

## INTRODUCTION

Social scientists generally stay away from anything labeled “fuzzy” because their work is so often described this way by others, especially by scholars in the “hard” sciences. My initial title for this book, *Fuzzy Social Science*, made so many of my colleagues cringe that I felt compelled to change it so that the adjective “fuzzy” applied to sets, not to social science. Today, however, “fuzzy” has a new usage that is unrelated to its original meaning of “unclear” or “muddled” (Kosko 1993; Grint 1997). When applied to a set, it signals that relevant objects can have varying degrees of membership in the set (Zadeh 1965).<sup>1</sup> For example, people can have varying degrees of membership in the set “tall”; investments can vary in their degree of membership in the set “low-risk.” Fuzzy sets are especially useful for categories that are imprecise, like “tall” or complex in construction, like “low-risk.”

Today, fuzzy sets are used in many different fields and technical arenas to address a variety of questions and problems, both mundane and abstract (Ross 1995). The range of applications is ever-expanding, reaching from artificial intelligence to washing machines and the stock market. Much of the literature on fuzzy sets is concerned with the problem of control—how to develop machines that “act smart” in the face of ambiguity or complexity (Dubois, Prade, and Yager 1993). Often these applications involve programming machines with expert human knowledge. Consider, for example, the problem of automating a tractor-trailer so that it can back up to a loading dock under a variety of different conditions. Much of the human knowledge that would

1. Kaplan and Schott (1951) presented many of the basic ideas of the approach well in advance of its unveiling as fuzzy logic in the mid-1960s.

go into this task comes in the form of fuzzy sets describing the various conditions an automated tractor-trailer might encounter (e.g., “slippery” pavement). Applications of this type have received a great deal of attention, and for most of the reading public “fuzzy sets” means “smart machines.”

Like others, I am fascinated by these applications. However, these uses of fuzzy sets do not strike close to home for social scientists, and their impact on the social sciences has been slight (Smithson 1988). The terminology of fuzzy sets does crop up occasionally, and a few scholars have attempted to apply some of the ideas (e.g., Smithson 1989; Zetenyi 1988), but for the most part scholars have not recognized the potential of fuzzy sets for transforming social science methodology. Part of the problem is that scholars have attempted to integrate fuzzy sets into their existing repertoires of methods without altering their usual practices in any way. In fact, however, to appreciate the power of fuzzy sets it is necessary to adopt a broader understanding of data analysis and its place in the process of social research.

Fuzzy sets offer researchers an interpretive algebra, a language that is half-verbal-conceptual and half-mathematical-analytical. Thus, the greatest value of fuzzy sets for social scientists is their potential for enlivening, intensifying, and extending the “dialogue” between ideas and evidence in social research. This dialogue is the systematic interplay between theory and data analysis that occurs in most studies—the back-and-forth between the use of data to improve theory and the use of theory to guide the exploration of data. Most theoretical arguments, as verbal formulations, deal with set-theoretic relationships. Because fuzzy sets also address set-theoretic relationships, they offer the opportunity for creating a very close correspondence between theory and data analysis. In short, with fuzzy sets researchers can analyze evidence in ways that directly reflect their theoretical arguments. Thus, as I note in the preface, fuzzy sets have the potential to transform research that is oriented toward “discovery,” toward gaining new insights about the world (Diesing 1971). My primary goal in this book is to show how to use fuzzy sets to strengthen discovery in the social sciences.

While most social science research involves discovery in some way, this type of inquiry is faltering in many areas of social science. I believe that the main problem is the dominance of “conventional” forms of quantitative analysis.<sup>2</sup> While there is nothing wrong with quantification

2. I use the qualifier “conventional” in a sociological manner. A particular way of analyzing social data based on core notions about cases and variables has become

and the social sciences certainly need analytic rigor, quantitative analysis—especially as it is usually practiced—often constrains the dialogue between ideas and evidence in unproductive ways (Richters 1997). For example, these methods often assume that cases are homogeneous (e.g., in how causal conditions operate), when researchers should be trying to pinpoint and understand their heterogeneity. For the most part, these limitations are overlooked or seen as strengths because they are core elements of the theory-testing template that structures much social science discourse. As I show in this book, however, it is possible to conduct a far richer dialogue between ideas and evidence using fuzzy sets.

My argument has three main pillars:

First, I argue that social scientists interested in discovery must relinquish many of the “homogenizing assumptions” that undergird conventional quantitative analysis. These homogenizing assumptions structure how social scientists view populations, cases, and causes and thus constrain the dialogue between ideas and evidence in ways that limit discovery. I argue that researchers should instead focus on “diversity” using analytic strategies that are more common in qualitative inquiry. These strategies are easy to implement when the number of cases is small—the usual situation in qualitative inquiry. However, they are rarely used when *Ns* are large because of analytic difficulties (Ragin 1997). This book, especially part 1, formalizes and extends these “diversity-oriented” techniques to investigations involving large *Ns*.

Second, I argue that it is possible to use fuzzy sets to extend and deepen diversity-oriented research strategies. In a nutshell, diversity-oriented research attends to heterogeneity and difference, especially to differences in kind, using a “configurational approach” to social phenomena—viewing cases as specific configurations of aspects and features (i.e., configurations of set memberships). This approach searches for heterogeneity within “given” or preconstituted populations and conceives of “difference” in terms of kinds and types of cases, replacing the conventional view of difference as variation (i.e., as deviation from the mean). Fuzzy sets augment the configurational approach by allowing degrees of membership in types and kinds. Thus, the incorporation of fuzzy sets allows for “variation” without forsaking the core emphasis on types and kinds of cases.

Third, and most important of all, I argue that the link between theory and data analysis in the social sciences can be greatly improved using

dominant in sociology, political science, and related disciplines over the past several decades. The dominance of this way of analyzing data has conventionalized it.

fuzzy sets for the simple reason that fuzzy sets can be carefully tailored to fit theoretical concepts. Before explaining this position, I briefly describe, for those new to the topic, exactly what a fuzzy set is.

### *Fuzzy Sets*

A conventional (or "crisp") set is dichotomous: An object (e.g., a survey respondent) is either "in" or "out" of a set, for example, the set of Protestants. Thus, a conventional set is comparable to a binary variable with two values, 1 ("in," i.e., Protestant) and 0 ("out," i.e., non-Protestant). A fuzzy set, by contrast, permits membership in the interval between 0 and 1 while retaining the two qualitative states of full membership and full nonmembership. Thus, the fuzzy set of Protestants could include individuals who are "fully in" the set (fuzzy membership = 1.0), some who are "almost fully in" the set (membership = .90), some who are neither "more in" nor "more out" of the set (membership = .5, also known as the "crossover point"), some who are "barely more out than in" the set (membership = .45), and so on down to those who are "fully out" of the set (membership = 0). It is up to the researcher to specify procedures for assigning fuzzy membership scores to cases, and these procedures must be both open and explicit so that they can be evaluated by other scholars.

While it might appear to most social scientists that a fuzzy set is merely the transformation of a binary variable into a continuous variable, this understanding is not correct. Indeed, this common misperception of fuzzy sets may explain why social scientists have been so slow to grasp their analytic power and significance. A fuzzy set is much more than a "continuous" variable because it is much more heavily infused with theoretical and substantive knowledge. Despite the adjective "fuzzy," compared with the conventional variable, a fuzzy set is more empirically grounded and more precise.<sup>3</sup>

To grasp the critical difference between fuzzy sets and continuous variables, consider the contrast between a conventional measure of degree of Protestantism, applied to individuals, and the fuzzy set of Protestants. Imagine that the conventional measure is based on a variety of

3. Please note that there is an important distinction between fuzzy concepts and fuzzy membership (Lakoff 1973, 1987). A concept may be fuzzy in the more conventional sense of unclear, and membership in the set implied by the concept may be fuzzy because the concept is fuzzy. In this work, however, I am concerned with concepts that are relatively clear in meaning, at least to the researchers who use them, and the degree of membership of relevant objects in the sets implied by such concepts.

indicators of Protestant behaviors, attitudes, and beliefs and that these different indicators strongly correlate with each other, justifying their combination into a single index of degree of Protestantism. Assume further that this scale is both valid and reliable. But where on this scale is a full-fledged Protestant? Where on this scale is a full-fledged non-Protestant? Where on this scale is the cut-off value (or values) separating those who are more in the set of Protestants from those who are more in the set of non-Protestants?

To answer these questions, the researcher needs not only an index of Protestantism, the fine-grained measure just described, but also a good base of substantive knowledge about Protestantism and a solid grasp of its theoretical relevance—why degree of membership in the set of Protestants matters and how it should be assessed. Without this infusion of theoretical and substantive knowledge, the fine-grained measure of Protestantism remains vague and imprecise—*uncalibrated*. From a fuzzy-set perspective, the conventional variable's great capacity for making fine-grained distinctions with respect to relative levels must be grounded in knowledge; otherwise, its precision is wasted.

