Five

Finding the Objects to Study

Theorizing demands that we think about the real social world abstractly and in general terms. But to do the empirical research upon which theories rest, one must answer the mundane but essential questions: "Where do I go?" and "What do I look at?" Answering these questions requires both a clear definition of the actual social units about which data will be collected and a strategy for sampling from the universe, or population, composed of all such units. Defining these units of study is in part a matter of deciding what types of objects are theoretically relevant to the problem, and the choice of a sampling strategy is in part determined by the kinds of theoretical and empirical generalizations required to solve the research problem. However, research involves a variety of practical constraints as well as theoretical demands, and these must also be considered.

The studies that are feasible at any given moment are confined to particular places, times, persons, and variables. Each study is only a sample of the much larger universe of studies that might logically follow from a research problem's theoretical formulation. This chapter addresses the ways in which researchers employ different definitions of the objects of study and different sampling strategies to cope with empirical inquiry's varied demands for—and constraints upon—generalization from data and how a multimethod approach may improve our ability to generalize.

The Dilemma of the Few and the Many

The research process requires trade-offs between researchers' ultimate theoretical objectives and the various constraints that determine the feasibility of particular types of studies. Because of this, research inevitably focuses social scientists' immediate attention more narrowly than their broader theoretical interests imply. However, few social psychologists are really curious only about the 20 undergraduates recruited to participate in an experiment, just as few urban sociologists are chiefly concerned about the particular neighborhood of a middle-sized city selected for a case study, and few network analysts are interested solely in the 1500 people selected in a sample survey for the study of their friendship choices. Instead, they are mostly interested in what the findings about those 20 undergraduates, that single neighborhood, or those 1500 individuals might tell us, respectively, about other young adults or even people in general; about this type of neighborhood, or even all neighborhoods; or about the friendship patterns of the eight million people in the metropolitan area from which these 1500 were selected, or even of people in the entire country.

The objects we study are interesting primarily because our findings may apply more generally to a larger class of similar objects. The objects of study and the objects of interest correspond, respectively, to the sample and the universe. Whether the findings from the objects studied will hold true for the objects of interest is a question of the *representativeness* of the sample: How well do these few represent the many? The answer to this question will depend upon how the units and the universe were defined and the way in which the sample was selected. We will see that different styles of research characteristically treat these issues in different ways.

Sampling and the Costs of Precision

Research, like all human activity, involves costs. Sampling is one way of reducing those costs. The precision of our knowledge is directly related to its cost; that is, greater precision requires greater outlay. The degree of precision sought is, therefore, never absolute but instead relative to the needs and purposes of the research, and utility criteria often determine our information's precision. In everyday life, the degree of precision required is similarly determined. For instance, to decide whether or not to carry an umbrella we usually look out the window rather than conduct a meteorological study. In science, however, the degree of precision is in principle determined by the demands of knowledge rather than by practicality. This is perhaps the source of the popular image of the mad scientist ignoring all costs, moral as well as material, in the "irrational" pursuit of truth. It may also be why such scientists are often depicted as independently wealthy. Actual scientists are neither rich nor crazy and are acutely aware of the costs of their research. Consequently, they are willing, in a rational trade-off, to accept approximations to knowledge.

Sampling, furthermore, facilitates other kinds of precision. By studying a few selected units rather than the whole universe, one may put additional resources into sharpening measurement by acquiring more and also more accurate data about fewer units. However, reducing measurement error may increase sampling error. This dilemma is similar to the Heisenburg "uncertainty principle" in subatomic physics, wherein one may know with great precision the momentum of a subatomic particle but not simultaneously its position, or conversely, its position but not its momentum. Thus, one may spend years getting to know a single case in great detail, as with an individual in psychoanalysis or a community or organization in fieldwork ethnographies. But one may end up with a less convincing argument for generalizing one's findings than other researchers who may ask the same individual only a few questions in a survey or look up a few published census statistics on a neighborhood.

Units and Universes

If, as the poet Pope said, "The proper study of mankind, is man," then social scientists should study people. However, we must further ask "What is it about people we wish to study?" Rather than looking only at individuals as the units of study, we may be interested in the *groups* in which people live and work, the *interactions* they engage in ranging from murder to love, or the *settings* in which they are born and die. All are, of course, related in varying ways to people and their social behavior, but groups, interactions, and settings are the units, respectively, not persons.

Only by clearly defining our units of study can we select appropriate styles or methods to collect data about those units. If we have defined our units of analysis as eighteenth-century revolutions, for example, then participant observation is obviously of limited use compared with content analysis of archives and written accounts. If we are interested in the current problems of rape, defining the unit of analysis may present more problems. Will our units of analysis be the individual rapists or the victims, will it be the rape itself as an interaction, or will it be communities' varying rates of rape? The unit of study is intrinsically linked to the definition of the problem and the specific propositions and hypotheses that one wishes to develop or test.

