Durkheim's Suicide and Problems of Empirical Research

Author(s): Hanan C. Selvin

Source: *American Journal of Sociology*, May, 1958, Vol. 63, No. 6, Emile Durkheim-Georg Simmel, 1858-1958 (May, 1958), pp. 607-619

Published by: The University of Chicago Press

Stable URL: https://www.jstor.org/stable/2772991

DURKHEIM'S *SUICIDE* AND PROBLEMS OF EMPIRICAL RESEARCH[1]

HANAN C. SELVIN

ABSTRACT

The theoretical content of *Suicide* has been thoroughly discussed, but relatively little attention has been paid to the ways in which Durkheim tested and refined his theories with empirical data. Durkheim's empirical analyses are multivariate: each additional variable is systematically incorporated into the relationships that have previously been studied. The analyses are also replicative: the same relationship is studied in widely varying contexts. A careful study of the consequences of these procedures, both intended and unintended, yields valuable insights into current problems of theoretically oriented empirical research.

Sixty-one years after it first appeared in print, Émile Durkheim's *Suicide*[2] is still a model of sociological research. Few, if any, later works can match the clarity and power with which Durkheim marshaled his facts to test and refine his theory. The stature of this work is even more impressive when one remembers that Durkheim lacked even so rudimentary a tool as the correlation coefficient. Yet the methodology of *Suicide* is important to those now engaged in empirical research, not merely to historians of sociology. Durkheim recognized and solved many of the problems that beset present-day research. Others he formulated so lucidly—perhaps because he did not exile his methodology to appendixes—that their solution is relatively simple with the tools now available. To use a fashionable term, this paper will explore some of the methodological "continuities" that stem from *Suicide,* much as the essays based on *The American Soldier*[3] and *The Authoritarian Personality*[4] have done with these more recent classics.

"Methodology" has several meanings to sociologists. To some it means questionnaires, interviews, punched cards—the hand tools of research. To others, such as Durkheim himself and Parsons,[5] it is the assumptions and concepts used in constructing a theory. Here it will be used to mean the systematic examination of the procedures, assumptions, and modes of explanation in the analysis of empirical data.[6] This focus on Durkheim's methodology is not meant to minimize the importance of his theoretical insights; the value of methodological investigations, after all, is that they lead to more effective theorizing about social behavior. But Durkheim's theoretical development has been discussed by many authors,[7] including others in this *Journal,* while his analytical procedures have not received the attention they deserve.

[1] I am deeply indebted to Paul F. Lazarsfeld for acquainting me with the kind of analysis on which this paper is based.

[2] Émile Durkheim, *Suicide,* trans. John A. Spaulding and George Simpson (Glencoe, Ill.: Free Press, 1951).

[3] Robert K. Merton and Paul F. Lazarsfeld (eds.), *Continuities in Social Research: Studies in the Scope and Method of "The American Soldier"* (Glencoe, Ill.: Free Press, 1950).

[4] Richard Christie and Marie Jahoda (eds.), *Continuities in Social Research: Studies in the Scope and Method of "The Authoritarian Personality"* (Glencoe, Ill.: Free Press, 1954).

[5] Émile Durkheim, *The Rules of Sociological Method,* trans. Sarah A. Solvay and John H. Mueller (Glencoe, Ill.: Free Press, 1938); Talcott Parsons, *The Structure of Social Action* (Glencoe, Ill.: Free Press, 1949), pp. 20–27 and chap. ix.

[6] Paul F. Lazarsfeld and Morris Rosenberg (eds.), *The Language of Social Research* (Glencoe, Ill.: Free Press, 1955), p. 4.

[7] Parsons, *op. cit.*; Harry Alpert, *Émile Durkheim and His Sociology* (New York: Columbia University Press, 1939), Parts II and III; Émile Benoit-Smullyan, "The Sociologism of Émile Durkheim and His School," in Henry Elmer Barnes (ed.), *An Introduction to the History of Sociology* (Chicago: University of Chicago Press, 1948), chap. xxvii. His chapter contains a considerable bibliography.

MULTIVARIATE ANALYSIS

Central to Durkheim's methodology is his use of what has been called *multivariate analysis:* "the study and interpretation of complex interrelationships among a multiplicity of characteristics."[8] Much of the empirical analysis in *Suicide* can be viewed as the progressive introduction of additional variables. It will be useful to examine one of these analyses in detail, for it includes several of the procedures to be considered in this paper.

The first chapter on egoistic suicide (Book II, chap. ii) begins with the relation between religion and suicide rates for three groups of countries—the predominatly Protestant, the mixed Protestant and Catholic, and the predominantly Catholic.[9] But, as Durkheim points out, this comparison includes countries with radically different social conditions and requires consideration of the relation between religion and suicide *within* each country. Bavaria, the German state with the lowest proportion of Protestants, has the lowest proportion of suicides. And, in what may seem a mere piling-up of instances, the provinces within Bavaria also exhibit this same relationship: "Suicides are found in direct proportion to the number of Protestants and in inverse proportion to that of Catholics." Prussia and the Prussian provinces are the site of a similar analysis. Then the analysis is repeated for a third country: Switzerland. Here Durkheim takes advantage of the fact that both French- and German-speaking areas contain some cantons that are largely Catholic and others that are largely Protestant. This allows him to hold constant the effect of language as well as nationality ("race") while examining the effect of religion on suicide.

