AMERICAN SOCIOLOGICAL REVIEW

THE PRESIDENTIAL ADDRESS: MEASUREMENT AND CONCEPTUALIZATION PROBLEMS: THE MAJOR OBSTACLE TO INTEGRATING THEORY AND RESEARCH*

H. M. BLALOCK, JR.
University of Washington

American Sociological Review 1979, Vol. 44 (December):881–894

In one sense the theme of this paper is obvious. Sociologists face extremely tough intellectual and practical tasks owing to the ambitious nature of our common objectives and the complex reality with which we deal. These tasks will require a *concerted* effort of scholars with diverse substantive, theoretical, and methodological interests and persuasions. Yet in many respects we seem badly divided into a myriad of theoretical and methodological schools that tend to oversimplify each other's postions, that fail to make careful conceptual distinctions, and that encourage partisan attacks.

Rather than dwelling on these divisive issues within our profession, it is crucial that we learn to resist overplaying our differences at the expense of common intellectual interests. There will obviously be disagreements over appropriate strategies, as well as ideological and subdisciplinary differences. But an idealization of conflict and dissensus is self-defeating. Some conflicts will inevitably occur and, if constructively resolved, may result in benefits to the discipline. But I think there has been too great a tendency to exaggerate these benefits, without recognizing the inherent dangers of endless theoretical and methodological debates and a further fractionating of our field.

One particularly disappointing feature of our discipline is that we have not had the productive interplay between theory and research called for so eloquently by Merton (1968) several decades ago. This interplay, if it ever comes about in a systematic way, will require us to grapple with a number of extremely complex problems that I shall merely list before narrowing my remarks to two issues that illustrate the need for analyses that are simultaneously theoretical and methodological. My list is as follows:

1. Reality is sufficiently complex that we will need theories that contain upwards of fifty variables if we wish to disentangle the effects of numerous exogenous and endogenous variables on the diversity of dependent variables that interest us.

2. Many social changes are either very rapid compared to the intervals of observation or are continuous rather than discrete, so that temporal sequences cannot easily be inferred or linked to given historical events.

3. Realistic models of naturally occurring social phenomena must be nonrecursive or contain highly specific assumptions about lag perods or distributed lags.

4. Many important theoretical variables are highly intercorrelated, though perhaps the empirical associations among them will be underestimated due to random measurement errors. Resolving this multicollinearity problem will require a combination of large samples and good measurement.

5. Human actors and social systems tend to be nonhomogeneous with respect

*Address all communications to: Hubert M. Blalock; Department of Sociology; University of Washington; Seattle, WA 98100.

to parameters in structural equations, implying that they will not respond similarly to changes in other variables. This will have major implications not only for our theories but also for measurement decisions, whenever effect indicators are being used, and for micro–macro analyses where aggregation decisions are needed.

6. Many groups and contexts have fuzzy boundaries. Standards, such as group norms or role expectations, also tend to be imprecise and subject to dispute. Measurement that depends in some essential way upon these fuzzy boundaries or standards thereby becomes exceedingly difficult.

7. The linkage of micro- and macrotheories involving different units of analysis is problematic unless simplifying assumptions can legitimately be made. In particular, aggregation–disaggregation problems are made difficult whenever there are nonnegligible contextual effects, nonlinearities, or unknown measurement errors.

8. All measurement is to some degree indirect and therefore requires untested assumptions of a causal nature, but this problem is especially serious whenever one-to-one linkages between constructs and indicators cannot be assumed, whenever replications under standardized conditions are impossible, whenever homogeneity properties facilitating indirect measurement cannot be assumed, and whenever the ratio of unmeasured to measured variables is high.

9. Given the practical roadblocks to data collection that will continue for the foreseeable future, any piece of research will necessarily involve large amounts of missing information, thereby requiring either implicit or explicit assumptions and the neglect of numerous variables thought to be theoretically important.

Although the development of theory is important in its own right, I believe that the most serious and important problems that require our immediate and concerted attention are those of conceptualization and measurement, which have far too long been neglected. I have reached this conclusion having come at the matter from two very different perspectives. The first is through an examination of the implications of random and nonrandom measurement errors for data analysis and theory testing, and the second is through frustrating efforts to make sense of the theoretical and empirical literature in one of our substantive fields, that of race and ethnic relations. Both these endeavors leave me with the realization that these conceptualization and measurement problems are much more complex than I had previously thought. In fact they are so complex, and their implications for analysis so serious, that I believe that a really coordinated effort in this direction is absolutely essential.

Clearly, we need theories that are sufficiently general to integrate our fragmenting discipline into reasonably coherent bundles. These theories must be precise enough to yield predictions that are both falsifiable and that extend beyond common sense. We might hope that our theories and analyses can also be reasonably simple, but for reasons elaborated elsewhere (Blalock, 1979) I do not believe we can simultaneously achieve generality, accuracy, and simplicity. Therefore we must give up one or another of these desirable characteristics. If we opt for simplicity, and if social reality is in fact complex, we shall inevitably be misled.

Given the limitations imposed by our meager resources and missing data, it is crucial that we carefully examine what these imply in terms of linkages between theory and research. Missing variables force us to use highly indirect measures, improper aggregation operations, and crude background factors as indicators of experience variables. For practical reasons, many of these missing variables must remain unmeasured. Thus we must substitute a series of implicit or explicit assumptions about how these variables operate. But assumptions can either be made blindly for convenience or after one has carefully tried to identify the missing variables and think through their implications for the theory in question. The latter course is much more frustrating and disillusioning, but it is the surest way to make genuine progress in pinpointing inadequacies in existing theories and data.

