the outcome (positive cases only)	b. Identify a single combination of causal conditions shared by positive cases; verify that it is not found among relevant negative cases (candidates for the outcome)	c. Classify cases according to the outcome each exhibits; identify a single combination of causal conditions specific to each outcome.
Multiple-Conjunctural Causation	d. Identify the different combination of causal conditions linked to a single outcome (positive cases only)	e. Identify different combinations of causal conditions exhibited by positive cases; verify that these combinations are not found among relevant negative cases (candidates for the outcome)	f. Classify cases according to the outcome each exhibits; identify the different combinations of causal conditions specific to each outcome.