A clear statement of the problem and a clear definition of units of study would seem to imply distinct methods appropriate to gathering data about those problems from those units. However, the constraints and limits of research often restrict one to certain feasible data sets that may in fact require a redefinition of units and a recasting of the theoretical problem. The link between how a problem is formulated, the definition of units, and the nature of one's data cannot be logically broken. But, it can be extended by recognizing that multiple data sets and different units of analysis may permit different theoretical formulations of the problem being investigated, and *vice versa.*

Units and Variables

In defining units of analysis, care must be taken to distinguish between the units of study and a study's variables. Variables are characteristics of the units of study, which may vary from unit to unit. This may seem to be an easy distinction, but if we return to the example of rape, certain complexities can be demonstrated to sensitize the researcher to confusions that might exist on this point in other research. Suppose that one is interested in studying something about the settings in which rapes occur. Should settings be taken as the unit of study and rape treated as a variable, or should rapes be treated as the units of analysis and characteristics of the settings taken as variables? The answer depends both upon the specific questions one hopes to answer and upon the nature of the data one can collect. If one wants to know where rapes do and do not occur, then settings should be defined as the unit of analysis, and rapes should be treated as a variable across settings. If instead one is interested in differences in settings that may be related to other characteristics of rapes, such as when or to whom they are most likely to occur, then rapes would be defined as the units of study, and characteristics of the settings would become variables.

One way to keep these issues clear is to construct a data matrix in which units and variables are the two dimensions. See Figure 5.1.

In a field study of a community or an organization, there might be only a single unit about which to collect data, but in a sample survey the

		Variables		
		Var. A	*Var. B*	*Var. C*
	Unit 1			
Units	*Unit 2*			
	Unit 3			

Figure 5.1 A Data Matrix of Units and Variables

number of units may be in the thousands. Furthermore, in any given piece of research one might collect data on different types of units simultaneously, as when a fieldworker or an experimenter studies the group as a whole as well as the individuals, subgroups, or dyads that comprise the group. As units of study shift, so too does the nature of the data collected and the types of problems that can be explored.

The data matrix is a heuristic tool that describes a range of trade-offs between the number of units studied and the number of variables studied for each unit. For example, the case study may be thought of as focusing on only a single row but on many columns of the data matrix; that is, collecting many variables (the "rich data" for which case studies are valued) on a single case. By contrast, large-scale sample surveys (national opinion polls, for example) collect data on one or a few variables (one or a few columns) across many different units (rows). The trade-off between the number of cases studied (rows) and the number of variables studied for each case (columns) represents a real-world constraint on research design. In short, though one might ideally prefer to study many variables about many units, theoretical demands and limited resources (time, money, and personnel) may narrow the choices.

However, the data matrix also suggests the possibility that a single piece of research may fall at a variety of points within "the attribute space of research design" defined by the matrix. For example, the *comparative* case study begins to expand the number of columns to two, three, or more cases moving in the direction of styles relying on larger samples. Alternatively, surveys or other large sample designs may select not one universe to sample from but several, which then become different cases to be compared. Multimethod research designs clearly offer many possibilities here. For example, longer in-depth interviews may shrink sample size (the columns) but expand the rows (the number of variables studied) compared with telephone or mail surveys having a larger number of sampled units but fewer questions and fewer variables.

Types of Units

It is useful to list briefly the more common types of units and the corresponding styles or methods of research commonly used to study each.

Individuals. Individuals, considered as units, have doubtless been studied with the widest range of methods, from direct observation and interviewing, to analyses of archives, and experiments. The application of so many different methods to the study of individuals suggests that the linkage between the way in which problems are stated and the methods selected to study them may be attributable more to convention than to intrinsic properties of methods or the nature of the units studied.

Attributes of Individuals. Characteristics of individuals are more often treated as variables than as units of study. However, just as medical researchers may focus on individual's specific organs as the unit of investigation (e.g., kidneys, of which most *individuals* have *two*), so social scientists may focus on individual's specific characteristics. For example, in attitude research social psychologists are often interested in different attitudes as the units of analysis, recognizing that attitudes themselves may have variable properties such as saliency, intensity, and direction. Experimental and survey research have been commonly used to study these properties. However, field observation of behavioral expressions of attitudes as well as content analyses of documents and artifacts may also be employed.

Actions and Interactions. The units of analysis are often defined as specific actions or interactions. Take the foregoing example of rape. Rather than focusing on the criminal or his victim as individual units of analysis, we may focus instead on the rape as an action or behavior, considering either the rapist's acts, as in many criminological studies, or the victim's response, as in many victimization studies. However, we may also define rape as an interaction, and by so doing define yet another class of variables for investigation (degrees of force used and resistance encountered or mutual prior acquaintance).

Observation, in either field or laboratory settings, is the method that most directly measures behavior. Interviews may provide self-reports of behavior, or written records may contain accounts of specific actions, but they are only behaviors in themselves if one focuses on the activities of self-reporting and record keeping. By being aware of these nuances in the definition of units of analysis with respect to behaviors, one may avoid numerous problems of inference in later stages of analysis. Furthermore, it may alert one early on to select methods that will more directly measure behaviors.

Residues and Artifacts of Behavior. When we are unable to observe individuals or their behaviors directly, we may instead exploit the fact that much behavior leaves residues. Residues are physical products or artifacts that may be selected as units of study and from which we may make inferences about the behavior itself. Webb, Campbell, Schwartz, and Sechrest (1966) discuss at length the creative use of "erosions" and "accretions" in the physical environment as indicators of past human actions. For example, defining smudges on the glass surrounding exhibits at a museum as units may enable one to make inferences from the smudges' frequency and height about the popularity of different exhibits among different age groups. Similarly, records in the files of an organization may be defined as units of study as indicators of the behavior of the individuals whose behavior is being recorded (say, a worker's production record), and also the individual doing the recording (a foreman or a manager).