All the preceding analyses are based on data for nations or other large aggregations; thus the discussion of Bavaria cites the relatively high rate of suicides in provinces with high proportion of Protestants and the low rate in provinces with many Catholics. The implications of this procedure will be considered later; here it is important to note only that Durkheim recognized the difference between relationships based on aggregate data and those based on individual data, for he goes on to say that "in a fairly large number of cases the number of suicides per million inhabitants of the population of each confession has been directly determined." And he presents data on the suicide rates by religion for twelve periods of time in five countries, as well as some fragmentary data for France.

After disposing of the "deviant case" of Norway and Sweden, Durkheim considers the low suicide rate among Jews. As compared with Protestants and Catholics, Jews are more likely to live in cities and to pursue intellectual occupations—both conditions that are associated with higher suicide rates. Therefore, Durkheim reasons, if the reported rate of suicide among Jews is lower, despite these conditions, the "true" Jewish rate must be even lower than the figures reveal it to be.

As this passage makes plain, multivariate analysis meant more to Durkheim than simply considering the separate relationships between suicide and the several independent variables—religion, nationality, and language. Each new variable is progressively incorporated into the preceding analyses, so that several variables are considered jointly. The methodology of mutivariate analysis is most clearly seen in the case where a relationship between one independent variable (say, religion) and the dependent variable (suicide) is "elaborated" by the introduction of a third variable or "test factor" (say, nationality). Lazarsfeld, Kendall, and Hyman have defined three major types of elaboration: explanation, interpretation, and specification.[10] Explanation is the attempt

[8] Lazarsfeld and Rosenberg, *op. cit.,* p. 11.

[9] This and the following two paragraphs are taken from *Suicide,* pp. 152–56.

[10] Paul F. Lazarsfeld, "Interpretation of Statistical Relations as a Research Operation," in Lazarsfeld and Rosenberg, *op. cit.,* pp. 115–25; Patricia L. Kendall and Paul F. Lazarsfeld, "Problems of Survey Analysis," in Merton and Lazarsfeld, *op. cit.,* pp. 133–96, esp. pp. 135–67; Herbert H. Hyman, *Survey Design and Analysis* (Glencoe, Ill.: Free Press, 1955), chaps. vi and vii.

to "explain away" the apparent meaning of an observed relationship. For example, the association between religion and suicide might have been a manifestation of nationality, since countries like Germany have both a high suicide rate and many Protestants. Looking into this possibility, Durkheim finds that the original association between religion and suicide persists when national differences are taken into account; nationality is therefore not an explanation of this relationship.

Once convinced that nationality and language do not explain away the association between religion and suicide, Durkheim turns to the interpretation of this relationship: what is the chain of variables connecting two such disparate phenomena as Protestantism and a high suicide rate? A spirit of free inquiry, according to Durkheim, is the most important link in this chain: Protestantism fosters free inquiry and free inquiry in turn leads to a higher rate of suicide.

Although Durkheim lacked the statistical techniques to develop these ideas rigorously, he saw their central place in theoretically oriented research. The relationship between two variables ". . . may not be due to the fact that one phenomenon is the cause of the other but to the fact that they are both the effects of the same cause, or, again, that there exists between them a third phenomenon, interposed but unperceived, which is the effect of the first and the cause of the second."[11]

Specification, the third mode of elaboration, identifies the conditions under which a relationship holds true in greater or less degree. For example, the effect of religion on suicide is less in the German cantons of Switzerland than in the French. Since specification appears in many forms in *Suicide* and since its role in the development of sociological theory differs from the other modes of elaboration, it will be considered at some length.

Specification leads to the development of multivariate theories of behavior in a way that is not true of explanation and interpretation. The aim of specification is to construct three-variable relationships—to say that, as in the example just cited, the effect of religion on suicide is greater in one place than in another. Note that this statement cannot be decomposed into a set of two-variable relationships. Explanation, on the other hand, involves a three-variable association only as an intermediate step, either toward rejecting the apparent finding or toward affirming its provisional meaning; in either case, the result is not a three-variable relationship. Interpretation, likewise, uses the three-variable association only to produce a series of two-variable relationships, to show that these relationships are linked by the variables they have in common.

This greater complexity of specification, its essential three-variable nature, leads to more complex problems in analysis. Durkheim's successes and failures in coping with some of these problems are instructive. Three problems will be considered: (1) the joint effects of group and individual characteristics on individual behavior; (2) the theoretical problems stemming from the statistical concept of "interaction"; and (3) the question of when to stop an empirical analysis.

The variety of analyses that come under the heading of specification is suggested by Hyman's classification; among other ways, one can specify a relationship according to the interest and concern of respondents, the time and place at which it occurs, or the conditions and contingencies on which it depends.[12] Durkheim's analysis provides still another type based on the "units of analysis." This type of specification of which there are several varieties, has been called "contextual analysis":[13] it involves the joint effects of an individual characteristic and a group characteristic on rates of individual

[11] *Rules,* p. 131.

[12] Hyman, *op. cit.,* pp. 295–311.