In the hands of a social scientist, therefore, a fuzzy set can be seen as a fine-grained, continuous measure that has been carefully calibrated using substantive and theoretical knowledge relevant to set membership. This infusion of knowledge transforms rankings that are almost entirely relative in nature (e.g., degree of Protestantism) to ones that show degree of membership in a well-defined set (e.g., degree of membership in the set of Protestants). As I show subsequently, this infusion of knowledge often redefines portions of the range of a conventional continuous variable as irrelevant. For example, the range of variation in the Protestantism index above a certain value may be irrelevant to the fuzzy set of Protestants because these scores may all signal full membership in the set—scores of 1.0. Likewise, the range of variation in this index below a certain value also may be irrelevant to the fuzzy set of Protestants because these scores may all signal full nonmembership in the set—scores of 0 (for a detailed discussion of these principles, see chapter 6).

### *Fuzzy Sets in the Dialogue Between Ideas and Evidence*

Fuzzy sets are especially valuable in the back-and-forth between theory and data analysis precisely because they are heavily infused with theoretical and substantive knowledge. This infusion of knowledge makes it possible to tailor fuzzy sets to theoretical concepts. This book elaborates various uses of fuzzy sets as interpretive tools, showing how they trans-

form and enrich the dialogue between ideas and evidence. For illustration consider the following features of fuzzy sets:

*Fuzzy sets combine qualitative and quantitative assessment in a single instrument.* All fuzzy sets consist of two qualitative states, full membership and full nonmembership, and all the quantitative variation that exists between these two qualitative states. Further, fuzzy sets distinguish between objects that are “more in” versus “more out” using the crossover point (.5). For example, the fuzzy set of “religious fundamentalists” has four main parts: those who are fully out of the set (fuzzy membership = 0), those who are not fully out of the set but still more out than in (fuzzy membership > 0 but < .5), those who are more in than out of the set but still not fully in (fuzzy membership > .5 but < 1.0), and those who are fully in the set (fuzzy membership = 1.0).<sup>4</sup>

This integration of qualitative and quantitative assessment in a single instrument gives fuzzy sets an important advantage over conventional variables in the dialogue between ideas and evidence. Consider the proposition that religious fundamentalists are politically conservative. In conventional social science, this statement has two main interpretations with respect to the meaning of religious fundamentalists. The first is that being a religious fundamentalist (a qualitative state) increases either the probability of being politically conservative or the level of political conservatism. The second is that the level of religious fundamentalism (a continuous variable) correlates either with the probability of being a political conservative or with the level of political conservatism.<sup>5</sup> In the first formulation there are two qualitative conditions: one embracing all religious fundamentalists, the other embracing everyone else. In the second formulation, there are many different levels of religious fundamentalism defined primarily in relation to each other (i.e., above versus below the average level). Both ways of interpreting social science concepts are considered valid. Typically, researchers simply choose the formulation that seems most sensible given the question at hand.

With fuzzy sets, however, researchers are not forced to choose be-

4. Cases also may reside at the crossover point (fuzzy membership = .5), the point of maximum ambiguity in whether a case is more in or more out of a set. Thus, this location could be considered an important “fifth” main part of a fuzzy set. (See my discussion of “five-value” fuzzy sets in chapter 6.)

5. Social scientists usually prefer measures with fine-grained distinctions. Thus, the favored formulation would be to treat both religious fundamentalism and political conservatism as variables with many different levels and to correlate them as interval-scale measures.

tween strictly binary and completely relative formulations because fuzzy sets embody both qualitative states (i.e., full membership and full nonmembership) and variation by level (i.e., degrees of membership between 0 and 1) in a single instrument. Thus, using fuzzy sets it is possible to operationalize multiple interpretations of a concept. For example, the proposition about religious fundamentalists (that they are politically conservative) could apply to all individuals displaying some degree of religious fundamentalism, even very minimal levels (i.e., fuzzy membership > 0). The expectation would be that there should be some pattern of correspondence between degree of membership in the set of fundamentalists and the outcome (political conservatism) observable even for those with weak membership in the set of religious fundamentalists. Alternatively, the statement could apply only to individuals who are more “in” the set of religious fundamentalists than “out” (i.e., those with fuzzy membership > .5). The expectation here would be that some pattern of correspondence between fundamentalism and conservatism should be visible for those individuals with membership scores exceeding .5. Or, the statement might apply only to individuals who are full-fledged religious fundamentalists, with a pattern of correspondence observable only for individuals with fundamentalism scores of 1.0. Fuzzy sets permit the investigation of all three interpretations simultaneously.

Take another example: Some political scientists have argued in various ways that “democratic countries don’t make war with other democratic countries” (see Russett 1993). But countries vary in the degree to which they are democratic (Bollen 1993)—in the degree to which they belong in the set of democratic countries. Does the statement apply, in varying degrees, to all countries with at least some presence or level of democracy (for example, a dictatorship that permits political parties)? Or does it apply only to countries that are more democratic than not? Or does it apply only to full-fledged democracies? With fuzzy sets it is possible to address these varied interpretations of the proposition with a single instrument and to draw conclusions that are more nuanced than is possible with conventional variables.

Thus, by combining qualitative and quantitative assessment in a single instrument, fuzzy sets make it possible for researchers to address varied interpretations of social scientific concepts in an explicit manner. Almost all concepts refer simultaneously to qualitative and quantitative differences. Concepts are typically expressed qualitatively in theoretical discourse (e.g., the statement about democratic countries above) but also allow for varying degrees (e.g., variation in degree of democracy).

These two aspects of social scientific concepts often spawn great controversies, with social researchers talking past each other. For example, a quantitative analyst might document a cross-national relationship, using data on many countries, between degree of "democracy" and some other variable, and feel confident about its breadth. A qualitative analyst might scoff at this analysis because it includes a vast majority of undemocratic countries. Fuzzy sets remedy this type of analytic chaos by joining qualitative and quantitative assessment, forcing researchers to distinguish between qualitative and quantitative aspects when they operationalize their concepts. Thus, researchers who use fuzzy sets can establish a very close correspondence between the content of the concepts appearing in theoretical statements and propositions, on the one hand, and the analysis of empirical evidence, on the other.

*Fuzzy sets provide tools for the assessment of set-theoretic relationships.* Set-theoretic relationships are implicit in most social science discourse. To state, for example, that "religious fundamentalists are politically conservative," is not to say that the only path to political conservatism is through religious fundamentalism. Rather, religious fundamentalism is viewed as one of several possible routes to political conservatism, which certainly allows for the possibility that most political conservatives are not religious fundamentalists. While this set-theoretic relationship might not be perfect, it would not be unreasonable to expect religious fundamentalists to be a rough subset of political conservatives. Essentially, the expectation here is that most who are religious fundamentalists are also politically conservative, even though most who are politically conservative may not be religious fundamentalists. Thus, in this example, the variable treated as causal (religious fundamentalism) could be seen as an imperfect subset of the variable portrayed as the outcome (political conservatism).

Consider another rough subset relationship. Most professionals have advanced degrees, even though not all those with advanced degrees are professionals (many with advanced degrees drive taxis; others own their own businesses; some are full-time parents, and so on). Thus, professionals are a rough subset of people with advanced degrees. In this example, the variable that is treated as the outcome, having a career as a professional, is a rough subset of the variable that is seen as causal, holding an advanced degree. In other words, the set-theoretic relationship in this example is the reverse of what it is in the previous example, at least from a causal point of view.

At the level of discourse, the two statements are parallel in structure.

We state the subset first. (1) "Religious fundamentalists are politically conservative." (2) "Professionals have advanced degrees." However, in causal terms the two statements are the reverse of each other. The causal condition comes first in the first statement and second in the second statement. In conventional forms of quantitative analysis, there is no simple way to take account of these different set-theoretic relationships. Most social scientists would assess both statements by correlating the variables in question. If religious fundamentalism and political conservatism correlate, this evidence would be cited as support for the causal argument implicit in the first statement. If possessing an advanced degree and having a professional career correlate, this evidence would be cited as support for the causal argument implicit in the second statement. In short, the fact that rough set-theoretic relationships are clearly embedded in the two statements, along with the fact that these two set-theoretic relationships are the opposite of each other, at least from a causal point of view, would be ignored by practitioners of conventional methods.

The issue of set-theoretic relationships is central to the link between theoretical argumentation and empirical analysis and therefore is not something that social scientists should ignore. Note that the two statements presented above involve very different kinds of causal reasoning (see Mackie 1974, 1985). In the first statement, the causal condition (religious fundamentalism) is one of several possible routes or paths to the outcome (political conservatism). It may not be one of the more important routes; it is simply a route. In the second statement, having an advanced degree is a prerequisite for most professional careers, a virtual necessary condition. Having an advanced degree is not one of many possible paths to a career as a professional; it is the major hurdle on the (almost) only route. These different causal structures account for the opposite set-theoretic relationships embedded in the two statements, with the causal condition the subset in one statement and the outcome the subset in the other.