In the sections that follow I shall discuss two very different though serious problems that illustrate the complexity of the type of analysis that I believe is needed in the face of such missing information. These problems are: (1) the plethora of theoretical definitions of generic behaviors and their implications for measurement, and (2) the confounding of measured and unmeasured variables when individuals are aggregated in macrolevel analyses.

Both types of problems illustrate an important kind of temptation, namely that of substituting relatively simple operational indicators for theoretical constructs without paying careful attention to the underlying measurement model and required simplifying assumptions. In the case of behaviors, we note a tendency to define variables theoretically so as to facilitate generalizability at the expense of realism with respect to simplifying assumptions. In the case of aggregation by spatial criteria we encounter the need to specify unmeasured variables and mechanisms linking location in physical space to whatever dependent variable is being investigated.

The researcher, constrained by serious data limitations, usually finds it convenient to sidestep these issues. The theorist, trying to make sense of diverse empirical studies, is then confronted with an almost hopeless task and may be tempted to use the empirical information either selectively or anecdotally—or even to ignore it altogether.

THE MEASUREMENT AND CONCEPTUALIZATION OF BEHAVIORS

We have recently made considerable progress with respect to data *analysis* but relatively little with respect to data *collection*, and in particular our ability to observe, categorize, and measure behaviors. Even if one does not accept this assertion, I assume there is consensus on the need to improve our measurement of behaviors. There inevitably will be numerous practical obstacles to observing behaviors as they actually occur, but the problems I shall discuss are conceptual or theoretical and would occur even under the most ideal circumstances.

Human behaviors are extremely diverse, so much so that if we were to try to explain each one separately the situation would become hopeless. One way to resolve this problem would be to limit ourselves to a very restricted number of behaviors, but this is obviously not the course we are following. In a few instances behaviors pose no special measurement problems aside from observability. For the most part, these revolve around simple biological or economic needs. Certain other task-related activities are also rather directly linked to these basic needs, so that even though they may vary from culture to culture, they are easily classified.

Many behaviors of greatest interest to sociologists, however, are not of this nature and tend to be confounded in the real world. This is particularly true for behaviors that require a social partner or that are conditional on properties of a social system. We recognize that there are many different "forms" of these behaviors, which are often very different in terms of manifest or directly observable characteristics. Thus one may achieve status in a variety of ways, such as killing enemies, saving lives on the operating table, tackling opponents on a football field, or making vague political promises.

How do we get a theoretical handle on these diverse behaviors so as to group them into a much smaller number of conceptual ones? Although there are undoubtedly more, I am aware of four strategies, all of which rely on theoretical assumptions that usually remain implicit: (1) a linkage is assumed between the behavior and some motivational state, which usually appears in the theoretical definition; (2) there is an assumed causal linkage between the behavior and some consequence, which is an integral part of the definition; (3) the behavior is defined in terms of some general social standard with which it is compared; and (4) there is an assumed linkage between the behavior and other variables that cause this behavior to be repeated, with replication being an essential component of the definition.

Each of these definitional strategies thus requires simplifying assumptions that will be more or less realistic, depending upon the complexity of the setting, the motivations of the actors, and the reactions of other actors who may also affect outcomes or ways in which behaviors are repeated. We should therefore not be surprised to find each definitional strategy being accompanied by certain theoretical biases that help the social scientist justify whatever simplifying assumptions are most convenient for that strategy. The more complex the behavior, the more crucial it is to uncover such biases and to state assumptions explicitly.

1. Behaviors Defined in Terms of Internal States

The lack of a perfect correspondence between attitudes and behaviors has been well documented. But it may not be so obvious that many general types of behaviors are *defined* in such a way that some internal state becomes an essential ingredient in the definition, so that measurement requires assumptions about these internal states. For instance, "aggression" may be defined as behavior intended to injure another party or "altruistic behavior" as any form of behavior intended to benefit someone other than the actor, without regard to the consequences for that actor. "Avoidance" may be defined as any behavior the purpose of which is to reduce contact with another actor, and "exploitation" may be conceived as the use of another actor for one's own ends.

What simplifications seem necessary in using this definitional strategy? Let us illustrate with the aggression example. If there were a closed set of behaviors that could be listed, each of which is clearly linked to the injury of another party, one could supply the observer with the names of these familiar behaviors. Two kinds of difficulties are encountered, however. First and most important, many behaviors serve several ends at once. In fact human beings are remarkably adept at creating situations in which actors can kill several birds with one stone. Aggression may be instrumental in weakening the competitive position of an opponent or in attaining status among one's peers. This means that the same behaviors can be classified in different ways if the theoretical definition is stated in terms of a postulated internal state. Thus several different types of behaviors may be hopelessly confounded. One theorist may refer to aggression, a second to exploitation, and a third to competitive behavior, all "observed" in the same way.