[13] Hanan C. Selvin and Warren O. Hagstrom, "Contextual Analysis: The Joint Effects of Group and Individual Characteristics on Behavior" (in preparation). I am indebted to Robert K. Merton for first calling my attention to the necessity of specifying the social context of a relationship between individual characteristics.

behavior. In discussing the lower suicide rate among married people, Durkheim points out that in France the difference between the married and the single (his "coefficient of preservation") is greater among the men, while in the Grand-Duchy of Oldenburg it is greater among the women.[14] That is, the social and cultural differences between France and Oldenburg are manifested in two essentially different ways: (1) They exert a *direct* effect; the over-all suicide rate is noticeably higher in France than in Oldenburg. (2) They exert an *indirect* effect; the *relationship* between sex and suicide is different in France and in Oldenburg. In other words, national characteristics have a differential impact on the sex-suicide association in the two countries, the difference between the sexes being greater in France than in Oldenburg.

Methodological devices like contextual analysis are more than ingenious ways to manipulate data. As Merton has emphasized, they are important in opening new directions for theory.[15] Durheim's contextual analysis raises questions about the ways in which group and individual characteristics interact to affect behavior. For example, under what conditions do national characteristics produce such a marked reversal in the association between individual attributes and behavior?

The negative side of this case can also be found in *Suicide:* where Durkheim lacked adequate statistical techniques, he was occasionally led into theoretical contradictions. At one point he asserts that "the relation between the aptitude for suicide of married persons and that of widowers and widows is identically the same in widely different social groups, from the simple fact that the moral condition of widowhood everywhere bears the same relation to the moral constitution characteristic of marriage."[16] But Durkheim's data on Oldenberg and France lead to the opposite conclusion—that the relation between the suicide rates of married persons and widows and widowers was *not* the same in the two countries. What Durkheim lacked and what has since become available is a precise conception of statistical interaction, the ways in which the association between two variables depends on the values of a third variable.[17]

Durkheim's treatment of statistical interaction and of the theoretical relationships that it measures is notably inconsistent. Sometimes, as here, he ignores the presence of interaction in his data. Elsewhere, he correctly notes its presence, remarking, for example, that seasonal differences in suicide are less pronounced in cities than in rural areas.[18] And in another place he assumes, without any evidence for or against his assumption, that the interaction of temperature and location is zero: ". . . if the temperature had the supposed influence, it should be felt equally in the geographical distribution of suicides."[19]

One possible reason for Durkheim's inconsistency is that he had not formalized his analytical procedures. In effect, each time he came to a case of specification, it had to be reasoned through from the beginning. Formalizations such as the Lazarsfeld-Kendall-Hyman types of elaboration enable the analyst to recognize the same principle at work in different instances and therefore to treat them similarly. Such formalizations also reveal a link between apparently disparate phenomena; thus the formal model of elaboration also illuminates the seemingly unrelated problem of when to stop an analysis.

Durkheim's treatment of "race" and suicide provides a case in point.[20] Arguing that the high rate of sucide in Germany "might be due to the special nature of German

[14] *Suicide,* pp. 177–80.

[15] Robert K. Merton, *Social Theory and Social Structure* (Glencoe, Ill.: Free Press, 1949), chap. iii.

[16] *Suicide,* p. 307.

[17] The phenomenon of statistical interaction has been given many different names (e.g., specification, conditional relationship, differential impact, differential sensitivity, and non-additivity of effects).

[18] *Suicide,* p. 120.

[19] *Ibid.,* p. 113.

[20] *Ibid.,* pp. 86–87.

civilization," he decides to "see whether the German retains this sad primacy outside of Germany."[21] To this end he examines the suicide rates in the provinces of Austria-Hungary, in which German-speaking people range from 1.9 to 100 per cent, and finds "not the least trace of German influence" on the suicide rate. However, a close examination of Durkheim's data, particularly of the five provinces that have high proportions of Germans and disproportionately few suicides, leads to quite different conclusions. These provinces—Upper Austria, Salzburg, Transalpine Tyrol, Carinthia, and Styria—comprise the western part of present-day Austria. If these five contiguous provinces are removed, the Spearman rank correlation for the remaining ten provinces is 0.95, indicating an almost perfect relationship between the suicide rate and the proportion of German-speaking people.

The important point here is not substantive but methodological. Durkheim stopped his analysis as soon as he found a "zero" relationship. This procedure is perhaps more common in research today. Small associations are considered a signal to turn to other matters, especially when the associations are not statistically significant. The reasoning behind this assumption is never made explicit, but it would seem to be that, if two variables are not associated when other items are left free to vary, they will not be associated when these other items are "held constant." That is, if the total association between two variables is zero, the partial associations will be zero. Sometimes this is true; often it is not. Hyman's passage on the "elaboration of a zero relationship"[22] indicates that this may happen when the two partial relations are approximately equal in size and opposite in sign. For example, a surprisingly small association between job satisfaction and participation in community organizations resulted from a positive association between participation and satisfaction among members of the working class and a negative association of approximately the same size in the white-collar class.