Conventional methods (e.g., correlation) and conventional measures (e.g., interval- and ratio-scale variables) cannot be used to address set-theoretic relationships. Indeed, such relationships are completely outside their scope. Yet, set-theoretic relationships are the bread and butter of theoretical argumentation and discourse; they are embedded or implied in most theoretical statements. As I show in this book, it is possible to address set-theoretic relationships with fuzzy sets. For example, it is possible to assess whether the set of religious fundamentalists is a

rough subset of the set of political conservatives, using measures that attend to fine-grained differences in both religious fundamentalism and political conservatism.

The features of fuzzy sets just discussed are only two of their many aspects that make them especially powerful for linking ideas and evidence in social research. I explore these features in part 2 of this book, after first laying the proper foundation in the diversity-oriented approach.

#### *Fuzzy Sets and Diversity-Oriented Research*

Fuzzy sets have much to offer all of social science. However, their value is greatly enhanced when used in tandem with the diversity-oriented approach. This approach brings many of the special features of qualitative analysis, including an interest in set-theoretic relationships, to research involving large *N*s. Indeed, the pairing of these two—fuzzy sets and the diversity-oriented approach—is central to this book. Without the frame of the diversity-oriented approach, the value of fuzzy sets to social science is limited. Most conventional data-analytic techniques use linear algebra and thus do not address set-theoretic relationships. Likewise, without fuzzy sets, the scope of social phenomena that is amenable to the diversity-oriented approach, with its emphasis on configurations of set memberships, is restricted to categorical social data. In short, the key link between fuzzy sets and the diversity-oriented approach is that they both deal with sets and set-theoretic relationships.

Imagine the response of a conventional quantitative researcher to the two features of fuzzy sets just discussed. The embedding of qualitative states (i.e., full membership and full nonmembership) on continua looks like a bad idea. Why compress the two ends of a continuous variable, converting values at one end to scores of 1 (full membership) and values at the other end to scores of 0 (full nonmembership)? From the conventional viewpoint it is almost always a mistake to throw away variation. Linear models work best when the extremes of the distribution of a continuous variable are well populated with cases (Chatterjee and Price 1991). Likewise, the conventional quantitative researcher has little use for set-theoretic relations. In the conventional view, the proximate goal of quantitative analysis is to explain variation in one or more dependent variables. It does not matter whether an independent variable is a rough subset of the outcome (e.g., religious fundamentalism and political conservatism) or the reverse (e.g., holding an advanced degree and having a career as a professional). If either of these rela-

tionships exists, then the causal condition should correlate with the outcome.

The diversity-oriented approach, by contrast, focuses directly on set-theoretic relationships. Full elaboration of this approach is the goal of part 1 of this book; here I mention a few of its main features. At its core, this approach is concerned with exploring and constructing types and kinds of cases, with a special concern for causal heterogeneity—the different paths to an outcome. Thus, this approach builds on and extends the techniques I developed in *The Comparative Method* (1987). In the diversity-oriented view, what may appear to be a single population may be packed with many different types, subtypes, and mixed types of cases. Types are most visible when cases are viewed not as collections of analytically distinct variables but as configurations of set memberships. Once cases are understood in terms of their configurations of set memberships, it is possible to examine their heterogeneity, especially with respect to the different combinations of causal conditions linked to an outcome. For example, a researcher might study the different configurations of background characteristics linked to political conservatism without assuming that the same causal conditions operate in the same way in all cases. For some types of cases “rural background,” for instance, might contribute to political conservatism; for others, it might work against it. In the diversity-oriented view, the important point is not that causal conditions may have opposite effects in different contexts, but that the researcher must view cases as configurations—as combinations of aspects and conditions.

With categorical data, it is relatively easy to see cases as configurations of set memberships because of the straightforward correspondence between categories and sets (Ragin 1987). Using ordinal, interval, and ratio-scale data, however, this one-to-one correspondence evaporates, and it becomes more difficult to view cases as belonging to types, that is, to view them as instances of configurations. Without the aid of fuzzy sets, the diversity-oriented approach is limited to social data that are wholly categorical in nature. Using fuzzy sets, however, it is possible to address all kinds of data and all aspects of cases, even those involving fine-grained differences in degree or level. Thus, fuzzy sets provide a way to extend the view of cases as configurations of set memberships, first presented in *The Comparative Method*, using many different kinds of evidence.

While the use of fuzzy sets is “detachable” from diversity-oriented research and vice versa, the two work best hand in hand. The bond

that joins them is their common focus on sets and the analysis of set-theoretic relationships. This book shows how the diversity-oriented approach amplifies the social scientific relevance of fuzzy sets and how fuzzy sets amplify the range of the diversity-oriented approach.

#### *Plan of the Book*

Part 1 of *Fuzzy-Set Social Science* describes the central features of diversity-oriented research and culminates with an example of the approach. As I present this approach, I also discuss aspects of conventional quantitative analysis that interfere with its use as a discovery tool: (1) its dependence on populations that are constituted prior to data collection and analysis; (2) its heavily variable-oriented discourse, a framework that is antithetical to the analysis of cases as interpretable configurations of aspects, and (3) its additive-linear view of causation, an understanding that depends upon strong homogenizing assumptions about cases, which, in turn, make this approach insensitive to causal complexity.

The foundation for my critique of conventional quantitative analysis is provided by qualitative, case-oriented social science. In this approach, researchers conceive populations as meaningful sets of cases that often must be formed and conceptualized in the course of an investigation. They view cases as configurations of aspects and seek to understand them at the level of the specific instance. They see causation as conjunctural and plural—causes may combine in different and sometimes contradictory ways to generate the same outcome. I present diversity-oriented research as a middle path between qualitative, case-oriented research and quantitative, variable-oriented research. In some respects, part 1 of *Fuzzy-Set Social Science* can be seen as a rebuttal of King, Keohane, and Verba's treatise on methodology, *Designing Social Inquiry* (1994). They argue, in essence, that the way to improve qualitative research is to make it as much like quantitative research as possible. I argue the opposite, namely, that many of the special features of qualitative research should be integrated into quantitative research.

Chapter 1 situates diversity-oriented research between the quantitative analysis of patterns observable across many cases (i.e., conventional variable-oriented research) and the qualitative, in-depth study of small Ns (i.e., case-oriented research). I offer a general discussion of these two strategies and of the gulf that separates them, arguing that diversity-oriented research bridges this gulf. Instead of giving priority to either cases or variables, diversity-oriented research emphasizes "types" and "kinds," a formulation that views cases configurationally, as combinations of aspects.

Chapters 2 through 4 address the three main pillars of quantitative social science—populations, variables, and causation. I focus on the contrasts between conventional quantitative methods and the diversity-oriented approach. For populations, the contrast is between the conventional understanding of populations as given or "preconstituted" and the diversity-oriented view of populations as categories that must be constituted and refined in the course of an investigation. For variables, the main contrast is between the conventional understanding of variables as analytically distinct aspects of cases and the diversity-oriented view of cases as configurations of aspects. This discussion builds on Paul Lazarsfeld's (1937) concept of "property spaces" and their use in the construction of types. For causation, the main contrast is between the conventional view of causation as a contest between independent variables to explain variation in an outcome and the diversity-oriented view that causation is often both conjunctural and multiple. In the conventional view, each single causal condition, conceived as an analytically distinct variable, has an independent impact on the outcome. In the diversity-oriented view, causes combine in different and sometimes contradictory ways to produce the same outcome, revealing different paths.

Chapter 5 formalizes the diversity-oriented approach with an application. The key focus of the application is the problem of studying causal complexity—how to use information on configurations of set memberships to assess causation that is both conjunctural and multiple. Causation this complex is difficult to assess. I show that by testing the sufficiency of all possible combinations of conditions drawn from the researcher's property space it is possible to address causal complexity in a way that shows the different paths to a single outcome.

While part 1 synthesizes much of my recent work on methodology, it is important to note that this presentation incorporates a number of important advances. First, it offers a clear exposition of the concept of property space and shows the links between property spaces and "truth tables," the centerpiece of the techniques I present in *The Comparative Method* (1987; see also Becker 1998). Second, it establishes the centrality of the analysis of both set-theoretic relationships and necessary and sufficient causation to social scientific inquiry. Third, it demonstrates that the assessment of necessity and sufficiency need not be "all or nothing," but instead may be probabilistic and partial in nature. Last, it shows that even when there is great causal complexity (defined as situations in which no single cause is either necessary or sufficient), the investigation of combinations of causal conditions is still possible.

While dramatic in scope, the methodology presented in part 1 is limited in the way it views and represents diversity. In conventional set theory, a case is either in or out of a set (e.g., the set of males, the set of democratic countries, the set of not-for-profit organizations, and so on); a case cannot be partially in or partially out. Everyday experience indicates that this “in-or-out” conception of set membership is too restrictive; membership in sets is often partial. Thus, while a great deal of diversity is revealed when researchers examine configurations of set memberships, a lot is hidden when membership in sets is constrained to be crisp—either in or out. In fact, the membership of most social phenomena in the sets social scientists use to characterize them is usually fuzzy, not crisp. That is, cases (e.g., countries) exhibit varying degrees of membership in categories (e.g., the set of democracies).