The method of content analysis is specifically geared to dealing with physical residues of symbolic behaviors as in newspapers, novels, or popular songs. In such research, great care must be taken to define the unit of analysis and the sampling strategy because the unit may be either the content of the material (e.g., the crimes reported in newspapers) or different physical forms of the media (e.g., the entire issue of a newspaper, the front page, or the story or article). For example, in the 1960s, a number of studies of urban riots used newspaper accounts as the source of data while defining the riots themselves as the unit of analysis. Later researchers questioned the validity of this research by shifting the definition of the unit of analysis from the riots to the newspaper accounts or articles themselves (Danzger, 1975). Where the former raised substantive issues about riots and riot cities, the latter raised issues about the operation of the press and the role of the media in depicting civil disorders. Such simple questions as the physical proximity of the newspaper to where the event occurred may produce a distortion in the data, as recent research has shown (Myers and Caniglia, 2004). Again, how one defines the unit of analysis may raise different theoretical issues even when, as in the preceding example, the data set itself remains the same.

Settings, Incidents, and Events. In addition to individuals and their behaviors, there are settings, incidents, and events. All definitions of units and of samples imply space and time. However, this class of phenomena is explicitly defined using space and time coordinates. Settings may range from private rooms to expansive public places. Often the analysis of settings is accomplished through observation or the recording of observations with the use of cameras and video equipment. This unit also has been widely used in multimethod research, in which observation may be paired with the analysis of archives about events or incidents that may have occurred in different settings, or with interviews and surveys to tap people's attitudes and feelings in the settings.

When defined as the units of analysis, and not just as a convenient sampling frame for getting at other units, characteristics of social settings have proven to be important variables in their own right. Interest in the physical settings of behavior, or "situated action," is seen in the development of several subfields, such as "environmental psychology." Newman's (1972) work on "defensible space" is an excellent example of multimethod research using settings as the units of study. Newman was interested in the design and settings of buildings (specifically, high-rise public housing) and their relationship to crime and residents' fear of victimization. Settings were observed, crimes were analyzed through records, and residents were interviewed about their reactions. By combining these methods, Newman could explore a wide range of variables in the most efficient manner, systematically exploring propositions that had been hinted at but left unexplored in previous research.

In contrast with settings, which are spatially defined, incidents and events are temporally defined units of analysis. They have a definite beginning and an end, have a known and perhaps significant place in historical time, and are recurrent. Public ceremonies, elections, city council meetings, and urban riots are all examples of this type of unit. Of course, each of these phenomena can also be studied by defining other characteristics as the units of analysis. For example, one might define rituals, electoral candidates, council members' voting records, or the cities in which riots may occur as the relevant units. However, focusing upon incidents and events temporally, as historians do, for instance, orients the researcher to variables in data that highlight the ordering and sequencing of behavior, the causes of beginnings and endings, and the conditions under which phenomena recur.

Studying incidents and events, especially in their recurrence, leads the researcher to state dynamic propositions (or theories of process) and to search for data that will correspondingly capture the phenomena at several points in time. One immediately thinks of historical archives, census data over the years, or successive waves of interviewing in panel surveys as appropriate methods. Even the classical experimental design measures phenomena at two points, before and after the experimental intervention. Cross-sectional, one-shot surveys also often include retrospective data about the respondents and their experiences. Moreover, a combination of methods may be employed, and multiple methods may be especially useful in studying past incidents and events in comparison with the present. For example, Erikson (1976), in his postanalysis of a flood disaster, uses both interviewing and archival research to recapture the flow of events and to determine their impact on the community at the time of his study.

Collectivities. The final general class of units to be considered deals with human beings in the aggregate. Such analyses begin from the recognition that the whole is greater than the sum of its parts. Collective units—such as families, organizations, communities, professions, or nations—have unique characteristics or properties that cannot be arrived at simply by adding up the traits of the individual members. For example, individuals do not have power structures, only collectivities do.

Defining collective units often involves "boundary problems" of deciding where the organization or community ends, of deciding who is in it and who is not. One may take participants' definitions; for example, the list of an organization's members. But often, the members of a unit of social structure may be unaware of the collective unit itself or of its constituent members. As Kadushin (1966) has shown in his study of social circles, or as Fischer (1982) and Wellman and Leighton (1979) have shown in their studies of urban networks, circles and networks may have distinct and varying properties and be studied as units of analysis even when only the researcher (but not the participants) is able to objectively define them.

The problem of defining the boundaries of collective units of study is very similar to the problem of defining the universe from which to draw a sample. In both cases, care must be taken to establish logical and measurable criteria for membership. In the former case, the purpose is to define units for study; and in the latter, to define the appropriate universe from which to sample.

Furthermore, there are numerous ways to define most collective units of analysis. For example, the debate on how one should identify communities has occupied the literature of urban sociology for over half a century. The way in which a unit is defined is closely linked to the nature of the data one collects and the propositions one explores. However, different definitions of units can be used in the same research, with appropriate but different data sets being collected, and either different or the same propositions being tested. For example, Hunter (1974), in a study of urban neighborhoods in Chicago, defines community in three different ways and collects different data to test *different* propositions that relate to these different definitions of the units of study. In contrast with Hunter's work, Kasarda (1974) demonstrates how the *same* proposition may be explored across shifting definitions of the units of analysis. He tests the relationship between the size of social units and their administrative ratio for different types of units ranging from organizations to communities to nation-states.