To my knowledge, Hyman's is the only published discussion of this problem. It may be useful, therefore, to make two further points suggested by Hyman's brief treatment. First, he implies that this kind of relationship is uncommon and even accidental. Actually, it may occur frequently under certain conditions—for example, in the kind of contextual analysis discussed above, where people are assigned to groups instead of being born into them or choosing them themselves. A study of leisure-time behavior in army training companies found that many small or zero associations between behavior and an individual characteristic, such as marital status, resulted from opposite and approximately equal associations in companies with different "leadership climates."[23]

Second, still another type of elaboration of a zero relationship may be ranged alongside the two identified by Hyman. A zero association between two variables may occur even when both partial associations are in the same direction. The hypothetical example in Table 1 shows that this case would interest the student of political behavior. At both levels of education, the people with more information tend to choose the Democratic party, yet the "collapsed" table of information and party affiliation without regard to education will show that, among both the more-informed and the less-informed, 50 per cent are Democrats. The two partial associations are positive (and about

[21] By today's standards Durkheim's table shows a moderately high degree of association. The Spearman rank correlation is 0.57. In general, Durkheim regarded anything much less than perfect rank correlation as "independent." The reason why he could demand and find such high levels of association, while survey researchers are content with much less impressive relationships, has to do with the differences in the numbers of cases on which the associations are based. See the discussion of "grouping" in G. Udny Yule and M. G. Kendall, *An Introduction to the Theory of Statistics* (14th ed.; New York: Hafner, 1950), pp. 313–14.

[22] *Op. cit.*, pp. 307–10.

[23] Hanan C. Selvin, *The Effects of Leadership Climate on the Nonduty Behavior of Army Trainees* ("University Microfilms Publications," No. 19.256 [microfilmed Ph.D. dissertation, Columbia University, 1956]), Appendix F.

the same size); the total association is zero. Unrealistic as this example may be (although it could describe a university town with a Democratic newspaper), it does demonstrate the importance of looking into those zero associations that theory or previous research suggests should not have been zero. A zero association between two variables may therefore result from any one of three different conditions in the partial relationships: zero associations in both partials, equal and opposite associations, or associations in the same direction. Only the first of these is a signal to stop the analysis.

THE FUNCTIONS OF REPLICATION IN EMPIRICAL RESEARCH

Another multivariate procedure that is conspicuous in *Suicide* and that deserves intensive inquiry is *replication,* the systematic restudy of a given relationship in different contexts.

Again using the first chapter on egoistic suicide as an illustration, we find that the original table relating religion and suicide is followed by no less than the *seventeen* replications in three pages. Why so many replications? Durkheim's answer is not altogether clear. He does demand that the facts cited to support a thesis be "numerous enough not to be attributable to accidental circumstances—not to permit another explanation—to be contradicted by no other fact."[24] But these principles are not systematically explained. This section will therefore examine some of the uses of replications in *Suicide* and will consider what functions replications may serve in current research.

[24] *Suicide,* p. 95.

Durkheim's lavish use of replications is all the more noteworthy in contrast with modern survey research, where a relationship is often demonstrated only in a single table. One reason why Durkheim used so many replications is undoubtedly that his data came from official records; it cost him little more to study suicide in six countries than in one. On the other hand, in contemporary surveys the researcher must gather his own data, often at great expense, so that one community is a practical limit. However, as will be seen, this explanation rests on too narrow a conception of replications; a careful study of Durkheim's procedures will show that there are abundant opportunities for replication in every survey. A second reason for demonstrating a hypothesis in only one table is the belief that a close adherence to modern techniques of statistical inference—tests of significance, confidence intervals, and the like—guarantees the statistical soundness of the conclusions. This point of view has recently been vigorously attacked and staunchly defended.[25] Instead of repeating these arguments here, it will be shown that properly conducted replications may achieve the same ends more effectively. First, however, it is necessary to clarify the different kinds of replications.

TABLE 1

INFORMATION AND PARTY AFFILIATION
EDUCATION HELD CONSTANT
(Hypothetical Data)

	Less Educated		More Educated	
	Much Information	Little Information	Much Information	Little Information
Per cent Democratic.....	38	24	74	63
N.....	(200)	(100)	(100)	(200)

"Unit replications" are the most frequent in *Suicide;* in these the original finding is

[25] Hanan C. Selvin, "A Critique of Tests of Significance in Survey Research," *American Sociological Review,* XXII (1957), 519–27; David Gold, "Comment" and Hanan C. Selvin, "Reply," *American Sociological Review,* XXIII (1958), 85–86; James Beshers, "Comment" and Hanan C. Selvin, "Reply," *American Sociological Review,* Vol. XXIII (April, 1958).

re-examined for different groups of subjects —for example, the excess of military over civilian suicides is confirmed for eight different countries of Europe.[26] Durkheim further replicates this finding within the Austro-Hungarian Empire; the military-civilian difference persists in the various military areas.[27] These two examples suggest that unit replications may be divided into two subtypes: "external replications" and "internal replications."

In external replications the conclusions of one study are tested independently in another study, usually conducted by a different investigator.[28] Cross-cultural comparisons are a familiar example. Demographers and others who work with official records also conduct external replications as a matter of course. But this procedure is difficult and expensive for anyone who must gather his own data.

Just the opposite is true of internal replications, in which a finding is restudied for smaller groups within the original set of subjects. In the passage on military and civilian suicides there are several external replications, based on the different countries, and several internal replications for areas within one country.[29] Although these replications have a geographical basis, a finding may be replicated in groups formed on any set of variables. In the chapter on anomic suicide (Book II, chapter v), Durkheim presents a table of suicide rates by occupation for eight countries. Here the various countries act as external replications for his original finding on occupations and suicide. But the same table could also serve as an internal replication of the national differences considered in the earlier chapter on egoistic suicide, the comparisons between over-all national suicide rates being replicated for the comparable occupational groups within the countries.[30] Of course, all replications are not equally desirable; some criteria for choosing replications will be taken up after considering the uses of replications.