Part 2 shows how to use fuzzy sets to enrich social research, building on the foundation of the diversity-oriented approach presented in part 1. A central argument of part 2 is that diversity has two main aspects. The first aspect of diversity is that which is based on “differences in kind” arising from categorical distinctions. The second aspect of diversity involves the varying degrees of membership that instances may exhibit in the categories and types used to characterize them. Conventionally, social scientists think of the first kind of diversity as nominal-scale and the second kind as ordinal, interval, or ratio-scale. I demonstrate in part 2 that fuzzy sets capture both kinds of difference simultaneously—in the same analytic breath—and that for this reason they offer especially powerful analytic tools for social scientists. Most social scientific concepts invoke qualitative and quantitative distinctions at the same time. This fact can be seen in everyday sociological concepts like “rich,” which invoke a categorical distinction but also permit degree of membership, and in sophisticated ideal-typic notions such as “bureaucratic.” The key to understanding the power of fuzzy sets is to see that it is possible to specify qualitative breakpoints on continua and to incorporate these qualitative breakpoints directly into the analysis of evidence that varies by level. With fuzzy sets, researchers can analyze set-theoretic relationships while still attending to phenomena that vary by level or degree. They do not have to forfeit the study of variation by level in order to study cases as configurations or to explore causal complexity.

My presentation of fuzzy-set methods follows the template of part 1. After offering a detailed presentation of the nature of fuzzy sets in social research (chapter 6), I address the fuzzy-set approach to the problem of viewing cases as configurations of set memberships and the

associated problem of constituting populations. Because of their deep complementarity, these two topics are joined in chapter 7. I demonstrate several advantages of fuzzy-set methods over the crisp-set configurational approach presented in part 1. For example, using fuzzy sets, cases vary in the degree to which they are instances of a location in a property space, and this information can be used to aid the evaluation of the usefulness of the property space.

Chapters 8 and 9 discuss, in turn, the analysis of necessary and sufficient conditions. I demonstrate that the core of the analysis of necessity is an evaluation of the argument that an outcome is a subset of a cause. Likewise, the core of the analysis of sufficiency is an evaluation of the argument that a cause or causal combination is a subset of an outcome. Thus, the study of causal complexity is, in essence, the study of set-theoretic relationships. These chapters demonstrate that the fuzzy subset relationship is central to the analysis of causal complexity when using data that vary by level or degree.

Chapter 10 provides two applications of these arguments, both involving country-level data. The first application is an analysis of causal conditions relevant to “severe IMF protest” (mass demonstrations and riots in response to austerity programs mandated by the International Monetary Fund). The second is an analysis of causal conditions relevant to “generous welfare states” in advanced industrial democratic countries. Chapter 11 summarizes part 2 by addressing the use of fuzzy sets as interpretive tools.

## DIVERSITY-ORIENTED RESEARCH

BETWEEN COMPLEXITY AND GENERALITY

Two main traditions can be distinguished in the history of ideas, differing as to the conditions an explanation has to satisfy in order to be scientifically respectable. The one tradition is sometimes called *aristotelian*, the other *galilean*. [. . .] As to their views of scientific explanation, the contrast between the two is usually characterized as causal *versus* teleological explanation. The first type of explanation is also called mechanistic, the second finalistic. The *galilean* tradition in science runs parallel with the advance of the causal-mechanistic point of view in man's effort to explain and predict phenomena, the *aristotelian* tradition with his effort to make facts teleologically or finalistically understandable.

—von Wright, *Explanation and Understanding*

## INTRODUCTION

Social scientists often face a fundamental dilemma when they conduct social research. On the one hand, they can emphasize the *complexity* of social phenomena—a common strategy in ethnographic, historical, and macrolevel research—and offer sensitive, in-depth case studies of specific instances. On the other hand, they can make broad, homogenizing assumptions about cases and document *generalities*—patterns that hold across many instances. Research strategies that focus on complexity are often characterized as “qualitative,” “case-oriented,” “small-*N*,” or “intensive.” Examples of this include in-depth case studies of the social cohesion of large immigrant communities in Manchester, the origins of the Korean War, and the construction of gender in sports bars in Peoria, Illinois. Strategies that focus on generality are often characterized as “quantitative,” “variable-oriented,” “large-*N*,” or “extensive.” Some examples of this are broad-based studies of the connection

between diversification and profitability across large multinational corporations, the link between technology and scale of destruction in armed conflicts, and the correlates of alternative gender identities.

This methodological divide is somewhat artificial and overly simplistic. Still, it has provided ample fodder for conflicts and controversies in many academic fields and arenas, from the pages of introductory textbooks to hiring and firing decisions at top social science departments (Shea 1997). While the contrasts between these two styles of social research are substantial, it is easy to exaggerate their differences and to caricature the two approaches, for example, by portraying quantitative work on general patterns as scientific but sterile and oppressive and qualitative research on small *Ns* as rich and emancipatory but soft and subjective. It is important to avoid these caricatures because the contrasts between these two general approaches provide important leads both for finding a middle path between them and for resolving basic methodological issues in social science.

In this chapter, I am mostly interested in these two approaches as different ways of connecting ideas and evidence, especially evidence that comes in the form of *multiple cases* or *instances*. How do we learn from multiple instances? How should we conduct *cross-case analysis*? The case-oriented strategy addresses a relatively small number of cases in an in-depth manner, paying attention to each case as an interpretable whole. In essence, this strategy is an extension of the single-case study to multiple cases with an eye toward configurations of similarities and differences. In this approach, in-depth knowledge of cases provides the basis for constructing limited generalizations that hold for the cases studied. These generalizations may or may not have wider relevance. The variable-oriented strategy, by contrast, is to look for broad patterns across many cases, usually by correlating aspects, and to draw inferences based on these broad patterns. In the variable-oriented strategy, small *Ns* are seen as untrustworthy because they are more likely than large *Ns* to provide distorted representations of broad, population-wide patterns. For the case-oriented researcher, confidence comes from depth; for the variable-oriented researcher, it comes from breadth.

As I show in this chapter and in part 1 of this book, there is a middle ground between the two strategies of depth and breadth. This middle ground emphasizes the study of diversity, an approach that emphasizes seeing cases as configurations of aspects and disaggregating populations into types. The main purpose of this chapter is to situate diversity-oriented research between two methodological traditions, case-oriented

and variable-oriented. Prominent in this examination is a discussion of practical differences separating case-oriented and variable-oriented research—basic contrasts in how researchers go about the seemingly mundane task of drawing conclusions from data (see also Myles and Huberman 1994). By focusing on the practical level, it is possible not only to understand why practitioners of these two methodological styles often talk past each other, it is also possible to see how basic ideas about cases, causes, populations, and variables constrain and shape the dialogue between ideas and evidence. I argue that the homogenizing assumptions of the variable-oriented approach pose obstacles to discovery in social research and that these obstacles are addressed in the diversity-oriented approach. While this chapter sketches many of the key features of diversity-oriented research, full presentation of the approach is the purpose of part 1.

#### *Variable-Oriented versus Case-Oriented Research*

Two empirical observations serve as a useful starting point in the analysis of these two methodological styles. First, in most social science sub-disciplines there is an inverse relation between the number of cases and the number of aspects of cases (or “variables”) that researchers study. Second, researchers tend to gravitate toward the study of either many cases and relatively few variables or many variables and relatively few cases. These two research styles offer alternate ways of constructing representations of social life from multiple instances. Of course, there are other ways to construct representations from multiple instances, but these two are the most common in social science.

The inverse relation between the number of cases and the number of variables in a study can be seen clearly in an extreme contrast: the in-depth study of the lives of a small number of individuals and a survey study with thousands of respondents. The case-study researcher examines many aspects of each case and attempts to construct a representation of each individual from the interconnections among the aspects of each case. In effect, the case-study researcher’s goal is to show how the different aspects mutually constitute the whole case and then to compare and contrast the different wholes. A survey 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 variables. This researcher constructs a generic representation based on patterns observed across many cases (e.g., using correlations among variables). In contrast to comparative case-study

research, which is deep but not very wide (i.e., “intensive”), survey research is wide but not very deep (i.e., “extensive”).<sup>1</sup>

Social scientists who study cases in an in-depth manner often see empirical generalizations simply as a means to another end—the interpretive understanding of cases.<sup>2</sup> In this view, a fundamental goal of social science is to interpret significant features of the social world and thereby advance understanding of how existing social arrangements came about and why we live the way we do. The rough general patterns that social scientists identify simply aid the understanding of specific cases; they usually are not viewed as predictive. Besides, the task of interpreting and then representing socially significant phenomena (or the task of making selected social phenomena significant by representing them) is a much more immediate and tangible goal. Empirical generalizations and social science theory are important, but their importance derives primarily from their service to the goal of interpretive understanding. By contrast, those who study patterns across many cases with an eye toward formulating generalizations believe that the fundamental goal of social science is to advance general, explanatory theories addressing wide expanses of the social terrain. In this view, social scientists should uncover general patterns, refine their theories, and use this abstract, theoretical knowledge to advance the common good. Thus, systematic theory is seen as the centerpiece of good social science as an end in itself.