The preceding discussion of types of units and corresponding research styles is presented not as a codified system matching one type of unit to a given type of problem and then to a corresponding "best method." We do stress that care must be taken to define units of study so that logical consistency will exist among types of units, the nature of the data, and the form of one's propositions. But our underlying theme is that a multimethod approach, because it gives access to different types of units and appropriate data, may increase the logical consistency of our work as well as suggesting new and important avenues for both research and theory.

Units of Observation and Units of Analysis

We can define units of analysis as those entities *about* which we collect data and about which we want to generalize or make inferences. Observational units may be defined as those units *from* which data are collected. Following common practice, however, a study's units of analysis may be different from its units of observation. For example, one may collect data from a housewife (unit of observation) about the size of her family (the family being the unit of analysis). Sometimes there is confusion at both data-collection and data-analysis phases about the unit of analysis, and it usually centers upon ambiguity as to whether one is making inferences about individuals or groups (collectivities), and this in turn is often related to the failure to distinguish between units of observation and units of analysis.

Ecological Fallacies, Structural Effects, and Hierarchical Embeddedness

Two methodological and analytical issues in the social sciences relate to this distinction between units of observation and units of analysis. The first is the often noted "ecological fallacy," or the inappropriate drawing of analytic inferences about *individual-level* correlations or relationships from aggregate data on *collective-level* units of observation. For example, if one finds that cities with a greater proportion of their population between ages 13 and 18 have higher crime rates than cities with a lower proportion of teenagers, one cannot conclude from this that teenagers are more likely to commit crimes than other age groups. This involves shifting from collective-level units of observation (cities' age distributions and cities' crime rates) to individual-level units of analysis (individuals' ages and individuals committing crimes). Recent statistical procedures have been developed to provide some "range" of estimates of individual level relationships from collective-level data (see Hammond, 1973).

A more direct way to deal with the problem is to use multiple methods to gather data at both individual and collective levels in order to permit analysis at both levels. This is not to suggest that the object of such research is always to obtain individual-level correlations. Sometimes the aggregate level may be sufficient and appropriate for the problem at hand (for example, selecting cities and using indicators such as age distribution in deciding where to target money for anticrime activity). However, selecting multiple units of analysis and using multiple methods of data collection is a direct way to deal with the problem. In addition, it may raise intriguing theoretical issues in explaining the similarity or differences in the relationships that are found between the two levels.

A second issue related to the distinction between units of observation and units of analysis is that of structural effects (Blau, 1960; Davis, Spaeth, & Huson, 1961). The same data may be collected on individuals and aggregated to a group level (e.g., age of individuals and *average* age for the group), and these may interact with one another in predicting to some individual-level dependent variable (e.g., committing a crime). The question of structural effects is often posed as whether individual characteristics or group characteristics predict behavior better. Structural effects may likewise be analyzed when data are collected on unique properties of the groups themselves that are not reducible to individuals and not simply aggregated from individual-level data (e.g., duration of the group versus an individual's duration as a member). As with ecological fallacies, multiple methods can be usefully applied to gather both the individual- and group-level data required to address this complex class of theoretical questions.

Third, the idea of ecological effects has recently been extended both statistically and methodologically by combining a number of different embedded levels of analysis and different and distinct data sets. Much of the current research on "neighborhood effects" in education, for example, is an attempt to disentangle personal and family variables (such as age or number of sibs) from school effects (such as average class size) and from local community variables (such as crime rate). Distinct data sets and distinct variables appropriate to each level of analysis are combined in estimating the unique contribution of each factor and also their joint effects or interaction effects on a student's educational outcome (Duncan & Raudenbush, 2001). Research by Sampson and Raudenbusch (1999) on systematic observation of street activity related to crime similarly uses an embedded approach in which the units of observation ranged from individual behaviors to physical characteristics of face blocks to whole neighborhoods and local communities. Data sources on the different levels or embedded units of analysis included videotaping of street behavior, coding of physical characteristics of the street block, and gathering of archival crime data and census data on the neighborhoods and local communities. Again, a multimethod strategy of combining different units of analysis and different data sources proved invaluable for capturing the complexity of the phenomenon being studied.

Types of Sampling

Social scientists employ two major types of sampling, each of which includes several subtypes. First, there is probability sampling, which includes simple random samples, systematic samples, stratified random samples, and multistage cluster samples. Second, there is nonprobability sampling, which includes quota samples, purposive samples, and convenience samples.

Probability Sampling

The key to probability sampling lies in our ability to assess the probability that a given set of units that make up the sample would be drawn from the universe of such units by chance. In making this assessment, we rely upon "sampling theory." This theory is based upon the *idea* that one could take repeated samples of the same population and compare the samples. For example, if you were to select different samples of 10 students from a population of 100 students, 50 of whom were males and 50 of whom were females, most of the time the randomly selected samples would have 5 males and 5 females. But sometimes there might be 6 males and 4 females, or vice versa. And rarely, though possibly, you might by chance select a sample in which all 10 were males. These repeated hypothetical samples from the same population are called the *sampling distribution.* In practice, we usually select only one sample. But given the hypothetical sampling distribution, we can assert that most of the time our sample will reflect the "true" characteristic of the population as a whole (in the preceding example, the percentage of males and females).

Probability sampling does not ensure that one truly knows what is going on in the population, but it provides a known probability of error. It allows one to say, in effect, I'm not absolutely sure that this is true, but I'm 95 percent confident that it is. With nonprobability sampling, there is also a possibility of error, but here we are left in the dark as to how much confidence to place in generalizing from what we have studied to what we have not. However, there are off-setting advantages, as we will see, to using certain types of nonprobability samples.

