METHODS OF MEASURING DIFFERENCES IN SOCIAL CLASSES

By: Theodore R. Anderson, Yale University

1. Statement of the Problem

A social class, as the term is used in modern sociology, is a complex phenomenon. It consists of many social categories (such as occupations) which bear a particular relation to a fixed scale (such as a prestige, an income, or a composite scale). This paper will not discuss social classes directly, but rather will focus on the problem of measuring differences between social categories out of which class groupings may arise. Socially, how different are two occupations, or two ethnic groups, or two religions? This paper will present one means of answering these questions.

Suppose for the moment that a matrix describing the social difference between each pair of, say, occupations, were available. It is highly unlikely that such a matrix would have unit rank; that the differences could be resolved into a space of one dimension or into a single directional scale. Two or more dimensions would almost certainly exist. It is convenient to call the space characterized by these hypothetical dimensions a social space. In keeping with this terminology, the difference between any two categories may be described as the social distance between the categories. Using these terms the problem discussed here is that of measuring the social distance between any two categories which exist within a social space.

A means of measuring the social distance between categories should prove useful to a variety of research problems. A matrix of such social distances represents one means of describing a social structure, such as a community, in quantitative terms. Such a description, in turn, would permit studies of temporal change in structural characteristics. At present it is difficult to do more than to speculate about answers to many structural change questions of considerable interest. For instance, is anything like a polarization of occupational roles into opposing camps occurring within the United States, or is the direction of change toward greater homogeneity? An effective measure of social distance could be used to provide an answer.

Such a measure should also prove useful in developing what might be called a social ecology. For example, social mobility might be subjected to the same analysis, via mathematical models, that is currently occurring in connection with physical migration. At the present time, breaking a social mobility matrix into its theoretical components is virtually impossible, largely due to the absense of a sound multidimensional measure of social distance.

It is clear that a measure of social distance will not, in itself, solve the above problems. Rather, it will merely make their solution possible. It is also clear that a complex social structure will contain many parameters other than those characterizable as social distances. Never the less, a sound measure of social distance should have considerable utility. This paper will present such a measure after briefly considering other possible ways of solving the problem. The proposed solution will be illustrated using relatively simple empirical data. Finally, an example will be presented showing how the derived distances may be used to solve a research problem.

2. General Considerations

The term social distance, as it is used here, is equivalent to the phrase, behavioral difference. That is, if the members of two groups perform precisely the same behaviors in all contexts, or behavioral domains, they occupy the same position within social space. The more similar are the behaviors, the closer in space are the groups. The problem of measuring social distance, or behavioral difference, reduces initially to the problem of selecting a behavioral domain within which to observe similarity and difference. By a behavioral domain is meant some population of behaviors such that all the behaviors have something in common. That which is common, for instance, might be a relation to a specific institutional order. Thus, all behaviors bearing upon property might form a domain, or all behaviors bearing upon the mass media of communication.

In general, there are two basic types of behavioral domains. First, there are those sets of behaviors through which individuals in one category or group are related to individuals in other categories or groups. Such behaviors include interactions between groups, interchanges of goods, services, and people, and evaluations of group members by members of other groups. The second type includes all other behavioral domains. To be most useful, measures of social distance should be generated from the second type of domain and not from the first type.

The reason for this statement is simple. Virtually all of the important hypotheses which might be tested in part through the use of social distance measurements involve the relational domains (interactions, interchanges, and evaluations). Do members of groups which are closer in social space interact more frequently? Do members of groups which are far apart feel more hostile toward each other? Is social mobility related to social distance? To be able to answer questions such as these it is crucial that the measure of social distance be generated out of behaviors which are conceptually independent of the ones in question.

Of course, each of these questions implies an hypothesis which could form the basis of a measure of social distance. Thus, a matrix which states the frequency of interaction between social categories could be transformed into one stating the social distance between each pair of categories using the assumption that the frequency of interaction is inversely related to social distance. Virtually all existing measures of distance in social space are based upon some such transformation. Such measures are presumably reasonably accurate. However, to the extent that they use intergroup relations to generate the social distances, they cannot be used to test relational hypotheses. It is the opinion of this author that social distance should be measured within other domains because most of the hypotheses worth testing lie in the relational domains.

No attempt will be made here to survey the literature of existing measures. Such measures are either drawn from relational domains (*e. g.*, the Bogardus measure of social distance (1)) or are measures along only one direction in social space (*e. g.*, the vast number of measures of socio-economic status). The measure proposed here avoids these limitations, in that it is based on behaviors not in the intergroup domains and is a general, non-directional, measure of distance.