1. 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. While representation and inference are both important in social science, 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 social scientific generalization. Those who study the complexity and specificity of social phenomena tend to be skeptical of empirical generalizations; those who study patterns across many cases do so with an explicit eye toward constructing such generalizations.

2. Theoretically decisive case studies, of course, seek theoretical generality as opposed to empirical generality. For example, researchers often choose specific cases because they are extreme in some way and thus present social phenomena in a “pure” way (e.g., Dumont 1970). Other researchers choose specific cases because they seem completely ordinary and thus should not skew findings or results in any particular way (e.g., Becker et al. 1961). Overall, the relationship between case studies and theoretical generalization is difficult and tortuous, and researchers must make a special (and strong) case to support the connection.

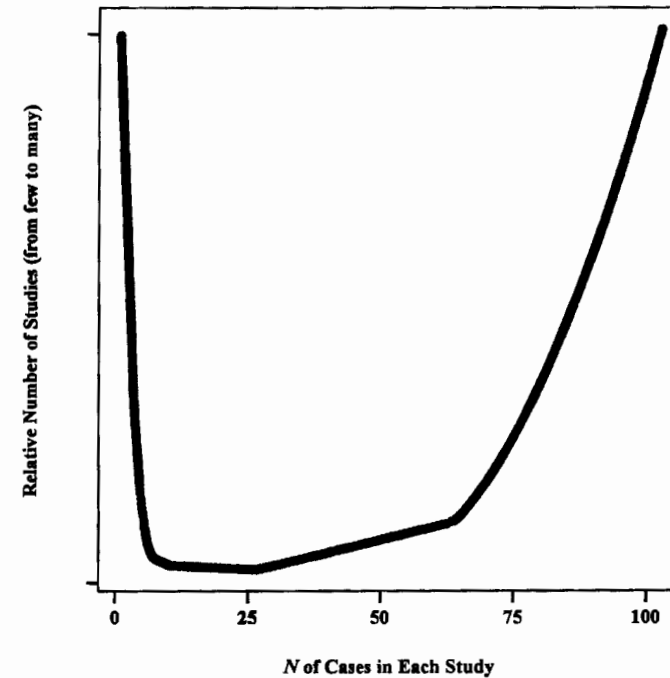


Figure 1.1 Plot of relative number of studies against  $N$  of cases in each study.

This bifurcation of social scientists with respect to the issue of generalization is evident in their published work. The examination of almost any research area in the social sciences reveals that there is a sharp divide separating those who do small- $N$  qualitative studies from those who do large- $N$  quantitative studies. In comparative sociology and comparative politics, for example, a frequency distribution showing the number of studies with different size  $N$ s reveals a clear U-shaped pattern, illustrated in figure 1.1 (see also 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 conduct studies of ten or twenty countries, but many who study one or two (case study research) or more than fifty (i.e., enough to permit the use of conventional quantitative methods). This U-shaped distribution is replicated in many other research areas.

## AN EVER-WIDENING GULF

A number of factors reinforce both this division and, more generally, the tendency for social scientists to gravitate toward one or the other styles of cross-case analysis:

1. 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 eighty countries to know each country's politics well. Imagine a survey researcher trying to keep track of the life histories 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. Empirical intimacy with cases comes at a very high price. Thus, social scientists who do in-depth research tend to restrict themselves to relatively small *N*s. Those who conduct quantitative analyses of large *N*s usually must forfeit the opportunity to gain in-depth knowledge of the cases they study.
2. Social scientists tend to identify themselves and each other as qualitative or quantitative and they 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 these differences among students. Finally, "invisible colleges" (Crane 1972) have formed based on these two methodological orientations. Thus, the social organization of disciplines tends to reinforce the U-shaped distribution illustrated in figure 1.1.
3. Social scientists have a limited repertoire of methods. Case-study methods are geared toward the acquisition of in-depth knowledge of a small number 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 track of configurations of similarities and differences across more than a handful of cases.<sup>3</sup> 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, makes statistical signifi-

3. Consider for example, the fact that a study of only 10 cases has 45 paired comparisons contained within it (Ragin 1987; Griffin and Ragin 1994). A study addressing the logically possible combinations of only 8 presence/absence attributes involves the consideration of 256 different configurations, most of which must be addressed as "thought experiments" (Weber 1949).

- cance more probable. It is very difficult to achieve statistical significance with  $N = 20$ ; it is much easier with  $N = 200$ . There are few techniques explicitly designed for intermediate-sized *N*s (however, see Ragin 1987; Ragin, Berg-Schlosser, and De Meur 1996). These methodological limitations push researchers toward the two ends of the *N*-of-cases dimension.
4. Most social scientists value both kinds of knowledge—broad, theoretically relevant knowledge (e.g., the general conditions that favor "rational" decision making) and culturally or historically specific knowledge (e.g., the origins of affirmative action as a national policy in the United States). Case-oriented methods are best suited for questions that are about culturally or historically specific phenomena. Variable-oriented methods are typically used to identify broad, theoretically relevant patterns. From this perspective, the U-shaped distribution illustrated in figure 1.1 appears to some 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).
  5. Social scientists hold conflicting views regarding how connections within cases should be studied. In small-*N* research, aspects of cases are understood in the context of the separate wholes they form. This configurational view sees parts as mutually constitutive and interconnected within a given case. In large-*N* research, by contrast, aspects of cases are connected primarily in cross-case analysis: A strong pattern of covariation across cases (e.g., a strong correlation between "human capital" variables and income) is the main evidence used to support an argument that aspects of cases are empirically connected in some way (e.g., people with more human capital use it to "get ahead").
  6. When researchers attempt in-depth analysis of an intermediate-sized *N* (say,  $N = 20$ ), they often are confronted with what seems like unmanageable diversity. Because they can achieve some in-depth knowledge of cases when they study intermediate-sized *N*s, they are able to identify intricate patterns and other subtleties that would be invisible in a quantitative, variable-oriented analysis. But it is difficult to represent this complexity with conventional techniques. At the two ends of the *N*-of-cases dimension (see figure 1.1), by contrast, the answers to the questions social scientists ask tend to be more straightforward, often univocal. A single case, in the end, typically gives a single answer. The commonalities across a small number of cases (say, three to five) likewise may provide the basis for

a single answer. At the other end of the *N*-of-cases dimension, researchers base their answers on patterns observed across many cases—evident, for example, in correlations—which average out the idiosyncrasies of individual cases, which in turn yield representations of generic processes—another form of univocality.

Given the prevalence of the U-shaped distribution of number of studies by number of cases and the many 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 of 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.

#### *The Problem of Bridging These Two Worlds*

While easy to explain, there is no simple way to justify this U-shaped pattern. It is true that social scientists value both kinds of knowledge—in-depth knowledge of cases and broad statements about patterns that hold across many cases. It is also true that these two forms of knowledge are different, as von Wright (1971) indicates in the passage quoted at the beginning of this chapter. But in-depth knowledge of cases is often dependent on knowledge of patterns that hold across many cases, and vice versa. Case-oriented researchers, for example, often cite general patterns that they themselves have not documented to explain case-specific phenomena (e.g., citing the “well-known” tendency for peasants to be highly risk-averse to explain the failure of an agricultural diversification program). If these general patterns turn out to be without empirical support, then the case-specific argument is suspect (e.g., political corruption, ignored by the researcher, may have felled the diversification project). Likewise, it is very difficult to explain broad patterns across many cases without reference to case-level phenomena (Abbott 1991). Variable-oriented researchers regularly cite unobserved case-level mechanisms to explain the cross-case patterns they document. If these mechanisms cannot be confirmed at the case level, then the variable-oriented conclusions are suspect (Collier 1993, 1998).

Consider also the fact that a deep complementary relationship between case-oriented and variable-oriented research augments their mu-

tual dependence. With variable-oriented techniques, for example, it is very difficult to address questions about actors' motives and subjectivities or to observe event sequences and causal connections (Rueschemeyer 1991; Rueschemeyer and Stephens 1997). Case-oriented methods, by contrast, excel in these areas. With case-oriented techniques, however, it is difficult to gain confidence that findings are general in any way. This is one of the central strengths of the variable-oriented approach (Goldthorpe 1991, 1997; King et al. 1994; Lieberman 1992, 1998; Lijphart 1971, 1975).<sup>4</sup>

The complementarity and mutual dependence of these two worlds of social research undermines the idea that there can be a convenient division of labor between the two camps. As much as they might want to, the two camps cannot ignore each other. Perhaps this mutual dependence, combined with an unhealthy measure of mutual suspicion, explains the periodic eruption of conflicts over method in academic journals, departments, and other arenas.<sup>5</sup> While it seems unlikely that these two worlds will ever merge, it is clear that their tendency toward mutual suspicion is unhealthy (Shea 1997).