Simple random sampling, though heralded because of its seeming simplicity and because of its closeness to fitting the major assumptions of "probability sampling theory," is nonetheless one of the least used strategies. The primary reason is that often we have no way of identifying beforehand all the units that make up the universe, or population, from which the sample is to be drawn. If one can identify all the units, then the procedure is indeed simple. Merely assign a number in sequence to each unit, and then, by using a random number table, select those units whose numbers come up until one has the sample size desired.

Systematic sampling is a widely used probability sampling strategy that relies upon a list of units from the population, usually a list generated for other purposes, such as a roster of members of an organization or a list of registered voters. Rather than assigning a number to each member on the list and selecting the sample by using a random number table, one merely picks a random place to start on the list and then selects every *N*th unit (every fourth, fifth, or twenty-first name, depending both upon the size of the list and the proportion of the population one wishes to draw into the sample). One must be careful that the list does not display a pattern that matches the interval of selection such that a biased sample would result. Furthermore, one must be especially careful to understand how and why the lists were generated, who from the population might be systematically excluded from the list, and who might be selectively included. To justify the use of a given list and to understand the origin of the list often requires research by other methods, such as observation and interviewing of those who have generated the list.

Stratified random sampling can provide even greater representativeness of a population than simple random sampling. However, stratified random sampling presupposes another data source, because it requires some prior empirical knowledge of the population. If one knows the proportion of certain categories (or strata) within the population (such as the proportion of white or black), then by randomly selecting units within each category in the same proportion as in the population, one may actually ensure a more representative sample than might by chance be generated by a simple random sample. For example, it would not be possible, by chance, to end up with an all white sample as might happen with simple random sampling.

Multistage cluster sampling is a hybridized sampling strategy that relies on at least two different scales of "units" in which the smaller scale units (e.g., people) are distributed in a number of large scale units (e.g., neighborhoods or organizations). One first randomly selects a number of larger scale units as a first stage sample (e.g., a sample of neighborhoods or organizations). Then, from each of these larger scale units one selects a random sample of the smaller scale units. Often this sampling strategy is used to reduce the cost of data collection so that one does not need to go to every organization or every neighborhood to collect data on a sample that is statistically representative of the population. In effect, this is a multimethod strategy, in that different scales and different units of observation have been identified. However, too frequently these different scales of units are not fully included as such in the analysis. We have seen that certain significant empirical and theoretical issues such as "ecological fallacies" and "structural effects" might be explored by means of this sampling strategy.

Nonprobability Sampling

Quota sampling is a nonprobability sampling strategy that is a precursor to more fully developed forms of probability sampling and is explicitly concerned with trying to select a representative sample. Quota sampling assumes a known distribution of certain important characteristics of the universe or population, such as the proportion of male and female, white or black, in different age categories. One then selects people who have those characteristics until the sample proportions match the proportions of those characteristics found in the universe or population as a whole. Ideally, for example, one would end up with the same proportion of black men in their forties as there are in the population.

One of the major problems with quota sampling is determining which specific population characteristics should be taken into consideration in developing quotas. Often these are characteristics that prior research has shown to be significant in relation to the variables presently being studied. For this reason, quota sampling is very dependent upon prior research findings in the field. Another problem is that one cannot assume that the sample is representative of the population for the characteristics not explicitly taken into consideration. In the preceding example, one might have sampled those black men in their forties near a military base, and so they might overrepresent one particular occupation in the sample. However, if the sampled characteristics are theoretically significant, this may be presented as a concern overriding the statistical concern of representativeness. But if so, one is never sure that these associated but unsampled characteristics are not, in fact, the determining variables in one's analysis.

Like quota sampling, purposive sampling relies on the researcher's prior theoretical and empirical understanding of the universe with respect to the issue under study. Also, like stratified sampling, purposive sampling often attempts to include particular categories or subgroups of the population, but it differs in that rather than trying necessarily to select subgroups that are fully representative, purposive sampling may select only certain subgroups that represent theoretically meaningful variation. We will see, for example, that the testing of grounded theory relies on a form of purposive sampling in which the units selected are theoretically defined as important and not statistically determined to be representative. Purposive sampling, in short, is a claim on the part of the researcher that theoretically significant, not necessarily statistically significant, units have been selected for study. It is, therefore, incumbent upon the researcher to justify the selection process with a qualitatively different, but an equally rigorous, rationale as that used by those arguing for statistical representativeness.

Convenience sampling is perhaps the most widely used, but the least statistically or theoretically justified, sampling strategy in the social sciences. Convenience sampling, as its name implies, means that the researcher has studied a particular unit or set of units because it is readily at hand. For this reason, it is a relatively low-cost technique. Usually, convenience sampling is associated chiefly with participant observation and experimentation in which researchers frequently study easily accessible settings and subjects who are close to home.

However, convenience sampling may be defined more broadly to include any research in which the composition of the sample is determined primarily by consideration of the costs of access to the data rather than by one of the more rigorous sampling techniques discussed previously. It can even be extended to include a convenient definition of the universe from which a sample is drawn rather than a theoretical definition of the appropriate universe. So defined, a study based upon readily accessible organizational records, or easily available public data such as the census, or media reports of conflict may also be thought of as employing a convenience sample.