3. The proposed Solution

Social distance is essentially equivalent to behavioral difference. Groups are socially distant to the extent that their members behave differently. There is, however, no general measure of the amount of difference between unlike activities. Therefore, the key indicator of social distance between two groups is the difference in the proportion of persons performing the same activity in each group. In particular, consider a set of mutually exclusive and exhaustive categories within a population of persons. Occupations, religions, and ethnic groups are examples. Consider two such categories, say the *ith* and *kth*. Consider also a set of behaviors, which need not be mutually exclusive (the same person may perform one or more of the behaviors). Let the *jth* behavior be one of this set. Finally, let the proportion of persons in the *ith* and *kth* categories who perform the *jth* act be $P(ij)$ and $P(ik)$. The social distance between i and k is assumed to be closely related to the difference between these two proportions.

The difference between these proportions, however, is an index of social distance only to the extent that it is shared by or common to other behaviors (in the factorial sense). Social space may be characterized according to this assumption by creating the matrix specified by the typical element, $P(ij)$, where i ranges over all categories and j over all behaviors. This matrix may be called the data matrix.

Consider a column of the data matrix; that is, a matrix consisting of the elements, $P(1j)$, $P(2j)$, $P(3j)$, ..., $P(nj)$. This matrix specifies the relative frequency with which the behavior j occurred within each of the various categories under study. In other words, it specifies the relative distribution of this particular behavior through the categories. This matrix may, thus, be called the *pattern* of the behavior, j. Each behavior, of course, has such a pattern. The analytical problem is to discover a set of independent patterns which reproduce the common elements in the original set of observed patterns.

The procedures, usually called factor analysis, for producing such a set of independent behavior patterns are well known and need not be discussed at length here. The correlations between all possible pairs of behavior patterns are first determined. This correlation matrix is then factor analyzed. The factors which emerge are the independent, common behavior patterns from which the correlation matrix may be reproduced, and from which the original behavior patterns (the data matrix) may be reproduced, save for specific elements.

Such a set of independent behavior patterns constitutes what is here called a social space. The position of each category within this space may readily be determined from the factor loadings. The result is a characterization of social space, and a description of the position of the categories within it. The matrix of distances between each pair of categories may be determined either by simple algebra using the data matrix and the factor loadings or by measuring distances directly on a representation or map of the social space. Social structures may be identified either through the distance matrix or in terms of the overall shape of the categories in the social space.

4. An Illustration

To facilitate an understanding of this solution, the method was applied to a set of readily available data. This empirical analysis is primarily an illustration. In particular, it is based upon too few observations to be highly reliable in detail as a description of social space. Further, the factor analysis is only approximate. For illustrative purposes, however, the analysis is entirely adequate.

The population, in the illustration, consists of employed persons living in 10 randomly selected census tracts of Akron, Ohio in 1950. Living in a tract is a behavior. Thus, there are 10 possible behaviors corresponding to the 10 census tracts. These behaviors happen to be mutually exclusive, though they need not be. Social categories are defined as the various occupations (professional, managerial, etc.) in the one column census classification, cross-classified by sex. The problem is to measure the social distance between these occupations. (2)

The first step is to create the data matrix (see Table 1) which indicates the proportion of persons in each occupation living in each tract. The next step is to correlate each pair of columns in this matrix, thus producing the correlation matrix (see Table 2). From the correlation matrix it is clear, for instance, that tracts A-5 and A-8 are very similar in occupational distribution ($\underline{r} = .79$). This fact may be confirmed by examining the first two columns of the data matrix. The third step is to factor analyze the correlation matrix. The factor analysis (via the centroid method) of this matrix yielded two factors (see Table 2) which represent orthogonal directions in social space. The final step is to use these factors to create a map of social space (shown as Figure 1).

Some comments about the procedure and the findings are in order. First, the factors themselves are, initially at least, only arbitrary directions in social space. The distance between each pair of occupations remains constant through any orthogonal rotation. No particular meaning need be attached to each factor; that is, the factors need not be named as long as attention is focused on the relations among the occupations. If the problem is shifted from measuring social distance to explaining how existing distance relations come into being, then an identification of the factors would be appropriate. From this point of view, it is fairly obvious from the map that Factor I measures something that is very similar to the socio-economic status of the occupation. Factor II, based on only two behavior patterns (B-8 and C-6), is less reliably measured and more difficult to identify. It is possible that it is related to the institutional structure of the community, with service personnel low (from private household workers to professionals) and material goods personnel high (from operatives to craftsmen) on this factor. Such an interpretation is highly tentative at best, however.