The second obvious problem is that the relationships between the internal state and the behavior may be more or less direct, may involve differing time delays, and may be subject to distortion by either the actor or the observer, whose theories of social causation may differ. Some forms of aggression are very overt, immediate, and nonsubtle. Others are very delayed, so much so that the observer may not be around at the time they are enacted. Still others may be subtle and disguised as behaviors of a different type. In yet other instances, the actor may intend to injure another party but may fail because he or she does not adequately understand the motivation of that party. The observer, too, may misread the intent of the first actor.

Given these difficulties, what kinds of simplifications are we tempted to make? First, we may confine ourselves to simple laboratory settings in which actors' choices are restricted to a small number of alternatives, each of which is assumed simply linked to a postulated internal state. Aggression, altruism, or avoidance thereby may be identified with relatively simple operations such as that of pushing a blue rather than a red button. The measurement–conceptualization problem is then transformed into one of assessing "external validity" or generalizability.

A second simplifying strategy uses a restricted subset of behaviors most simply linked to the internal state referred to in the conceptual definition. If some forms of aggression are subtle, indirect, or delayed, these are excluded from the operational definition because they are difficult to interpret. The result is a nonrandom selection of behaviors biasing the findings in unknown ways. If educated persons are more likely to use subtle forms of aggression than less educated ones, then aggression among the former group will be underestimated.

A third strategy is closely linked with the second. One may limit oneself to a small number of "master motives" that underlie nearly all human behaviors, so that whenever a behavioral form can be located that supports one's theory, this master motive is invoked in labeling the behavior to confirm this theory. Thus if one believes that an intent to injure others is present in almost all human interactions, then nearly every form of behavior can be considered as a subtle form of aggression. Similarly, it is possible to infer an exploitative motive behind almost every behavior. Those who see status–seeking as a prime motivator may define whatever behaviors they see as instances of status–seeking, and to some degree they will probably be correct. Since most behaviors in complex settings involve mixtures of motivations, there is a wide–open opportunity to label any given behavior as an instance of many different kinds of generic behaviors defined theoretically in this fashion.

A fourth way to simplify the classification of behaviors is to accept the actor's word for his or her own motivation. Rarely are we so naive as to believe a respondent who claims a pure motive, but in effect this may be what actually occurs whenever we ask a respondent or witness to recall what has taken place. To do so, one must rely on popular vocabulary and common definitions, rather than scientific usage. This may result, to an unknown degree, in a generalization process involving the substitution of inferred generic terms for "directly observed" behaviors (such as a blow or spatial movement). The social scientist wishing to give precision to behavioral concepts that have popular meanings is thus faced with a dilemma. Either one must rely on popular definitions when events are being reported, or one must develop a more precise terminology that does not correspond to this popular usage. Whenever one wishes to generalize across cultures or languages, these problems become even more serious.

A fifth temptation is to use clues based on past behaviors to infer motivation. For instance, if an employer has not recently hired a sufficient number of blacks and then claims that current efforts are sincere, though unsuccessful, it may be inferred that this employer is discriminating against blacks (defined in terms of differential treatment *because of* race) if the ratio of black to white hirings does not reach some predetermined level. It may be, of course, that the past record also was a resultant of forces beyond the employer's control. The point is that intent is inferred on the basis of past practice or results, but without an explicit theory allowing for alternative explanations.

Whenever one wants a measure of the intensity of a behavior, we may add a sixth temptation to the list. This is to use an objective measure of the behavioral intensity or duration, without partitioning this among the several underlying motivations. For instance, suppose the intent to injure another person is only a very minor component of the actor's motivation. Perhaps self-defense or a desire to reduce competition is the major goal. Degree of aggression may be measured solely in terms of the behavioral act, thus bypassing the problem of measuring motivational strength (or utility) independently of the behavior itself. This may then affect one's theoretical interpretations. Suppose, for example, that the decimation of American Indians by white settlers was primarily based on the utilitarian goals of securing more land or protecting one's family. It would then be misleading to build an explanatory model, representing these actions as extreme aggression. The major point is that the relative importance of different underlying motivations needs to be kept distinct from that of the frequencies and intensities of behaviors unless it can be assumed that behaviors and motivations can be linked in a simple one-to-one fashion.

2. *Behaviors Defined in Terms of Consequences*

One may sidestep the problem of identifying and measuring internal states by focusing entirely on the consequences of the behaviors. Our discussion of this alternative strategy can be more brief since the issues once more illustrate our main point that there are numerous theoretical assumptions needed to carry out this kind of measurement strategy whenever the social situation is at all complex. First, *someone* must assess the consequences since a causal theory is being invoked to link the behaviors with some set of outcomes. But whose theory? The actor's? Other parties' in the situation? The supposedly neutral observer's? And which outcomes and using what time span? If there are both short-term and long-term consequences that are not identical, which should be used? And what if these outcomes are conditional on the behaviors of *other* actors in the social setting?

Once more, there will be pressures to simplify. One possibility is to confine oneself to very simple situations. Another is to dichotomize consequences as either occurring or not, thereby ignoring variations in degree. Still another is to ignore multiple consequences and look only at those consequences that are obvious and immediate in a temporal sense. Taking the example of aggression (now defined as behavior that results in injury), this would rule out many forms of delayed aggression or subtle types where the consequences seemed to be highly indirect. Again, this will result in biased measures to the degree that not all actors employ the same forms of aggression.