If convenience samples are employed for exploratory purposes or with the appropriate statistical or experimental controls for testing hypotheses, they may be quite valid within certain limits. However, a problem arises in that because they are convenient, the same types of samples tend to be overused. This limits both their potential for discovery in exploratory work and also their external validity in verificational studies. To remedy this, as Webb et al. (1966) have suggested, such studies may be augmented by using perhaps equally convenient but less stereotypic samples. With extensive data sources such as the census, this supplementation might seem to be unnecessary. However, as demographers know, even the census underrepresents certain segments of the population (e.g., the homeless) who may be of equal research interest and who may, therefore, need to be studied with other methods.

At yet another level, any sample may be considered to be a convenience sample to the extent that the researcher fails to sample from a theoretically defined, as opposed to a statistically defined, universe. A universe may be casually defined by convenience, while a sample of that universe may be rigorously drawn and defended statistically. From the viewpoint of testing theories and of accumulating theoretical knowledge, this is clearly an instance of misplaced concreteness. For example, researchers who have used the very rigorous sampling strategy employed in selecting individuals to be interviewed in the Detroit Area Studies of the University of Michigan would be hard-pressed to justify the selection of Detroit over any other American city on any other grounds but its proximity to Ann Arbor. Is this any less convenience sampling than the field researcher's decision to study the neighborhood surrounding her home or the experimenter's to study his students?

The multimethod strategy suggests that convenience must be addressed forthrightly as one among many rationales for choosing a particular sample of units for study and that its limitations must be recognized and compensated for. Heirich's (1977) research on religious conversion provides a good example of convenience combined with statistical and theoretical rigor. Over a period of seven years, he observed firsthand the growth of Catholic Pentecostalism near his university. Reviewing earlier research on conversion, Heirich (1977) notes that "Most were descriptive studies of odd clusters of converts, with little sense of how generalizable the description might be" (p. 657), and with no control group of unconverted persons to determine whether they too might be influenced by the same factors as the converts. To improve upon this earlier work, Heirich (1977) first drew a purposive snowball sample of local converts to Pentecostalism, attempting to interview the entire universe of members within his limited geographic area, by asking each respondent to name other converts. Then, to provide a control group, he drew a systematic probability sample of Catholic students (most converts were also students) from lists provided by the university and the Catholic students' organization. In this way, he was able to test hypotheses requiring comparisons between converts and a sample representing the pertinent population of nonconverts. He found that several hypotheses from earlier research on conversion were strongly supported by data from the converts alone but were then disconfirmed when tested with data from the control group as well.

Sampling and Generalization

In sampling, the central question is how generalizable, or representative, are the research results? Even if all other questions are adequately answered, a skeptic may still ask if one's findings hold true beyond the particular objects investigated in a given piece of research. An experiment may be valid for this class of introductory psychology students, but is it valid for other people? A field researcher's ethnographic account of family life in a slum community may be valid for that community, but is it valid for other slum communities? Or, a survey may be valid for a randomly selected sample of adults within a metropolitan area in a given year, but is it true of people in other metropolitan areas and at other times?

The multimethod perspective suggests new ways of thinking about these issues. First, both statistical and theoretical representation must be considered and counterbalanced in the sampling of subjects for study (neither purely statistical adequacy nor purely theoretical relevance is sufficient in itself). And second, the conventional dichotomy between probability sampling procedures usually associated with survey research and the nonprobability strategies associated with other research styles needs to be reconsidered.

These issues of different sampling strategies have recently come to the fore in discussions about survey research using the Internet. The fact that the Internet can be used to generate huge (20,000+) "convenience samples" relatively inexpensively when compared with traditional survey research samples forces one to confront the issues of bias and generalizability without denying the utility of such large data sets. James Witte (2004), a pioneer of Internet surveys, has explicitly called for a "multimethod strategy"—one that would systematically compare results from the convenience sampling of large Internet surveys with results from more controlled statistically representative random samples. Such multimethod sampling comparisons would allow one to simultaneously assess the generalizability and biases of the survey while still retaining the analytical power generated by its sheer size. Witte (2004) notes that "Methodologically, the goal of survey research is to collect data on a sample that represents a population. Randomness does not guarantee representativness; rather, it provides the means to quantify the level of confidence with which one can say that the sample does not represent the population. Survey 2001 [his large Internet survey] did not yield a random sample, and we do not "know" the selection probabilities for sample members. However, this *does not mean that the survey cannot yield representative social science data.* (italics in original). Although we do not "know" the selection probabilities our data allow us to "estimate" these probabilities. The survey collected data on standard demographic characteristics (e.g. gender, age, race, education), and combinations of these attributes can be compared with other data sources." (p. xviii-xix)

Theoretical and Statistical Representativeness

In atheoretical research, questions of representativeness are almost wholly statistical. For instance, with a public opinion poll, we would be interested solely in determining the probability of error in generalizing from the poll's sample to the larger population. But in theoretically oriented research, there is an additional consideration: How well does the population from which the sample was drawn satisfy the conditions under which the theory being investigated is thought to hold true? As Haas (1982) has observed, much social research presently fails to address this second type of question.