Replications, properly conducted, may serve at least two functions. They may provide a more valid "test of significance" where the usual tests cannot be legitimately applied, and they may lessen the seriousness of the "ecological fallacy," in which relationships between characteristics of individuals are wrongly inferred from data about groups.

REPLICATIONS AND STATISTICAL SIGNIFICANCE

Some reasons why tests of statistical significance are inapplicable in survey research have been noted in another paper.[31] Perhaps the most important reason—and the one that bears on the use of replications—is that the tests assume an experimental situation, in which differences between subjects can be randomized. Agronomists, for example, can convert constant differences in the fertility of their experimental plots into chance differences by randomizing the assignment of their plants; in effect, they toss a coin to determine whether a plant goes into the experimental plot or the control plot. But sociologists can seldom randomize. If Catholics are more numerous in Italy and Protestants in Germany, the sociologist must cope with these stubborn facts. He cannot randomly assign Protestants and Catholics to the two countries.

Randomization turns such systematic differences, or "correlated biases," into random differences. Tests of significance can then be used to measure the probability that the observed difference between experimental and control groups *could* have been produced by the accidents of randomization. Without randomization, these correlated

[26] *Suicide,* p. 228.

[27] *Ibid.,* p. 235.

[28] For a collection of external replications and an interesting treatment of the factors making for successful replications see Robert C. Hanson, "Evidence and Procedures Characteristics of 'Reliable' Propositions in Social Science," *American Journal of Sociology,* LXIII (1958), 357–70.

[29] *Suicide,* pp. 228–39.

[30] *Ibid.,* p. 258.

[31] Selvin, "A Critique of Significance Tests."

biases remain as possible sources of the observed difference between the groups, along with the experimental variable whose influence is being measured. Thus the interpretation of the difference in suicide rates between Italy and Germany as the result of religious differences must take into account the many other differences between the two countries—education, cultural norms, and so on, which may account, in part at least, for the relationship between religion and suicide. Some of these differences could have been removed by cross-tabulation, had Durkheim had the necessary data. Others, not necessarily the least important, could not. Consequently, it appears misleading to ask for the probability that the observed relationship may be "due to change" without first being reasonably certain that correlated biases did not produce it.

Replications, properly conducted, can take care of correlated biases and random errors at the same time. When Durkheim replicates within Germany the relationship previously studied among the various countries of Europe, he is removing the host of biases correlated with nationality. And when he further replicates within Bavaria and within Prussia, constant differences between these states no longer are mixed in with the religion-suicide relationship. Finally, his replications based on individual data remove some of the provincial differences. At each level of replication, additional correlated biases are controlled: the original relationship between religion and suicide gains more and more support.

Replications can thus lessen the likelihood that correlated biases may have produced the observed result. It is necessary only that the variable on which the replications are based be theoretically relevant (i.e., that there is reason to believe it is causally related to both the independent and the dependent variable). Now what of the random factors that may have produced the observed result? Here, too, replications are useful: they lead toward a valid test of significance which is simple to understand and easy to compute.

Consider once again the seventeen replications of the relationship between religion and suicide. How likely is it that this relationship resulted from some combination of random errors? That is, how often would one expect to find seventeen consecutive replications all showing Protestants more likely to commit suicide if, in fact, Protestants and Catholics are equally prone to suicide? Assuming that the replications are independent, that the outcome of one replication does not affect the outcome of another, the probability of such a series of replications is $(\frac{1}{2})^{17}$, a result significant at the 0.00001 level.[32]

Some cautions are in order. First, the validity of this computation, as of the ordinary test of significance, rests on the elimination of correlated biases. Since one can almost always conceive of relevant variables that have not been controlled and since there is no randomization to turn these correlated biases into random events, this condition is never fully satisfied. However, by applying this test only within series of replications, some, at least, of the correlated biases are controlled. The more variables that are controlled in any one replication, the more convincing this test of significance becomes.

Because he is dealing with official statistics that seldom are classifiable according to more than three variables, Durkheim's control over the correlated biases is relatively loose. Survey analysts, on the other hand, can more easily control these variables. For example, to test the hypothesis that socioeconomic status affects voting behavior, one would compare subgroups of high and low status that are similar on as many relevant variables as can be manipulated simultaneously. Such a test might compare two groups of young, single, urban,

[32] Samuel A. Stouffer, "Quantitative Methods," in Joseph B. Gittler (ed.), *Review of Sociology* (New York: John Wiley & Sons, 1957), pp. 45–46. Stouffer's use of replications in *The American Soldier* (Princeton: Princeton University Press, 1949), I, 92–95, is a notable example (he refers to replications there as the "method of matched comparisons"). See also Selvin, *The Effects of Leadership Climate,* pp. 43–47, 183–204.

Protestant, male high-school graduates, one group of high socioeconomic status, the other of low status. This comparison would be repeated for all possible combinations of these six variables; if enough of these comparisons are in the same direction as the original relationship, one can be reasonably confident that the original relationship is not the result of random errors, assuming, of course, that there are no other important correlated biases.