Whenever the consequences are conditional on the responses of other actors it will be especially tempting to simplify one's causal theory to obtain an unambiguous measure of the behavior. Suppose one defines discrimination as behavior that results in unequal consequences for classes of actors defined in certain ways, as for example by age, sex, or race. Suppose an employer makes a set of judgments that results in the hiring of disproportionately few blacks. Was the lack of hiring solely a consequence of the employer's action or also of the behaviors of the applicants for the position? It is tempting to try to get off the hook by crudely matching blacks and whites on "relevant" variables, as defined by the investigator, usually in accord with data availability considerations.

Furthermore, if the discriminatory behavior leads to some sort of response that jointly affects the outcome, then how does one define or measure the behavior without taking this response into consideration? How does one measure teaching effectiveness or leadership quality? The most tempting resolution is to assume away the problem by taking the second actor's behavior either as being totally dependent on that of the first or as having negligible consequences in its own right. Thus we often assume that minorities, children, and other relatively powerless actors are totally powerless, so that their own responses can be ignored.

Basically, this measurement strategy may tempt one to ignore all sorts of intervening and conditioning variables by grossly oversimplifying the causal connection between the behavior in question and the consequences that are being identified. There will be a vested interest in simplifying this set of consequences, just as the prior strategy creates one in simplifying the actor's motivational structure.

3. Behaviors Defined in Terms of Standards

Certain kinds of behavior are defined theoretically in terms of some social standard, which is often either rather vague or differently defined by actors having contrasting interests. For example, deviance is defined in terms of departures from social norms, which may be subject to dispute. In the case of criminal behavior the norms may be clearly stated in the form of laws that are enforced by official sanctions, but the laws themselves may vary from one jurisdiction to the next. Similarly, the notion of exploitation in an exchange relationship may be defined in terms of some standard by which equity can be evaluated. There are also a number of popular terms such as "mentally disturbed," "addiction," or "antisocial" behavior that presumably imply some sort of implicit standard.

In all of these instances an investigator who attempts to measure the degree of departure from such standards is faced with a dilemma. If reality is fuzzy, how is it possible to obtain precise measures? We have, it seems, a kind of sociological Uncertainty Principle that places an upper limit on the accuracy of measurement of such behaviors. How can one measure degree of conformity to imprecise norms? What if actors define a "fair" rate of exchange differently? Is there any meaning to a notion such as exploitation? The terms "conformity" and "exploitation" can be used in ideological writings, but can they become a legitimate part of a scientific vocabulary?

I believe it is possible to retain the essential features of the theoretical arguments that use such concepts, provided we make careful distinctions and somehow build the degree of fuzziness into these theories, as a separate variable. Whenever there is dissensus on group norms or on what constitutes a fair rate of exchange, this in itself becomes a datum of relevance to actors. Perhaps a measure such as a standard deviation can be used as a measure of such dissensus whenever the issue is unidimensional. When it is not, this in itself requires analysis because it will constitute an additional source of fuzziness for the actors concerned. Where the standards for a given subgroup are clear-cut but distinct from those of another, two separate variables can be delineated, as for example degrees of deviance from Group A norms and from Group B norms.

The temptation, here, is to substitute more precise standards for the true but fuzzy ones. One way to simplify the situation is to substitute some measure of average behavior for the norm, thereby giving it a definite meaning, although one that may differ from its meaning to the actors themselves. As is well known, there are two meanings to the word "normal," namely some measure of central tendency, on the one hand, and some idealized value, on the other. Insofar as these may differ according to the situation, our theories will then need to make the necessary distinction between the two types of standards.

Another alternative is to confine our operational measures to absolute values, using zero as the comparison point. Thus one may take suicide rates as a measure of deviance, but only if all suicides are socially defined as contrary to normative expectations. If certain suicides are not defined in this fashion, however, and if normative standards vary across the units being compared, then clearly suicide rates will not be an appropriate indicator of deviance. Unfortunately, many of our theories of deviance are not very precise as to the standard about which deviance is to be measured or whether the norm is to be defined in terms of a measure of central tendency or some legal or ethical standard.

4. Behaviors Defined in Terms of Replications

The fourth strategy, that of relying on replications, seems most common among behavioral psychologists and social psychologists who rely heavily on experimental designs involving repeated measurements. Given very simple settings and assumptions about motivating factors, such a strategy indeed makes sense. In generalizing beyond the laboratory setting, one obviously cannot rely so heavily on operational definitions of behaviors that require such replications. For example, if one defines reinforcing behaviors as those that are followed by later instances of the behaviors they are supposed to reinforce, one must rule out other causes of the replicated behaviors. Perhaps the behaviors are repeated because they are constrained or influenced by factors unknown to the investigator.

The more general point is that whenever several variables jointly affect a behavior, a reliance on the replication operation to measure a behavior will lead to both theoretical ambiguities and also empirical irregularities that make measurement much more difficult. In short, the research operations cannot be generalized readily to more complex situations in which replications occur under much less controlled circumstances. In making comparisons across such situations, both the measurement operations and the situations themselves will vary, so that theory and measurement become hopelessly confounded.

Whenever manifestly similar behaviors are rarely repeated in real-life situations we are faced with another kind of dilemma, the resolution of which will require theoretical assumptions. The observation period, being arbitrary in most instances, may in part determine the relative frequency of occurrences within a given population. If this proportion is very small, one will be confronted with a highly skewed response variable. This may be countered by defining the behavior in question as merely an instance of a larger class of behaviors that occur more frequently, but then problems of aggregation and homogeneity will arise. That is, the diverse behaviors that have been lumped together into the class may have different sets of causes or consequences.