> Under the hypothetico-deductive model [of knowledge], a research report would describe the population studied in terms of the general, analytic variables of the theory being tested. The description would establish the fact that the population satisfied the conditions under which a specific relationship was predicted to hold. Sociologists, however, almost never describe the populations they study in any such way. Instead, they are described in ways that permit a reader to judge whether and to what degree they are representative of historically specific societies of interest, an approach that makes sense only under the survey sampling model. (Haas, 1982, pp.108–109)

In short, statistical rationales for sampling are limited by the failure to identify and define universes theoretically. To the extent that the pertinent theoretical universe differs from the population actually studied, there is likely to be analytical and predictive slippage between whatever generalizations are advanced or tested in the research and the data upon which those generalizations are claimed to rest. We do not mean to suggest here either that statistically rigorous sampling techniques should be abandoned or that a survey sampling model of knowledge should replace the hypothetico-deductive model (Haas, 1982). Rather, we suggest that a multimethod perspective on research must recognize the degree to which even the most rigorous techniques may be embedded in a less rigorous selection process, and must be as skeptical about the significance of statistically proper but atheoretical samples as about theoretically relevant but empirically unrepresentative ones.

Methodological Restrictions on Universes

One of the major advantages of survey research is its capability to generalize to the population from which the survey sample was drawn. A major drawback of fieldwork and experimentation has often been their inability to do this because the persons or groups studied were chosen by nonprobability techniques. In nonreactive research that employs archives and records, the difficulty is often in defining the actual population from which the information was assembled.

By adapting the sampling techniques of survey research to the particular needs of the other styles of research, these limitations can sometimes be overcome. For example, probability samples of experimental subjects can sometimes be drawn from the population to which one wishes to generalize the experiment's findings. And greater care can be taken to sample natural groups and settings or the individuals or events in them, as Heirich's (1977) research illustrates. However, while borrowing the sampling techniques of survey research may help to solve some problems of generalization in social research, there are other problems that it may not solve. As Webb et al. (1966) have pointed out, the model of survey research pays little attention to the fact that "only certain universes are possible for any given method. A method-respondent interaction exists—one that gives each method a different set of defining boundaries for its universe" (pp. 23–24). What this implies is that the special demands of each research style's data-gathering techniques, including the techniques of survey research, lead their practitioners to study selectively certain universes of persons and groups while putting others beyond their reach.

For example, fieldwork tends to be conducted principally among relatively peaceable people in lower- to middle-level social positions, because they are people who are both less able and less likely to defend the perimeters of their groups than others who are either more prone to conflict or more elite. The interviews and questionnaires characteristic of survey research presuppose, among other things, relatively high levels of articulateness and literacy. Laboratory experiments and even many unobtrusive field experiments require relatively cooperative and compliant subjects for their successful completion. And the use of official statistics and archives may lead to study of those who have attracted, or at least failed to avoid, official attention, but may direct study away from either less prominent or more elusive people.

An important implication of these methodologically related population restrictions is that even if a method's sampling procedures are improved, the generalizability of its results may yet be questionable, because the *universes* from which the data are obtained may be systematically biased by the constraints of the characteristic data-collecting techniques. An unbiased sample of such a biased universe can be as misleading as a poor sample, especially if the nature or existence of the population bias is unknown. Fortunately, however, the different data-collection techniques associated with each style of research provide the opportunity to overcome these methodological restrictions on universes.

Dimensional Sampling

Sampling from a population to collect data and generalizing from a sample to a population are distinct but obviously interrelated sides of the same coin. However, there are stylistic differences among social science methods that emphasize varying degrees of precision and timing with respect to these issues. Field researchers often select a group because it is of interest or is accessible, and consider it sufficient to generalize from this group with decreasing confidence from a small number of similar groups to a larger number of different groups. Survey researchers, by contrast, are more interested in the precision of the sampling and in generalizing with greater confidence to a known population. Field researchers justify their lack of precision in sampling and generalization by pointing to the richness of data collected on a small number of units or even a single unit (the case study). They emphasize the synergistic "gestalt," and their more complete understanding of the structure and processes of the single case. Survey researchers justify the limited data they collect on their units (often data on individuals are reduced to a limited number of questionnaire items) by pointing to their confidence in the representativeness of what they have found. The field researchers are "hedgehogs" (they know a lot about a little), while the survey researchers are "foxes"—they know a little about a lot.

Several attempts at accommodation between the hedgehogs and the foxes have been proposed. Glaser and Strauss (1967), for example, proposed the "constant comparative method" as a strategy of sampling units in field research—a strategy that directly addresses the sampling question of what group or situation should I next study—or where do I go from here? Depending upon one's varying analytic goals, one selects either a similar group or a greatly different group. The strategy can be repeated ad infinitum until one has exhausted cases for all the theoretical dimensions that have emerged as significant in the research.

Arnold (1970) has suggested a somewhat different approach, which he calls *dimensional sampling.* This approach combines the survey researcher's concern about precision in defining units and selecting samples prior to data collection with the field researcher's concern about gathering "rich" data on a few selected, comparative cases. In dimensional sampling, one first selects the salient theoretical dimensions that have emerged in the literature and then uses these to construct an "attribute space" that defines a theoretical sampling frame or universe. This process is much like the survey researcher's attempt to define a population or a sampling frame in order to select a probability sample that will ensure unbiased representativeness prior to actually collecting data. In dimensional sampling, Arnold argues, one is much more likely to get an exhaustive, theoretically representative sample than the traditional comparative case study approach used in most fieldwork.

An example of the use of dimensional sampling is Hunter and Fritz's (1985) research on the power structures of community elites. Most studies of community power structures have taken one of two sampling strategies: the case study approach—as in the work of Floyd Hunter (1953) and Dahl (1961)—or a much larger sample of cities with fewer survey type variables—as in the work of Clark (1968). Hunter and Fritz (1985) combined these approaches by selecting four communities from a theoretical sampling space defined by two cross-cutting dimensions shown in previous research to be related to variations in community power structures, community size and complexity versus the social class composition of the community. The four communities they selected consisted of one which was small and poor, another large and poor, a third small and rich, and the final one large and rich. They then conducted a systematic survey of elites in each of the four communities and explored variations in power structures related to these two critical dimensions.