In controlling for correlated biases, the subgroups should be formed on relevant variables; in testing for statistical significance, the subgroups should be independent. Consequently, if both criteria are satisfied, replications will tend to accomplish both tasks simultaneously. Relevance depends on the implicit or explicit theory that is being used, and no rules can guarantee it. On the other hand, independence can in general be insured by seeing that no individual appears in more than one subgroup in a given replication. For example, it would not be efficient to replicate Durkheim's religion-suicide relationship by studying it first among married and single men and then among older and younger men in the same country, for some man will appear in both sets of comparisons.[33]

One empirical procedure used by Durkheim appears at first glance to be another form of replication. In demonstrating that the relationship previously found between time of day and the suicide rate really depends on the "intensity of social life," he musters a variety of indicators of social activity—accidents, rail travel, and express receipts.[34] Social activity and suicide turn out to be highly associated, thus supporting his interpretation of the time-of-day relationship. Although this seems to be another type of replication, serving the same ends as those discussed above, this kind of item replication lacks one essential element of the unit replications: independence of observations. In fact, one does not even seek independence in item replications. These replications help to demonstrate that an indicator is *valid*—in other words, that it means what the analyst says it means. The greater the variety of indicators of social activity that Durkheim can relate to suicide, the greater his assurance that social activity—and not some accidental correlate of it—is what accounts for the variations in suicide.[35]

Replications thus serve Durkheim as a means of testing and refining his empirical hypotheses. The preceding discussion has suggested that they may serve these functions even better in survey research, where large numbers of variables can be easily manipulated. Another function of replications, of less general application but often of crucial importance in studies based on "ecological" data, will be examined in the following section.

REPLICATION AND THE "ECOLOGICAL FALLACY"

Robinson[36] first called the attention of sociologists to the fallacy of assuming that associations computed from group means or group proportions are valid estimates of the associations that would be obtained from individual data.[37] Most of Durkheim's ta-

[33] Note that two or more time replications in the same table of *Suicide* are always separated by a period of years, thus increasing the independence of the replications.

[34] *Suicide,* pp. 118–20.

[35] Item replication is, in a sense, the inverse of scaling. In item replication, one begins with a concept and seeks a variety of indicators to clarify its meaning. In scaling, one begins with a set of items and asks whether there is a single underlying concept that accounts for them.

[36] William S. Robinson, "Ecological Correlations and the Behavior of Individuals," *American Sociological Review,* XV (1950), 351–57. This problem seems to have been discovered and treated independently in psychology, statistics, and economics, as well as in sociology, although not so fully as in Robinson's paper (see Edward L. Thorndike, "On the Fallacy of Imputing the Correlations Found for Groups to the Individuals or Smaller Groups Composing them," *American Journal of Psychology,* LII [1929], 122–24; Yule and Kendall, *op. cit.,* pp. 310–15; Kenneth H. Arrow, "Mathematical Models in the Social Sciences," in Daniel Lerner and Harold D. Lasswell [eds.], *The Policy Sciences* [Stanford: Stanford University Press, 1951], p. 134).

[37] Robinson's paper concentrates on linear regres-

bles are based on such fallacious reasoning. Thus he reports that the rate of suicide in departments of France varies according to the proportion of "persons of independent means."[38] This result is consistent with either of the following hypotheses: *none* of the people who commit suicide has independent means, or *all* of them have independent means. In this table the ecological association between characteristics of departments reveals nothing about the individual association between a person's wealth and whether or not he commits suicide.

Every case of ecological associations does not entail the ecological fallacy. Menzel has shown that ecological associations are not only permissible but necessary when the "unit of analysis" is a group rather than the individuals in it.[39] However, Durkheim never theorizes about wealthy and poor departments, only about wealthy and poor individuals. And if he were interested in group characteristics—at the level, say, of departments or provinces—why would he replicate for successively finer subdivisions within these groups, in two cases carrying the replications down to individual data?[40] It is clear that Durkheim was guilty of the ecological fallacy.

To say that Durkheim's procedures were fallacious does not detract from his conclusions.[41] They may be true, even if they do not necessarily follow from his data. In effect, Durkheim recognized this problem and attempted to solve it in the only way open to him—the systematic use of replications for units smaller than those in his original relationship. When he was able to carry the replications down to individual data, there was, of course, no ecological fallacy: the hypothesis stated for group data is confirmed for individual data. However, even when Durkheim did not have individual data, the procedure of replicating in smaller units may lead, as Duncan and Davis have shown,[42] to useful estimates of the individual association.

Durkheim's data cannot be used to illustrate this procedure, for a reason to be stated shortly. Consider, instead, the following ecological data from a hypothetical city in which there are 20 election districts of 100 voters. Each of the first 10 districts has 10 per cent Republicans and 20 per cent voting for Eisenhower. (These are group attributes; nothing is said about the number of individuals who are both Republicans and for Eisenhower.) The other 10 districts are 80 per cent Republican and 90 per cent for Eisenhower. However one chooses to measure association—by product-moment correlation, percentage differences in a fourfold table, or any other index—this is a perfect ecological association: all the districts with many Republicans are for Eisenhower, as against none of the districts with few Republicans. This is formally the kind of relationship studied by Durkheim.

Now consider any one of the first 10 districts. The number of Republicans for Eisenhower is not known, but it is easy to see that it must be between a maximum of 10 and a minimum of zero. The corresponding values for the other 10 districts are 80 and 70. If each district had the maximum possible number of Republicans for Eisenhower, the total for the city would be $(10 \times 10) + (10 \times 80) = 900$; the minimum total for the city would be $(10 \times 0) + (10 \times 70) = 700$. Table 2 shows the maximum and minimum values for the four cells of the individual association. If the association between Republicanism and preference for Eisenhower is a maximum in each district, then, for the city as a whole, 100 per cent of the Republicans and 18.2 per cent of the

---

sion and correlation, but the same problem arises, whatever the measure of association.