Another alternative is to aggregate over individuals assumed to be similar in certain respects, so that one then works with behavior rates as estimates of probabilities of engaging in the behavior. Obviously this requires a well-defined theory as well as data sufficient to classify such individuals into categories that are homogeneous with respect to the parameters of the equations and not merely a set of "objective" attributes of individuals, such as age, sex, or SES. Often these aggregating decisions are made on the basis of convenience or convention, with the theoretical rationale being only implicit.

Finally, one may lengthen the time span so that behavioral acts of a given type will be more frequent. But this not only causes inconvenience for the observer but also is likely to introduce heterogeneity into the situation. The individual's motivation may have changed, the method of data collection or observational procedure may have to be modified, and situational factors may also be changing. The relative gains and costs of these alternative resolutions will of course have to be assessed for each particular case, and this will require a number of untested theoretical assumptions.

To conclude this section on behaviors, in considering the implications of each of these definitional strategies the essential point is not that assumptions can or should be avoided but that they need to be made explicit. Furthermore, we see that each measurement strategy requires the use of theoretical assumptions, only some of which can be tested. Our own experience (Blalock and Wilken, 1979) in attempting to analyze selected basic concepts in the field of intergroup relations is that an apparently simple form of behavior, such as discrimination, aggression, or avoidance, requires for adequate conceptualization auxiliary measurement theories containing as many as twenty or thirty variables. I would be surprised if the same does not hold for other reasonably general social behaviors. It is no wonder, then, that our rate of progress in conceptualizing and measuring these behaviors has been slow and uneven.

THE CONFOUNDING OF VARIABLES IN AGGREGATING BY GEOGRAPHIC PROXIMITY

The literature on aggregation and disaggregation is both technical and discouraging in its implications, if one takes seriously the goal of integrating microlevel analyses, based on the individual as unit of analysis, with macrolevel studies where groups are the focus of concern.[1] Ideally, theories on the one level should be consistent, in some sense, with those on the other (Hannan, 1971:18–23). Furthermore, since some groups are nested within larger ones, and since in many instances group boundaries are fuzzy and therefore arbitrarily defined, it is also desirable to pass systematically from one aggregate level to another, as for example from counties to states.

In discussions of aggregation in the econometrics literature it is assumed that those who do the aggregating have a theoretical rationale for grouping individuals into behaviorally homogeneous aggregates. In most instances where sociologists use aggregated data, however, the grouping operation has already been done, usually with another purpose in mind. In these instances aggregation can hopelessly complicate one's analysis unless the criterion for aggregation can be fitted rather simply into one's theory. For instance whenever we are dealing with a corporate group as a unit of analysis it

[1] For three very different, though complementary perspectives on the aggregation problem, the reader is referred to the works of Firebaugh (1978; 1979); Hannan (1971) and Hannan and Burstein (1974); and Irwin and Lichtman (1976) and Langbein and Lichtman (1978). These sources also contain numerous additional references.

makes good sense to aggregate over individual members to obtain measures of group properties. Presumably, our interest will center on this group and other comparable groups as actors, as for example whenever business firms produce tangible products or state legislatures enact laws or allocate budgets.

In many other situations the picture is not this simple, however. Sometimes we may aggregate over a territorial unit that for certain purposes may be considered a corporate group (e.g., a state or county), but where the corporateness may not be an essential feature of the theory in question. For example, we may be studying crime rates in various counties, where county–level policies have virtually no impact upon these rates. Or we may have segregation indices based on block data that are available only for a central city, whereas the SMSA extends far beyond these arbitrary political boundaries. Or our theoretical interest may be on the microlevel, say, in understanding why individuals commit suicide or tend to avoid members of another group. Yet the data may be available only in aggregated form, as for example census tract data. In no sense can these territorial units be said to constitute true groups, nor is there any pretense that we are interested in highly coordinated behaviors.

In such instances we use the aggregated data because they are the only ones available. What can we say about the problems created when individuals are aggregated by spatial criteria? The answer depends upon the causal connections between the criteria used in grouping and the variables that appear in our theories (Blalock, 1964; Hannan, 1971).

The usual assumption is that the aggregation criterion, which we shall call A, is an independent variable in the model and that it is not operating to confound the effects of the independent variables under study. When we acknowledge the myriad ways in which spatial location may be linked to the variables of interest to sociologists we can anticipate the complications that such aggregation may produce. People are influenced not only by what goes on around them in the immediate present, but also in the past. They may have moved from one community to another, carrying with them those effects in the former residence that we refer to as "background influences." Furthermore, not all individuals are affected in the same ways by the variables in their immediate environment. Some may have lived in the area all their lives. Others may have moved into the area because of its local traditions, whereas others may have entered and resisted them.

To come to grips with the problems that such complexities create, it will be helpful to examine several models that are themselves oversimplifications of the actual processes at work. We begin with a model in which it is presumed that the territorial units are closed to migration and that contextual effects operate entirely within the boundaries that have been operationally defined.