Sampling and Synecdoche

Sampling is generally viewed as a process of selecting one or more cases for investigation for the purpose of generalizing to a larger universe of such cases. That is, one selects a few individuals, neighborhoods, or organizations in order to say something about a larger class of similar individuals, neighborhoods, or organizations. However, sometimes one selects cases in order to say something about a larger whole of which the sampled units are seen to be a functioning part, a microcosm. For example, one may study a number of groups, organizations, or institutions, as William Foote Whyte did in *Street Corner Society* (1955), in order to say something about a larger social unit of which they are a part. Whyte, for example, was not interested in simply generalizing to a universe of other small groups in Boston, or even small groups in American cities in the mid-twentieth century, but rather he was interested in saying something about an Italian slum community. In such a case, the groups are not just part of a statistical aggregate consisting of some universe of similar small groups, but instead they are seen as functioning parts of a larger whole, be it a neighborhood, a city, or an entire society. Similarly, Robert and Helen Lynd (1929) were not interested in studying *Middletown* (Muncie, Indiana) simply to generalize to a large universe of middle-sized cities in the United States, rather they had selected Middletown as prototypical, a functioning part or microcosm of American society, which, if they could understand it, would enable them to make generalizations about American civilization as a whole.

The use of such microcosms is not uncommon in social science research. Even if a large-scale sample survey is done throughout a single metropolitan area, analysts will often generalize their findings not only to other metropolitan areas (a larger universe of similar units) but to American society as a whole. We can refer to this type of sampling as a type of metaphor, namely *synecdoche,* in which the part is used to stand for the whole. For example, when a ship's captain shouts, "All hands on deck!" he does not expect to see a science fiction scene of disembodied hands scurrying about, but the sailors themselves fully connected to their functioning hands.

Generalizing by synecdoche is not, however, simply metaphor. It is a claim that the essential features of the larger social unit are reproduced in microcosm within the smaller social unit, and that by studying them in micro we might make inferences about the macrostructure of which they are a part. In this, it is not unlike the older theory of human reproduction, the homunculus, wherein human beings were completely formed but on a vastly smaller scale within the germ cells of their parents. In social science research, we might refer to this as socioculus.

Claims of generalizibility made by synecodoche stress not simply a statistical representativeness but most importantly a functioning parallelism, and such claims often include an explicit set of functioning linkages between the larger whole and the smaller part. This is the point at which sampling crosses into the realm of what we may call *contexted* sampling.

Claims for representativeness and generalizability stress that such and such is true in this given context or under this set of limiting and specified conditions. The argument is that, of course, one might expect to get different findings by focusing on a different part, by taking into consideration a different context, or a different set of conditions. These posited or asserted—but often unexplored—contexts and conditions are the constants, not the variables, in most analyses. However, these contexts and conditions can be turned into variables; all that is required is a selection and sampling strategy at a larger scale of analysis.

These observations on synecdoche and contexted sampling are themselves not unlike recent developments in the *fractal geometry* of Mandelbrot sets. The same question is seen to generate a pattern comprised of smaller units that reproduce the pattern of the larger unit, and the larger unit itself is seen to be a part of a still larger unit that has a similar pattern. Though the patterns are similar, they do show local variations at different scales. The fact that there is a similarity to the patterns across the different sizes of the units is called *scaling.* Furthermore, a most intriguing aspect of fractals across these different scales is that they are, throughout the scales, functionally connected. Perhaps there is a parallel "fractal geometry" to social structure.

Sampling Throughout the Stages of Research

A central thesis of this book is that multimethod research means more than simply triangulation or multiple methods of measurement. It is a perspective that permeates all stages of the research process from initial theoretical hunches to final publication. The same may be said of sampling. If sampling is seen as a rational selection process that has implications for the truth claims of one's research, then sampling is going on all the time. It is not restricted to the process of selecting units for observation, although this is its more technical and limited meaning. Sampling also enters into theory, in the selection of one or more general paradigms from among a universe of such paradigms; in the initial selection of concepts from a universe of concepts within a paradigm; and in the selection of a few key testable operationalized hypotheses from among the many that might be delivered from a single more abstract proposition. Furthermore, not only do we select units within universes, the more limited domain of sampling, we also select universes themselves. Multistage cluster sampling may be thought of as the social scientist's analogue to both the poet's and the physicist's recognition that there may be multiple universes out there. We could continue this logic to include selection of measuring instruments, items on a questionnaire, and ultimately the selection of the forms and outlets for publication (few researchers randomly distribute their findings). We are indebted to Allan Schnaiberg, who shared his ideas on this broadened meaning of sampling that he has developed in his course "The Logic of Social Inquiry" at Northwestern University.

Perhaps we have overextended the idea of sampling here, to the point where it is equated with mere selection and choice. However, we have done so purposefully. The rigor and precision with which people continue to debate various sampling strategies for selecting their units of observation should apply equally to the sampling that goes on throughout the research process. Viewing choices and selections at all stages, those critical decisions that have implications for the validity of one's research, as sampling questions to be considered in relation to a universe of options, may promote both a broader search and a more rigorous justification for the choices we finally make.