One way to link the two worlds is to build general knowledge of cross-case patterns from many, many case studies. This general knowledge can be gleaned from accumulated case-based knowledge or from strategically conducted studies of specific cases and situations (Lieberman 1985).<sup>6</sup> Rarely, however, is general knowledge constructed in this manner. Indeed, the two forms of knowledge usually are constructed

4. Because the two approaches are complementary, there are many ways to construct useful hybrid strategies (Janoski 1991). For example, a case-study researcher might apply a variable-oriented, quantitative method to strengthen some of his or her inferences and at the same time extend his or her knowledge of the case (King et al. 1994). This tactic is common among social historians and historical sociologists who use quantitative methods (e.g., Tilly 1986). Another hybrid strategy is to use a small number of case studies to flesh out and document patterns observed across many cases (Paige 1975).

5. Consider, for example, the report of deep divisions in political science reported in the *Chronicle of Higher Education* (Shea 1997). There is a deep division between the proponents of the rational choice perspective and those who favor in-depth knowledge of countries. This rift is especially pronounced in comparative politics, the branch of political science most concerned with area studies, the specificity of political systems, and their embeddedness in culture.

6. More generally, Lieberman (1985) sees small-*N* studies as strategic only and doubts their usefulness for inference. One could imagine, however, an entire case-oriented methodology organized around the idea of strategic selection of sites that could be used to buttress inferences.

independently, and few scholars have tried to distill and systematize the knowledge generated from many case studies.<sup>7</sup>

As an alternative, accumulated case-based knowledge could be used to evaluate the results of large-*N*, variable-oriented studies of cross-case patterns. If a deep chasm separates knowledge generated from accumulated case studies, on the one hand, and knowledge generated from the study of patterns across many cases, on the other, then one or both are suspect or at least need to be qualified or revised in some way. This comparison of results happens only occasionally, however, because the two ways of generating knowledge produce findings that are often incompatible and sometimes contradictory.

An excellent example of the problem of case-based knowledge contradicting variable-oriented knowledge is detailed in Dietrich Rueschemeyer, Evelyne Huber Stephens, and John Stephens's work *Capitalist Development and Democracy* (1992). Variable-oriented research has demonstrated a consistent link between economic development and democracy, documented via the strong correlation between these two variables (level of economic development and strength of democracy) in quantitative cross-national research. Typically this research has cited causal mechanisms directly linked to economic development (e.g., social mobilization) to explain the correlation. Case-study research, by contrast, has shown no such link. Rueschemeyer et al. argue that while it is true that economic development and democracy are strongly correlated across cases, the mechanisms producing this correlation are not directly tied to economic development per se. They study many countries in moderate depth, focusing on configurations of similarities and differences, and show that various combinations of class-based and institutional factors have shaped democratic institutions in different regions of the world. Contradictions of the type documented by Rueschemeyer et al. make it very difficult to bridge the two worlds of social research.

#### *The Impact of Analytic Procedures on the Disjuncture*

Despite the many interdependencies and complementarities of case-oriented and variable-oriented research and the dependence of both sets of researchers on the same body of social theory, there are very good reasons to expect disjunctures in the knowledge they produce. Primary among these reasons is the simple fact that the two approaches

7. Exceptions to this statement include such scholars as Barrington Moore, Jr. (1966), Reinhard Bendix (1978), and Immanuel Wallerstein (1974).

use very different analytic procedures to construct representations of social phenomena (Ragin 1994a). Consider, for example, the contrast between computing a correlation between a causal and an outcome variable, a popular analytic step in variable-oriented work, and identifying the causally relevant conditions linked to a specific outcome across a small number of cases, a well-worn analytic path in case-oriented research (e.g., Brinton 1965; Chirot 1994; Hobsbawm 1981; Wolf 1969). While both procedures seem deceptively simple and straightforward, these two ways of constructing representations of social phenomena from evidence involve sharply contrasting orientations toward cases, outcomes, and causes. That is, they differ fundamentally at a *practical* level. It is not necessary to invoke the contrasting emphases of the two worlds of social research—inference versus interpretive understanding—to account for the disjunctures in the knowledge they produce (as in Truzzi 1974).

#### *Cases*

When variable-oriented researchers compute a correlation between two variables, the cases become more or less invisible, and variables take center stage. Furthermore, the set of cases included in the computation must be fixed before the researcher can compute the correlation. Once this set is fixed, usually at the outset of an investigation, it is rarely altered. What matters most is that the cases (which are understood as “observations”) belong to the same general “population” and that they be drawn from this population with an eye toward randomness or representativeness or some combination of such criteria. If the relevant population is moderate in size (e.g., counties in North Carolina, members of the U.N., and so on), then all observations may be included in the analysis.

In a comparative case study of commonalities, by contrast, cases have clear identities and are usually chosen specifically because of their significance or their theoretical relevance (George 1979). Furthermore, the set of relevant cases may shift during the investigation because the researcher may decide that one or more cases do not seem like the others. For example, a researcher studying social revolutions might decide that the Mexican case does not fit well with the Russian, Chinese, and other cases and then drop it from the investigation. This flexibility is maintained throughout the investigation because the core concepts (e.g., “social revolution”) may be revised as the researcher learns more about relevant instances. In the end, the delineation of an empirical category and the clarification of core theoretical concepts relevant to

this category may be the primary achievements of the research effort—that is, beyond the new interpretive insights into cases offered by the researcher.

#### *Outcomes*

In correlational studies researchers usually identify a “dependent variable”—an outcome that varies across cases. Typically, such outcomes are aspects of cases that vary by level, for example, level of bureaucratization, level of democracy, and so on. Sometimes the outcome variable is categorical, indicating, for example, whether or not some event has occurred (e.g., debt renegotiation), and sometimes it is a frequency or a rate (e.g., the murder rate). The important consideration, in this procedure, is that the outcome must vary across relevant “observations.” The goal of the research typically is to explain, if possible, why each case has the value or level on the dependent variable that it does. Typically, cases are assessed relative to the average of all cases, using some measure of central tendency such as the mean as a statistical benchmark for evaluating each case. Such research is centrally concerned with the question of “why.” For example, a researcher might seek to explain why some countries are more democratic than average and others less so, why some organizations are more bureaucratic than average and others less so, or why some people have more education than average and others less so.

In a comparative case study of commonalities, by contrast, the outcome is often something that does not vary substantially across cases. In a study of peasant revolts, for example, cases are chosen precisely because they all display the same outcome—peasant revolts. Recall that the goal of this type of research is to identify common causal conditions linked to a specific outcome across a relatively small number of cases.<sup>8</sup> Thus, the focus is on cases with a specific outcome, not cases that vary widely in how much they have this outcome. While the outcomes in a study of this type will not be exactly identical (for example, not all peasant revolts are exactly alike), the researcher must demonstrate that the outcomes in the cases selected are in fact enough alike to be treated as instances of the same thing. Finally, unlike correlational studies, which are centrally concerned with the question of “why” (as in: Why

8. The study of the causal conditions shared by instances of an outcome is primarily directed toward the identification and evaluation of necessary conditions (see especially chapter 4).

some more than others?), comparative case studies are centrally concerned with the question of “how” (as in: How does it happen?). How do couples break up (Vaughan 1986)? How do community organizations get started? How do peasant revolts come about?

#### *Causes*

In a correlational study, causation typically is inferred from a pattern of covariation. If a variable thought to represent a cause or to be an indicator of a key causal condition is strongly correlated with the outcome variable, then the researcher may make a causal inference. Usually, the researcher will assess the relative strength of several causal variables at the same time. The typical goal is either to find out which one explains the most variation in the outcome variable or simply to assess the relative importance of the different variables. In effect, the independent variables compete with each other to explain variation. The causal variable that is least correlated with the other causal variables and most strongly correlated with the outcome variable usually wins the competition. In most investigations, each causal variable is considered sufficient, by itself, for the outcome or some increment in the outcome. That is, each one is considered an “independent” variable capable of affecting the outcome variable, regardless of the values of other causal variables.