A Closed-System Example

Suppose we are willing to assume that our criterion for aggregation, here a spatial one, operates only as an independent variable. Of course we do not imagine that location, per se, affects the variables of interest. Instead, one's spatial postion may be taken as a cause indicator of the unmeasured variables that are presumed to be the true causes of the variables in question. Take the model of Figure 1 as an illustration. Perhaps X_1 represents educational achievement, assumed to be a constant property of the individual once the process has been completed. Suppose X_2 represents a relatively constant type of personal value (say, egalitarianism) that has been developed over time as a result of socialization experiences linked closely

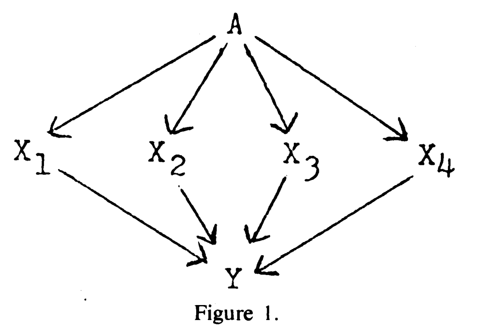

Figure 1.

with one's spatial location. Let X_3 represent another kind of attitudinal variable (say, one's attitude toward a specific minority) that is readily modifiable and therefore subject to changes in one's immediate environment. Finally, suppose X_4 represents a contextual variable (such as a set of sanctions) that operates in the immediate locale.

Now suppose that all these X_i affect a certain form of behavior Y. To simplify we shall assume that the effects are additive, so that the behavior Y may be represented by the equation

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon \quad (1)$$

In the model of Figure 1 we have drawn causal arrows from A to each of the X_i representing the argument that one's location in space in part determines the levels of these X_i as intervening variables.

In any realistic situation an investigator will be unaware of or unable to measure many of the X_i that affect Y. Suppose, for example, that only X_1 and X_4 have been measured and used in an incorrectly specified equation for Y. The least-squares estimates b_1 and b_4 of the parameters β_1 and β_4 will then be biased to the degree that the omitted intervening variables are correlated with X_1 and X_4. In the model of Figure 1 the intercorrelations among the X_i are due solely to A, implying that a control for A (if perfectly measured) would wipe out these interrelationships. Thus if we were to examine the data *within* a single territorial unit, we would find no association among the X_i, implying that even in the incorrectly specified equation

$$Y = a + b_1 X_1 + b_4 X_4 + e \quad (2)$$

the estimates b_1 and b_4 would be unbiased estimates of β_1 and β_4.

Of course this is a highly oversimplified model in which there are no other arrows connecting the X_i, whereas in actuality we would expect intercorrelations within each area. But this prototype model is presumably illustrative of more complex ones and involves the kind of assumption needed to justify controlling for residential area. The essential notion is that many causal factors are generally confounded together because of common residence. Therefore, a control for residence is expected to weaken these associations, if not do away with them altogether.

What is less obvious is that when we aggregate by location we do the very opposite of controlling for A. In grouping by A we put together people who are similar in their X_1 levels. But they will also be similar with respect to their X_2, X_3, and X_4 values. Suppose the X_i are labeled so that the relationships with A are all in the same direction, so that we may represent them by positive signs. Then persons who reside in a location where the X_1 values tend to be high will also have high X_2, X_3, and X_4 values. If we shift our analysis to the macrolevel, using the estimating equation

$$\bar{Y} = a^* + b^*_1 \bar{X}_1 + b^*_2 \bar{X}_2 + b^*_3 \bar{X}_3 + b^*_4 \bar{X}_4 + e^* \quad (3)$$

where the $\bar{X}_i$ represent mean values for the same X_i as represented in Figure 1, we may ask how the new least-squares estimates b^*_i may be expected to compare with estimates that would have been computed on the basis of individual-level data.

What happens in this case is that the $\bar{X}_i$ will be more highly intercorrelated than the microlevel counterparts X_i. If we have specified the model perfectly and if there is absolutely no measurement error in any variable, this will not lead to any systematic biases in the macrolevel estimates of the parameters. But because of the increased intercorrelations we encounter a multicollinearity problem that tends to increase sampling errors.

It is more important, however, to consider the implications of this confounding of intervening variables in instances where there are specification errors. Suppose we do not know all the X_i that cause Y and that are intercorrelated because of location. To be specific, suppose we have included only $\bar{X}_1$ and $\bar{X}_4$ in the equation for $\bar{Y}$. Our biases in parameter estimates will now be much more serious than in the micromodel discussed earlier. In effect, if we shift to group means but ignore certain of the causes of $\bar{Y}$, the effects of these omitted variables are even more con-

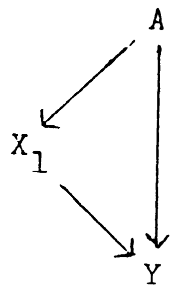

Figure 2.