Second, it is likely that there is at least a third factor underlying residential behavior. The small number of tracts included in this illustration, however, precluded even its tentative identification. Third, it will be observed that the occupations as a whole form a rough semi-circle in the space. This shape cannot be interpreted at the present time, but might possibly prove important in cross-cultural comparisons or in trend analyses.

Table 1. Per Cent of Persons (Classified by Occupation and Sex) Living in each of Ten Randomly Selected Census Tracts in Akron, Ohio, 1950.

Sex and Occupation	Census Tract										
	A-5	A-8	B-8	C-6	D-3	F-1	F-3	F-5	F-7	G-2	Total
Male											
Professional	2	3	4	3	1	19	26	15	3	24	100
Managerial	3	3	3	4	1	23	23	13	4	23	100
Clerical	7	10	11	8	3	7	25	14	8	8	101
Sales	4	5	4	4	2	13	27	16	6	19	100
Crafts	8	13	10	10	6	4	24	11	8	7	101
Operatives	12	18	10	13	12	1	16	6	10	2	100
Service a/	15	20	3	7	15	3	15	6	13	3	100
Laborer	15	23	3	8	21	3	11	4	10	3	101
Female b/											
Professional	12	3	5	3	1	18	19	16	4	18	99
Managerial	7	7	3	5	4	15	19	13	6	21	100
Clerical	5	8	9	8	3	8	26	17	7	10	101
Sales	7	11	7	8	3	6	26	16	9	8	101
Crafts	11	7	12	9	9	4	16	12	10	11	101
Operatives	16	19	9	14	14	1	13	5	9	2	102
Pvt hsld wkrs	12	23	2	2	21	6	8	3	9	15	101
Service	16	20	6	9	14	4	12	5	11	3	100

a/ Includes private household workers.
b/ Female laborers omitted (too few in number for stability).

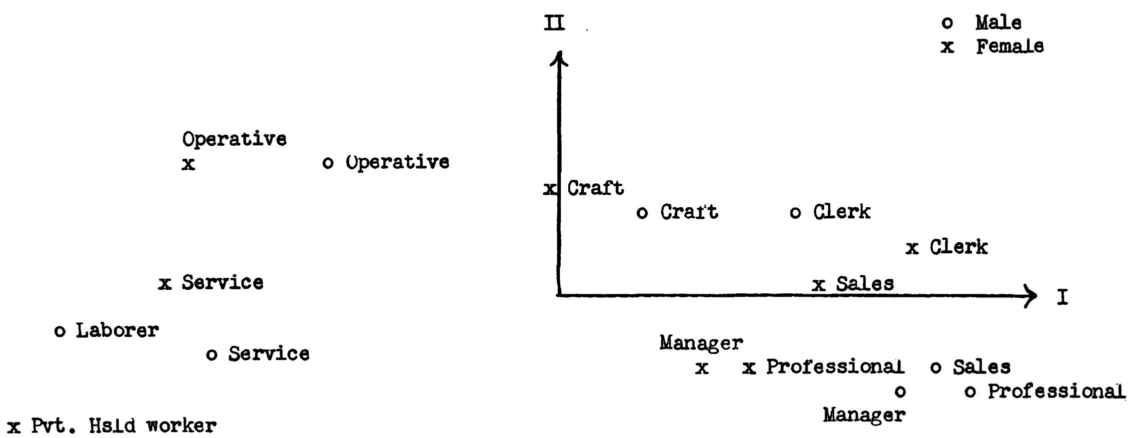
Source: 1950 United States Census of Population, Akron, Ohio Census Tracts, 1950 Population Census Report, Vol. III, Chapter 1. United States Government Printing Office, 1952. Figures adapted from Table 2.

Table 2. Correlation Matrix based on the Akron, Ohio Census Tract Data Matrix and a Rotated Factor Matrix which Approximately Reproduces the Correlation Matrix.

Census Tract	Correlation matrix[a]										Factor matrix[a]	
	A-5	A-8	B-8	C-6	D-3	F-1	F-3	F-5	F-7	G-2	I	II
A-5		79	05	23	81	-69	-86	-80	75	-74	-87	12
A-8			-06	44	94	-80	-77	-91	82	-77	-93	18
B-8				68	-16	-46	20	15	22	-44	12	76
C-6					29	-76	-19	-36	58	-82	-32	84
D-3						-69	-90	-95	74	-62	-98	-02
F-1							49	62	-90	92	70	-68
F-3								90	-59	44	93	19
F-5									-68	59	97	04
F-7										-86	-76	46
G-2											67	-70

a/ Decimal signs omitted.