In a comparative case study of commonalities, by contrast, causation is typically understood conjuncturally, in terms of combinations of conditions. The goal of this type of analysis is to identify the causal conditions shared by the cases. Causal conditions do not compete with each other, as they do in correlational research; they combine. How they combine is something that the researcher tries to piece together using his or her in-depth knowledge of cases. Because all the cases have more or less the same outcome, the usual reasoning is that the causal conditions shared by cases provide important clues regarding which factors must be combined to produce the outcome in question. Further, the researcher may argue that these common causal conditions are necessary for the outcome (see chapter 4). The search for such conditions is not a simple mechanical process of listing potential causal conditions and then seeing which ones are constant across relevant cases. Often the researcher tries to construct a composite portrait, often linked to a general narrative, from his or her in-depth knowledge of the cases included in the study. When constructing this argument, the researcher is especially sensitive to the possibility that a given causal requirement

(i.e., one of the necessary conditions) can be met in several different ways (see, e.g., Moore 1966).<sup>9</sup>

These and other practical differences in how case-oriented and variable-oriented researchers work with evidence to construct representations of social phenomena provide many opportunities for disjunctures in their findings. These practical differences also create a great deal of mutual suspicion. For example, from the perspective of variable-oriented work, the study of commonalities across a small number of instances is fraught with analytic sins and errors, among which are the following: (1) The number of cases is too small to warrant any kind of inference. (2) The procedure selects on the dependent variable (i.e., focuses on extreme or noteworthy cases, all with more or less the same value on the outcome variable). This practice deflates otherwise robust correlations (Geddes 1990; King et al. 1994; Collier 1995; Collier and Mahoney 1996). (3) The most important causal factors do not vary and thus are impossible to assess. Likewise, from the perspective of case-oriented work, the examination of the correlation between a causal and outcome variable across many cases is fundamentally flawed. For one, there are typically so many cases that there is no way for the researcher to know if they are all really comparable and thus belong together in the same analysis. Also, it is difficult to determine how something “comes about” by comparing cases with different levels of the outcome. The partial instances (i.e., those with lower scores on the outcome variable) are likely to provide many false leads. It is also pointless to try to isolate the “independent” effect of any causal condition when several factors usually must be combined for a particular outcome to occur (Ragin 1987, 1997).

While the practical differences between the two approaches reinforce the gulf between them, they also provide important clues for how to build a bridge. The best way to build this bridge is to find a middle ground where the concerns of both worlds can be met. This middle ground obviously involves compromises on both sides, but it should provide a way to address their practical differences, especially their different approaches to the construction of representations of social phenomena from empirical evidence.

9. This approach is not limited to the search for a single combination of causal conditions linked to an outcome. The researcher may identify several combinations, which in turn might prompt a reformulation of the outcome into types. I focus on the search for common causal conditions in this discussion to simplify the presentation and sharpen the contrasts in analytic procedures.

In the remainder of this chapter, I sketch this middle ground. I emphasize the study of *diversity*, a concept that bridges complexity and generality and provides the basis for a third world of cross-case analysis. The development of the diversity-oriented approach is the first of the three pillars of this book's central argument. The diversity-oriented approach drops many of the restrictive, homogenizing assumptions of conventional variable-oriented analysis without forsaking its emphasis on breadth—on studying many cases. By freeing researchers from these homogenizing assumptions, the diversity-oriented approach opens up the dialogue between ideas and evidence in social research, enlarging the potential for discovery.

#### *Diversity: Between Complexity and Generality*

The concept of *diversity* provides a way to balance the representational concerns of social scientists who lean toward the humanities—case-oriented researchers—and the inferential concerns of those who favor the hard sciences—variable-oriented researchers. Diversity is best understood as a synthesis that transcends these two opposing principles of generality and complexity. To study diversity is to take a broad view of social phenomena, as in much variable-oriented work, without imposing homogenizing assumptions at the outset of the research, the usual variable-oriented strategy (for example, the assumption that all cases are drawn from the same “population”). In short, the goal of diversity-oriented research is to find a middle path between treating analytic objects as members of fixed, homogeneous populations, on the one hand, and focusing exclusively on the specificity of individual cases, on the other.

When social scientists attend to the diversity of social phenomena, they make special allowance for kinds and types. Studying social phenomena in terms of their different kinds or types lies midway, conceptually speaking, between studying general patterns across “all” cases, on the one hand, and attending to the complexity of specific cases, on the other. Consider, for example, a researcher interested in military coups. The researcher could see coups as members of a single, all-inclusive population and investigate how they vary—in scale, intensity, duration, consequences, and so on. This study would focus on generality—on broad patterns observed across “all” cases. Alternatively, the researcher might select a relatively small number of coups for in-depth analysis and try to elucidate the specific character of these coups. The coups selected might be chosen because of their historical significance for a particular region of the world, or they might be especially well

suiting for answering a specific theoretical question, or they might be selected for some other reason. This study would focus on complexity, the specific character of the selected coups, and culminate in an in-depth representation of a relatively small, carefully delineated subset of coups.<sup>10</sup>

The study of the diversity of military coups would bridge the concerns manifested in these two common types of investigations. Rather than assuming that all coups are members of a single population or focusing on a carefully delimited subset of coups (or even on a single coup), a study of the diversity of coups would focus on their varied forms. After all, the broad category *military coups* may lack empirical integrity. Phenomena that superficially seem to be members of the same population and that are all conventionally labeled *military coups* in fact may belong in several different empirical categories. A study of the diversity of coups might culminate in a conceptual map showing types of coups and their varied causes and consequences, including mixed types. One conclusion of a study of the diversity of coups, for example, might be that under certain conditions (e.g., specific geopolitical contexts), some types of coups contribute to the long-term advance of democracy and social justice.<sup>11</sup>

It is clear that studying diversity is different from studying complexity, the usual focus of case-oriented research. In research that emphasizes complexity every case may seem unique, and researchers may spend considerable time and energy simply trying to find a handful of cases that are similar enough to permit comparisons. Often, case-oriented researchers resort to “contrast-oriented” (Skocpol and Sommers 1980; Skocpol 1984) and “particularizing” (Tilly 1984) comparisons precisely because the differences among their cases seem to greatly

10. Those who do small-*N* case-oriented studies sometimes go to great lengths to argue that the few cases examined in an investigation are the best cases to study for a given question, that these cases are the most representative of the general process in question, or that they are the most decisive relative to the theoretical issues at hand (see, e.g., Bendix 1978; Lipset 1950, 1963; Moore 1966; Skocpol 1979).

11. It is important to point out that allowing for multiple forms and types does not dictate that they *must* be found. Nor is it necessary in a diversity-oriented study to produce a conceptual map showing *all* types of a particular phenomenon (e.g., all types of military coups). The key consideration in studies that attend to diversity is that researchers constantly question key empirical categorizations and refrain from accepting any category boundary as fixed, at least not until the research is more or less complete.

outweigh their similarities. Diversity-oriented research, by contrast, tends to see social phenomena in terms of types and kinds and thus allows for middle-range generalizations—statements that refer to categories of social phenomena (e.g., types of military coups). In essence, diversity-oriented research emphasizes similarities among cases in the formulation of types and sees the specification of types, subtypes, and mixed types as an important means for understanding and explaining differences. Thus, like variable-oriented research, diversity-oriented research seeks to comprehend many cases at once.

Studying the diversity of social phenomena, however, is distinctly different from the variable-oriented study of covariation. For example, suppose a variable-oriented study of wars revealed that the greater the cultural differences between the opposing sides, the greater the loss of life (net of the independent effects of other causal conditions). While breathtaking in scope, this type of analysis might encompass strikingly heterogeneous cases. In order to examine the link between cultural differences and loss of life in wars, it would be necessary first to constitute and delimit the set of relevant wars as comparable cases—as members of a single population. But what is war? Should civil wars be included? Anticolonial uprisings? Armed rebellions? Wars that last three days? A week? Wars that involve less than one hundred fatalities? Should wars between preindustrial countries be included? Wars between preliterate peoples? How should the phenomenon, once defined, be delimited historically or geographically? Should it be?

Before the variable-oriented researcher can compute a single correlation assessing any possible connection between cultural differences and loss of life, the set of relevant cases must be well defined and delineated—it must be fixed. Once the set of relevant cases is established, conventional statistical techniques treat these cases as equivalent observations, like independent coin tosses. Thus, using these techniques presupposes a substantial degree of homogeneity among cases. Furthermore, once this population boundary is set, the question of homogeneity is typically not reopened. Any change in the boundary of the set of relevant observations would change the correlations among the variables used to characterize the cases included in the set. Fixing a population boundary in variable-oriented research tends to fix the assumption of case homogeneity as well. Thus, these procedures, which are central to this approach, mask the diversity of social phenomena, cloaking them in fixed “populations.”

In diversity-oriented research, by contrast, population boundaries are not taken for granted, nor are they fixed. Instead, they are fluid.

They can be revised up until the very end of a research project, as the investigator's knowledge of cases grows and deepens. Furthermore, in the diversity-oriented approach researchers are ever-cognizant of the possibility that the sets they study may need to be disaggregated into types or kinds and studied separately. Thus, in this third type of cross-case analysis, investigators may disaggregate and differentiate, as in case-oriented research, but not to the point that every case seems unique.

The concept of diversity advanced here seeks both generality and complexity in the emphasis on the delineation and systematization of types, forms, trajectories, and paths. It avoids the homogenizing bias of the variable-oriented approach, evident in its dependence on fixed populations, and the particularizing bias of the case-oriented approach, evident in its tendency to focus almost exclusively on differences and thus highlight the specificity of each case.

#### *Studying Diversity: Practical Considerations*

In the remainder of this chapter, I offer a brief sketch of several of the core features of diversity-oriented methods. This discussion also serves as an overview of the balance of part 1 of *Fuzzy-Set Social Science*, culminating in chapter 5 with an application of the diversity-oriented approach to configurational data on set memberships, using Boolean algebra (i.e., crisp sets). In part 2 of this work, I use insights from fuzzy-set theory to broaden and deepen this approach. The end product is a methodology appropriate for studying social diversity in its varied forms and manifestations.