[38] *Suicide,* p. 245.

[39] Herbert Menzel, "Comment on Robinson's 'Ecological Correlations and the Behavior of Individuals,'" *American Sociological Review,* XV (1950), 674; see also the discussion of the "modifiable unit" in Yule and Kendall, *op. cit.*

[40] *Suicide,* pp. 154, 175.

[41] Andrew F. Henry and James F. Short, Jr., have made the most recent attempt to formulate a comprehensive theory of suicide in their *Suicide and Homicide* (Glencoe, Ill.: Free Press, 1954).

[42] Otis Dudley Duncan and Beverly Davis, "An Alternative to Ecological Correlation," *American Sociological Review,* XVIII (1953), 665–66.

Democrats prefer Eisenhower, a difference of 81.8 per cent. This is something less than the perfect ecological association; there, every Republican district was for Eisenhower and every Democratic district for Stevenson. Similarly, if the within-district associations are as small as possible, then 77.8 per cent of the Republicans and 36.4 per cent of the Democrats are for Eisenhower, a difference of 41.4 per cent.

The ecological data for the districts thus lead to bounds for the individual association in the city as a whole. Durkheim's procedure of replicating in successively smaller units would likewise seem to generate bounds for the individual associations on which his analysis is really based. In fact, Goodman has shown that the smallest units give the closest possible bounds; nothing would be gained in the example above by aggregating the districts into larger areas.[43] Durkheim's reason for replicating at several levels instead of going directly to the smallest units is probably that his data were uneven (i.e., that he did not have comparable figures for the provinces in all his countries).

The only difficulty in applying this procedure to Durkheim's data is that it will not work. Goodman's analysis indicates that bounds can be inferred only when the points in the ecological scatter plot are located in at least two of the four regions in Figure 1. The two sets of points for the hypothetical election data are located in regions *A* and *D*, as shown. But if the proportion for Eisenhower is replaced by the suicide rate and the proportion Republican by any one of Durkheim's independent variables (say, the proportion of Protestants), then all Durkheim's data are located in region *B*. (This is the region in which the suicide rate is less than the proportion of Protestants and also less than the proportion of non-Protestants.) No amount of replication, therefore, will lead to bounds for the individual association between suicide and Protestantism or between suicide and any of Durkheim's independent variables. Suicide is too rare an event, compared with the rates of the other variables he used. Ecological reasoning in *Suicide* is not limited to geographical data.

TABLE 2

PARTY AFFILIATION AND VOTE
(Hypothetical Data)

Individual Candidate Choice		Individual Party Affiliation: Republican	Democratic	Total
Eisenhower.....	Max.	900	200	1,100
	Min.	700	400	
Stevenson.....	Max.	0	900	
	Min.	200	700	900

"A proof of the slight effect of marriage [on suicide] is the fact that the marriage rate has changed very little since the first of the [nineteenth] century, while suicide has tripled."[44] Here, again, the real question is whether married people are more or less likely than single people to commit suicide. With these data, however, a positive, negative, or zero individual association is possible. Data aggregated over time can lead to false interpretations, just as can data aggregated over geographical units. In both cases the difficulty is removed by studying the data within units rather than "ecologically." There is no ecological fallacy when the independent variable and suicide are examined for individuals within each geographical area or for individuals within each time period. In other words, both versions of the ecological fallacy can be considered as cases of "spurious association." This, it will be recalled, is an erroneous inference that is "explained away" by holding con-

[43] Leo A. Goodman, "Some Alternatives to Ecological Correlation" (Chicago: Statistical Research Center, University of Chicago, 1957) (dittoed). This comprehensive paper also includes another method for estimating the "individual regression" from ecological data under certain assumptions. Unfortunately, these assumptions are not met in Durkheim's data.

[44] *Suicide*, p. 185.

stant a variable that accounts for both the items being studied. The associations between region and religion and between region and suicide may account for the ecological association between religion and suicide; holding region constant removes any possibility of a false inference from this association. Similarly, the association between suicide and the marriage rate may lead to false inferences unless it is examined within relatively short time periods.[45] There is thus a close methodological connection between problems of ecological association and procedures of multivariate analysis.

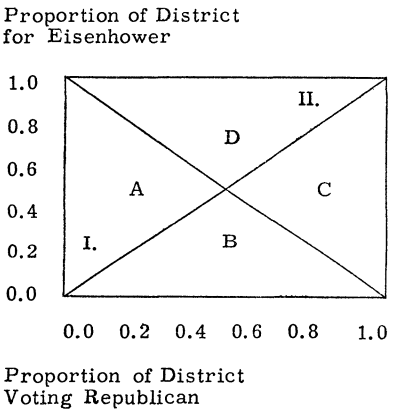

FIG. 1.—Conditions under which ecological associations yield bounds for individual associations.

This section has considered the uses of replications in testing sociological hypotheses and in making valid inferences from characteristics of groups to characteristics of their members. In both cases attention has been focused on the relationship of groups and subgroups; the description of the group, as given by Durkheim, has been taken for granted. But here, too, there are useful lessons to be learned from a careful examination of Durkheim's procedures.