Figure 1. Map of Social Space Based upon Factorial Structure.



Finally, it will be observed from the map that the high status occupations are more tightly clustered (especially among males) than are the lower status occupations. Thus, the distance from laborer to craftsman is as great as the distance from craftsman to professional. Male professionals, managers, and salesmen are very close together as compared to male laborers, servicemen, and operatives. This finding is not entirely consistent with existing studies of the social hierarchy, which tend to assume greater differentiation near the top of the scale. In interpreting the finding, however, it must be remembered that each occupation is a figure within this space such that the point on the map is the central tendency of the figure. It is possible that the upper occupations are less homogeneous (i.e., would have larger figures) than are the lower occupations. Comparing males and females in the same occupation gives some idea of the spread within each occupation. Discovering the shape of the figures which characterize each occupational category is a problem for future research.

5. An Application

A map of social space, and the distances which can be derived from it, are useful only in so far as they permit the resolution of problems which are otherwise difficult or impossible to solve. To illustrate how this map, or one like it based on more extensive data, might be used, it was decided to apply the P/D (or population divided by distance) model of interaction and interchange to a matrix of social mobility, using social distances derived from the map of social space. Rogoff's (3) 1940 intergenerational mobility matrix, which lists the occupations of about 10,000 sons in Indianapolis classified by the occupations of their fathers, provided the observations. (See Table 3 for details.) The expected values in Table 3 were generated from the model and the map of social space. The average accuracy of reproduction is about 92%,

Table 3. Social Mobility of Sons Observed by Rogoff, Indianapolis, 1940, and Expected by a Model Using Population Divided by Social Distance, where Distance was Raised to the 2/3rds Power.

Son's occupation	Frequency: Father's occupation: Professional Obs.	Professional Exp.	Managerial Obs.	Managerial Exp.	Clerical and sales Obs.	Clerical and sales Exp.	Skilled Obs.	Skilled Exp.	Semi-Skilled Obs.	Semi-Skilled Exp.	Service Obs.	Service Exp.	Unskilled Obs.	Unskilled Exp.
Professional	—	—	133	196	141	112	167	182	70	74	23	21	28	32
Managerial	40	68	—	—	83	100	117	157	63	60	25	18	20	26
Clerical and Sales	152	145	368	377	—	—	520	636	262	218	83	61	94	88
Skilled	100	77	172	194	165	209	—	—	279	308	78	75	111	105
Semi-skilled	59	66	238	154	179	149	734	643	—	—	125	141	216	198
Service	17	13	44	31	35	29	141	107	99	97	—	—	43	64
Unskilled	15	14	30	33	26	30	154	108	81	97	27	46	—	—
Per cent accuracy	91		90		90		91		94		91		94	

Source: (Of observations) Natalie Rogoff, Recent Trends in Occupational Mobility, Glencoe, Ill., The Free Press, 1953: Table in Jacket.

Each expectation is of the form, $aP/D^{2/3}$, where a is a constant of proportionality, P is the total number of sons in each occupation, and D is the social distance between occupations as measured from the map (Figure 1) of social space.

which means that only 8% of the observations would have to be shifted to achieve perfect reproduction. The fact that results as accurate as these occurred despite the unreliability of the measured distances, the shift from Akron to Indianapolis, and the shift from 1950 to 1940 (not to mention the unreliability in the observations) suggests that this method of measuring social distance has considerable promise.

It is important, however, not to exaggerate the importance of the findings to date. Before this procedure can be considered confirmed two types of research need to be performed. First, studies of residential behavior should be conducted in several cities to see if the spatial pattern of the occupations remains essentially constant. Second, studies within other behavioral domains should be conducted, again to see if the results are stable. Investigations along both of these lines are currently in the planning stage.

Footnotes

(1) See, for example, E. S. Bogardus, "Measuring Social Distance," Journal of Applied Sociology, 1925, 9, 299-308 and E. S. Bogardus, The New Social Research, Los Angeles, R. J. Miller, 1926.

(2) For an alternative method of analyzing the same data with roughly the same objectives see Otis Dudley Duncan and Beverly Duncan, "Residential Distribution and Occupational Stratification," American Journal of Sociology, LX, 5, March 1955, 493-503.

(3) Natalie Rogoff, Recent Trends in Occupational Mobility, Glencoe, Ill., The Free Press, 1953: table in jacket.