At first glance, the problem of studying diversity seems simply to recreate one of the basic dilemmas sketched at the beginning of this chapter—the difficulty of knowing a large number of cases in an in-depth manner. However, there are practical procedures useful for studying the diversity of cases that avoid the many homogenizing assumptions of the variable-oriented approach. To study the diversity of social phenomena it is necessary to formalize and extend several aspects of the case-oriented approach. Here, as before, I focus on practical aspects of constructing social scientific representations from empirical evidence on multiple cases. My goal in integrating these features into diversity-oriented research is to bring some of the logic and intensity of case-oriented research to investigations involving large numbers of cases. The basic problem is to examine similarities and differences across many cases while preserving the integrity of cases as complex configurations.

Diversity-oriented research borrows from case-oriented research the following features:

1. *Its understanding of cases as configurations:* In case-oriented research the different parts of a case are defined in relation to each other—in terms of the whole they form. For example, one aspect of a country's political system (e.g., its multiparty character) is understood in the context of other features of the country (e.g., its ethnic diversity). This way of approaching cases is not the same as using one aspect to "account for" another (e.g., as in using degree of ethnic diversity to explain why a multiparty system exists in some countries but not in others). It is a matter of interpretation: Having a multiparty system conveys different things about a political system depending on whether significant ethnic diversity exists. Similarly, at the individual level, the meaning of regular church attendance could be very different for those who are political activists versus those who are politically inert. In the configurational view, to change one important aspect of a case potentially alters the character of the case as a whole.
2. *Its view of populations as flexible, manipulable constructions:* Often in case-oriented work, a key research issue involves the conceptual identification of cases—what are they cases of? Rather than starting with a delineated population of observations or a sample drawn from a seemingly "known" population, the case-oriented researcher typically starts with orienting concepts and a handful of cases that, at first glance, seem relevant to the concepts. In the course of the research, there is often a reciprocal clarification of these concepts and of the boundaries of the empirical category of cases relevant to the concepts (Ragin 1994a; 1997). The set of relevant cases may shift substantially in the course of the research; it may subdivide into types; it may broaden; it may contract to a single case or even to a theoretical case. How the boundaries of this set change depends on the nature and direction of the dialogue between the researcher's ideas and his or her empirical evidence. The end product of such research is usually both empirically and theoretically relevant because empirical cases help researchers elaborate their theoretical ideas, which, in turn, deepen their understanding of their cases.
3. *Its emphasis on outcome-oriented investigation:* Case-oriented studies typically address specific qualitative changes in specific contexts. In some studies the qualitative changes are dramatic (e.g., countries with social revolutions, as in Skocpol 1979); in others the changes

emerge very slowly through time (e.g., countries experiencing similar historical transformations of certain institutional arrangements; see Bendix 1978). Very often the boundaries of an investigation are set initially by the universe of relevant qualitative changes (e.g., a study of instances of the emergence of strong national identities in postcolonial countries). Thus, case-oriented investigations are often outcome-oriented in the sense that the starting point of such studies is very often the identification of some qualitative change or aspect common to a set of cases. When case-oriented researchers confront diverse outcomes, a common practice is to grasp such diversity as divergent paths.<sup>12</sup>

4. *Its view of causation as conjunctural and heterogeneous:* In case-oriented investigations, explanations of outcomes typically cite combinations or “conjunctures” of causal conditions. The classic example of this kind of explanation is Weber’s (1978) explanation of the conditions that combined to give rise to rational capitalism in the West, a causal conjuncture that Weber argued was unique. John Stuart Mill ([1843] 1967) called this type of causation *chemical* because the outcome, a qualitative change, results from a combination of causal conditions. In case-oriented social science, attention typically is directed toward understanding how the different causal conditions combine in each case to produce the outcome in question. A common finding is that different conditions combine in different and sometimes contradictory ways to produce the same outcome. Thus, an interest in causal heterogeneity often accompanies a conjunctural view of causation. There is no presumption that the same causal factors operate in the same way in all contexts and all cases. The effect of any particular causal condition may depend on the presence or absence of other conditions, and several different conditions may satisfy a general causal requirement—that is, they may be causally equivalent at a more abstract level.

These practical features of case-oriented research, which are also central to diversity-oriented research, pose serious obstacles to the conduct of conventional variable-oriented social research:

The first feature—the idea that aspects of cases should not be viewed in isolation from each other—challenges conventional conceptions

12. When cases follow diverging paths from common origins, the researcher can use different paths to define types of cases and then examine similarities among cases following a given path and differences between sets of cases following alternate paths.

of the variable. In variable-oriented research, variables are typically viewed as analytically distinct and separable aspects of cases, and most statistical analyses assume that the meaning of a category or a value on a variable is more or less the same across all cases. The variable *years of education*, for example, should indicate the same thing for all cases. It should not indicate intellectual imprisonment for some and enlightenment for others—depending on these individuals’ other characteristics. The idea that each aspect should be examined in the context of other aspects of the same case wreaks havoc on most variable-oriented analytic procedures because this principle suggests that apparent similarities (e.g., having twelve years of education) may be illusory.

The second feature challenges conventional notions about populations. In variable-oriented research, populations typically are seen as empirically given—for example, the population of a country as shown in a census—and are rarely contested. Furthermore, as previously noted, most variable-oriented techniques (such as correlation and regression analysis) are predicated on population and sample boundaries that are more or less fixed at the outset of data analysis. To vary the population boundary throughout an investigation, in effect allowing it to be flexible and to shift, poses a practical impossibility.

The third feature—the emphasis on outcome-oriented analysis—argues that researchers should begin their research by focusing on the best and clearest examples of whatever outcome they are interested in. (This practice dovetails with the emphasis on keeping population boundaries fluid and establishing them in the course of the research.) From the perspective of variable-oriented analysis, however, this strategy is equivalent to studying a dependent variable that has little or no variation—another practical impossibility. In case-oriented research relevant negative instances are difficult to identify until the positive instances are well known and understood (Ragin 1997). Thus, in this approach it is common first to study causal similarities and differences among only positive instances. In statistical parlance this amounts to selecting on the dependent variable—a sin to be avoided (Geddes 1990; King et al. 1994).

The fourth feature—the focus on causal conjunctures and causal heterogeneity—poses major obstacles to the use of additive-linear models, the metatheoretical foundation of most forms of variable-oriented analysis. The guiding principle of such models is that it is possible and useful to assess the independent effect of each causal variable, net of the effects of all other causal variables. The assessment of the relative importance of different causal variables is the primary basis for

“theory testing” in the variable-oriented approach. However, if causation is complex and each outcome may result from various conjunctures of conditions, then it is impossible to estimate each cause’s “independent” effect. In variable-oriented work, causal conjunctures are assessed through analysis of statistical interaction. But interaction models present intractable estimation and specification issues, especially when higher-order interactions are present, when the number of cases is modest, or when their diversity is limited (Ragin 1987).

The many obstacles to variable-oriented analysis posed by these practical features of case-oriented work exist for one simple reason: *Conventional variable-oriented analysis is predicated on very powerful and consequential homogenizing assumptions.* These homogenizing assumptions are clearly visible in the understandings of cases, variables, populations, and causes that form the core of the variable-oriented approach, at least as it is conventionally practiced. As I show in the remainder of part I, diversity-oriented research incorporates key features of the case-oriented approach precisely to avoid these homogenizing assumptions. To make such assumptions in social research seriously undermines the study of the diversity.

## TWO

### CONSTITUTING POPULATIONS

The social scientist finds himself in continuous effort to combine two ideals: vision and precision. Clearly, he needs precise instruments if he wants to develop testable propositions. But the social world is very complex and doesn’t provide us with the well delineated objects from which the natural sciences start. Thus, he also needs a great deal of creative imagination—of vision—to decide on the objects about which such propositions should be developed.

—Lazarsfeld, foreword to *Constructive Typology and Social Theory*

#### THE CENTRALITY OF POPULATIONS TO CONVENTIONAL SOCIAL SCIENCE

The most common form of social science involves the examination of patterns across multiple cases, “cross-case analysis.” A researcher studying wars might investigate hundreds of wars; a demographer studying divorce might survey thousands of couples; an ethnographer interested in professional athletics might interview and observe dozens of athletes in a variety of sports. In variable-oriented work multiple instances of “the same thing” are highly valued because they create the possibility of quantitative analysis and statistical inference—if cases are plentiful and sampling is adequate. A quantitative analysis of wars, for example, might show that the greater the cultural differences between the opposing sides, the greater the loss of life, controlling for the effects of other relevant causal conditions. If observed, this pattern would have an important impact on social scientists’ view of war, and it could become a key element in their representations of the phenomenon. Cross-case analysis is also central to many types of case-oriented work. An ethnographic researcher, for example, might observe a variety of street ven-