A PROBLEM IN DESCRIBING GROUPS

The first section of the chapter on anomic suicide begins: "It is a well-known fact that economic crises have an aggravating effect on the suicidal tendency."[46] Durkheim then establishes that poverty is not the link between economic crises and suicide. He argues that poverty "tends rather *to produce the opposite effect*. There is very little suicide in Ireland, where the peasantry leads so wretched a life. Poverty-stricken Calabria has almost no suicide; Spain has a tenth as many as France."[47] Here is an association between groups (countries) and individual behavior, in which the relationship is attributed to *one* property of the groups—their poverty. But Calabria, Ireland, and Spain are not only poorer than France; they are also more Catholic than France and, as Durkheim was at pains to show in his analysis of egoistic suicide, less educated. There is no necessary reason why poverty should be singled out as the cause of the lower suicide rate in these countries. Religious or educational differences would have accounted equally well for the variations in suicide.

Essentially this same problem appears in different guise when one tries to infer the intensity of some form of behavior from data about its incidence. For example, Durkheim assumes that districts with high divorce rates have many families that are near divorce.[48] This statement may be true; it is not necessarily true. Two countries with the same average intensity of marital unhappiness may, because of their legal codes, have very different rates of divorce. To move legitimately from incidence to intensity in this way, one must take into account the other group characteristics that affect the dependent variable.

The source of these difficulties is clear. It is the oversimple description of a group according to a single variable. What can be done to avoid such problems? The obvious answer is to hold the other group characteristics constant by cross-tabulation, just as one does with individual characteristics. To study the influence of poverty on suicide, Durkheim would have had to find areas that were alike on other variables, such as

[45] This is explanation in which the "ordering principle" is not "time" but "level of aggregation" or "level of complexity" (cf. Kendall and Lazarsfeld, *op. cit.*, p. 196).

[46] *Suicide*, p. 241.

[47] *Ibid.*, p. 245; italics added.

[48] *Ibid.*, p. 263.

religion, urbanization, and education, and different only in relative wealth. As a practical matter, this is impossible. Murdock's cross-cultural comparisons are based on as many as 250 different cultures, but even this number is too small to allow more than one variable to be held constant in any table.[49] The smaller the groups, the easier such statistical manipulation of group variables becomes, at least in principle. But surveys that study the joint effects of group and individual characteristics on behavior seldom have the resources to include the hundreds of groups that would be necessary if group characteristics were to be manipulated by cross-tabulation. (Studies based on census data can easily handle thousands of groups, but they do not often have data on the joint effects of individual and group variables.)

If groups differing in only a single characteristic are practically impossible to find in survey research and if large numbers of groups are impossibly expensive, is there any alternative? The most attractive alternative—perhaps the only one—is to abandon the attempt to deal with one group characteristic at a time and to describe the groups with as many variables as necessary. Where the number of groups is too small for cross-tabulation, they should be described in terms of a multivariate typology.[50] In the first example above, for instance, Calabria, Ireland, and Spain would have been described not simply as poor but as poor, Catholic, and having a low level of education. Theoretical simplicity is thus sacrificed for theoretical and empirical accuracy. These countries would differ from France on three or more independent variables instead of one, but they would be described in all their relevant aspects.[51]

It is only fitting to close this argument for the typological analysis of group characteristics by noting that here, too, Durkheim pointed the way. In his chapter "How To Determine Social Causes and Social Types" (Book II, chap. i), he argues for the classification of suicides, not as the simple act of killing one's self, but according to "morphological types" determined by the psychological and other characteristics of each suicide. It is only because the necessary data were not available that he was compelled to use his "etiological types." The same logic and the greater ease with which group data may now be gathered suggest that groups be described as multivariate types rather than by a single variable.

Despite its occasionally archaic language ("suicidogenic current"), the empirical analysis in *Suicide* is as vital today as it was in 1897—perhaps more so, since the quantitative approach that Durkheim pioneered has since become widely accepted among sociologists. But overemphasis on the quantitative aspects of *Suicide* would be as dangerous as total neglect, if it furthered the current tendency to substitute technical virtuosity for hard thinking about empirical data, thinking that is guided by theory and is directed toward enriching theory. This, after all, is the essential message of *Suicide:* that methodology is valuable insofar as it springs from the needs of theory and that theory is most fruitful when it is continually tested and refined in empirical research.

UNIVERSITY OF CALIFORNIA (BERKELEY)

[49] George Peter Murdock, *Social Structure* (New York: Macmillan Co., 1949).

[50] Allen H. Barton, "The Concept of Property-Space in Social Research," in Lazarsfeld and Rosenberg, *op. cit.*, pp. 40–53.

[51] When the number of variables and the number of groups is as large as fifteen or twenty, it becomes economical to use factor-analytic procedures for determining the dimensions of the typology (see Andrew W. Halpin, *The Leadership Behavior of School Superintendents* [Columbus: College of Education, Ohio State University, 1956], pp. 3–10; Selvin, *The Effects of Leadership Climate,* pp. 11–39). The descriptions of group characteristics in these two works indicate that factor analysis is ideally suited to the study of groups (while it is, at best, a crude tool for the study of individuals). Briefly, the reason is that the correlations of group means and group proportions are virtually free from the idiosyncratic and random errors that attenuate the correlations of individual data, so that relatively few factors account for most of the variation in a large number of variables.