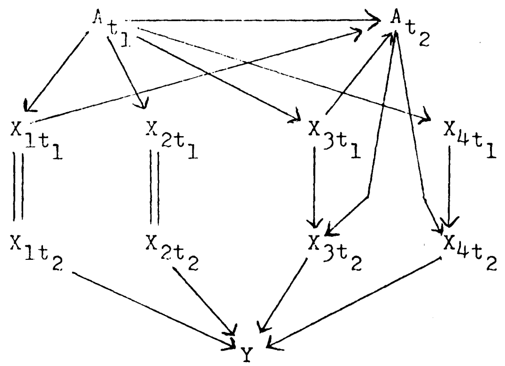

Figure 3.

founded with those of the intervening variables we have been able to include. Put another way, our aggregated model is more sensitive to at least these types of specification errors than is the micromodel, even where the location variable A has been ignored.[2]

For the model of Figure 1 we thus have three analysis possibilities. Our best option is to use microdata and to control for A. Our second best bet is to use microdata and to ignore A. In doing so, if we happen to leave out any of the intervening X_i we will confound their effects with the remaining X's. The third option is to obtain between-area data by aggregating, in which case we increase the intercorrelations among the intervening variables, thereby confounding to an even greater degree the effects of the omitted $\bar{X}_i$ with the causal variables in which we are explicitly interested.

We cannot say that aggregation will always have this effect, but to the degree that reality approximates the model of Figure 1 this will hold. In the extreme case where we have measured only X_1, the original model could be replaced by Figure 2 in which the arrow from A to Y has been drawn as direct. Here A is creating a partly spurious relationship between X_1 and Y, and should be controlled. But if we aggregate by A we are grouping by a cause of Y, and as I have shown elsewhere (Blalock, 1964) this produces a systematic bias in our slope estimate linking X_1 and Y, a bias that may be interpreted as resulting from the confounding with X_1 of all other effects of A that are also causes of Y.[3]

An Open-System Example

Now consider the somewhat more complicated but also more realistic situation in which persons are immigrating into and emigrating from each of the areal units. Here we must take A as a dependent as well as an independent variable. Of course the area is not "dependent" upon its residents. What we mean is that since our microunits of analysis are individuals or families, the particular area in which they are located is dependent upon their decisions. To study this kind of situation we now must bring in the time dimension and try to distinguish between contemporary and past influences, as well as internal states that we are willing to assume are stable over time as contrasted with those that may change as a result of immediate stimuli.

Consider the model of Figure 3. Here we distinguish between an individual's location at time 1, namely A_{t1}, and his or her location at time 2. Migration may or may

[2] Irwin and Lichtman (1976) stress that the essential criterion in deciding between a micro- and a macromodel is the relative degree of specification errors involved. Here, this criterion implies that the micromodel is to be preferred.

[3] Firebaugh (1978) discusses this kind of situation in terms of a general criterion for avoiding aggregation bias, namely that the association between Y and $\bar{X}$, controlling for X, must be zero if bias is to be avoided. In other words, $\bar{X}$ must not belong in the equation for Y, a criterion that will not be satisfied if $\bar{X}$ is a surrogate for other variables that have been omitted from the equation because of specification errors.

not have occurred in the interim. Following Stinchcombe's (1968) discussion of historical explanations we may draw an arrow linking A_{t1} to A_{t2}. What one does today, or where one is, influences tomorrow's behavior or location, if only in the sense that once a given pattern of behavior has been learned there is a vested interest in not changing it unless there are specific pressures to do so. For those who have not migrated, A_{t1} and A_{t2} will be identical. The degree of association between these two variables will depend on the proportion of migrants and, although not indicated in the diagram, this proportion itself could be one of the contextual variables that affect behavior Y, perhaps through the sanction system represented by X_4.

Suppose X_1 and X_2 represent variables that do not change over time. Therefore the change in location has not affected either of these variables. I have represented this by drawing in double lines without arrowheads to indicate that X_1 and X_2 remain identical at the two points in time. Suppose, however, that X_3 and X_4 may be affected by the new location as well as the old. Therefore I have drawn arrows to X_{3t_2} from both X_{3t_1} and A_{t2} (and similarly for X_{4t_2}), making the assumption that the changes produced by the change in location are almost immediate. Finally, the behavior Y at time 2 is taken as dependent upon the contemporary values of the X_i, as was also true in Figure 1.

Now suppose both migrants and nonmigrants are lumped together, as is often the case in microlevel analyses and as is practically always necessary for aggregated data. Again, if we have perfect measures of the contemporary values of all the X_i, we may estimate their separate effects without bias, though if they are too highly intercorrelated we shall have large sampling errors. But suppose there are specification errors, either in the form of poor measurement of some of the X_i or their omission from the equation. Previously we noted in Figure 1 that a control for A would remove all the intercorrelations among the X_i, so that if some were inadvertently omitted the estimates of the structural parameters for the others would remain unbiased. This will not be true, however, for the more complex model of Figure 3 unless both A_{t_1} and A_{t_2} are simultaneously controlled. If we looked only within A_{t_2} we would expect to find a correlation between X_1 and X_2 that would be some function of the proportion of immigrants, since these variables depend only on the factors operating during the earlier period. Presumably, X_1's correlation with X_3 and X_4 would be somewhat weaker, owing to the contemporary factors affecting the latter two variables. The correlation between X_{3t_2} and X_{4t_2}, assuming we are dealing with within-area data, will depend on the relative importance of contemporary influences as compared with earlier ones.

What happens when we aggregate using only the present location A_{t_2}? Once again, we do the very opposite of controlling for location and thereby tend to confound the effects of the four X_i. But we now also are grouping by a variable that may be dependent upon certain of the X_i. In Figure 3 I have drawn arrows from X_{1t_1} (say, education) and X_{3t_1} (say, attitude toward a minority) to A_{t_2}, presuming that these two X_i have influenced the decision to migrate. But if we aggregate by A (at time 2) we are manipulating a dependent variable in terms of the relationship between X_1 and X_3, and this will distort their relationship in an unknown way.

The models with which we have been concerned are grossly oversimplified and merely illustrative of the problems one encounters when aggregation operations are poorly understood. In a sense, aggregation by spatial units is understood in that the criteria for aggregation are clearly operationlized. But what we generally lack is a theoretical model connecting spatial location with the other variables in the system. Thus we achieve operational simplicity at the expense of theoretical clarity. The result is that we are unable to link our macrolevel aggregated data with the microlevel causal processes that may have produced these data. Put another way, if we wanted to insert the aggregation criterion into the causal model we would find that the model would have to be highly complex because one's spatial location is not simply related to the other variables in these models.

CONCLUDING REMARKS

These are but two among many possible illustrations of the need for careful conceptualization and attention to measurement problems and of the fact that theoretical and methodological issues are closely interconnected. They also suggest the importance of bringing implicit assumptions out into the open, even where the added variables in the model may have to remain unmeasured in any given piece of research. Unless this is done, many of these variables will remain confounded with measured variables. It will then be difficult to decide rationally as to the relative merits of alternative design strategies needed to unravel their interrelationships.

In particular, it is important to reemphasize how crucial it is to avoid the temptation to sidestep theoretical and conceptual issues by resorting to very simple operational procedures. I have illustrated this in terms of aggregation according to spatial criteria. A similar temptation also arises with time. Certain variables, such as education, experience, or investments, may be indirectly measured in terms of lapsed time, whereas the conceptual variables of real interest may be only very loosely defined theoretically. The literature on the age-period-cohort problem, for example, relies almost exclusively on calendar dates and lapsed time, as indicators of experience variables, as for example the assumed common experiences of persons born during a five-year period. Relatively sophisticated methodological techniques may be used to attempt to disentangle the separate effects of functionally interrelated variables, as operationally defined. But the true cohort or period effects remain only vaguely specified, as does the linkage between chronological age and maturation. Obviously, sophisticated data analysis techniques, alone, cannot resolve these and other problems unless the theoretical and measurement-error models are clearly specified (Glenn, 1976).

Initial efforts to specify models more completely and to theorize explicitly about linkages between measured and unmeasured variables are almost certain to have discouraging implications. We shall realize how many missing variables and hidden assumptions tend to be ignored in empirical data analyses, as well as theoretical interpretations of empirical results. This obviously carries with it the danger of inhibiting further work and encouraging a hypercritical appraisal of the sociological literature. I believe this is a risk we must take, however, if we are to create a really cumulative knowledge base.

This, in turn, leads me to one inescapable conclusion. Sociologists need to work *together* on these problems. We can ill afford to go off in our own directions, continuing to proliferate fields of specialization, changing our vocabulary whenever we see fit, or merely hoping that somehow or other the products of miscellaneous studies will add up. The plea, then, is for a sustained effort to clarify our theoretical constructs and self-consciously to ask ourselves how different strategies of conceptualization relate to problems of data collection and measurement.

There will still be plenty of room for differences in terms of the kinds of propositions we wish to state and test, the assumptions we are willing to make, the problems we study, the courses of action we recommend, and the theoretical and ideological biases with which we operate. In the proposed joint effort, there is a need for many different kinds of skills, interests, and knowledge bases to help solve technical issues, bring out implicit assumptions, and try to reach a working consensus on our conceptual apparatus and epistemic correlations.

If nothing else, such a concerted effort will better enable us to comprehend what each of us is trying to say and to appreciate more fully the complexity of the theories and analyses needed to understand a very complex reality. If we do not make this concerted effort, I fear that sociology in the year 2000 will be no more advanced than it is today, though perhaps it will contain far more specializations, theoretical schools, methodological cults, and interest groups than, even today, we can readily imagine.

REFERENCES

Blalock, H. M.
1964 Causal Inferences in Nonexperimental Research. Chapel Hill: North Carolina Press.
1979 "Dilemmas and strategies of theory construction." Pp. 119–35 in W. E. Snizek, E. R. Fuhrman, and M. K. Miller (eds.), Contemporary Issues in Theory and Research. Westport, Conn.: Greenwood.

Blalock, H. M. and Paul H. Wilken
1979 Intergroup Processes: A Micro-Macro Approach. New York: Free Press.

Firebaugh, Glenn
1978 "A rule for inferring individual-level relationships from aggregate data." American Sociological Review 43:557–72.
1979 "Assessing group effects: a comparison of two methods." Sociological Methods and Research 4:384–95.

Glenn, Norval D.
1976 "Cohort analysts' futile quest: statistical attempts to separate age, period and cohort effects." American Sociological Review 41:900–4.

Hannan, Michael T.
1971 Aggregation and Disaggregation in Sociology. Lexington: Heath-Lexington.

Hannan, Michael T. and Leigh Burstein
1974 "Estimation from grouped observations." American Sociological Review 39:374–92.

Irwin, Laura and Allan J. Lichtman
1976 "Across the great divide: inferring individual level behavior from aggregate data." Political Methodology 3:411–39.

Langbein, Laura Irwin and Allan J. Lichtman
1978 Ecological Inference. Beverly Hills: Sage.

Merton, Robert K.
1968 Social Theory and Social Structure. New York: Free Press.