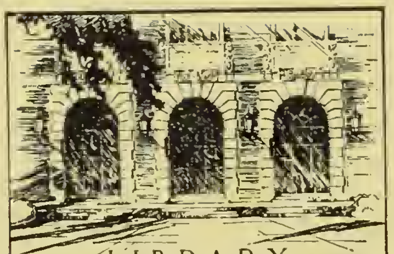


BUREAU OF RESEARCH AND SERVICE
College of Education
University of Illinois
Urbana, Illinois

SIMILARITY BETWEEN PERSONS AND RELATED PROBLEMS OF PROFILE ANALYSIS

Study performed under Contract N6ori-07135
with the Bureau of Naval Research

Lee J. Cronbach
University of Illinois

and

Goldine C. Gleser
Washington University

Technical Report No. 2
April, 1952

SIMILARITY BETWEEN PERSONS AND RELATED PROBLEMS OF PROFILE ANALYSIS*

Lee J. Cronbach, Bureau of Research and Service, College of Education,
University of Illinois

Goldine C. Gleser, Department of Neuropsychiatry, School of Medicine,
Washington University

Studies of personality and behavior are turning increasingly to a simultaneous consideration of several traits or characteristics, and a great many investigations attempt to deal with profiles or patterns of scores.

In this paper we bring together the procedures which may be used for describing relations between such patterns of multiple scores. A comparison of these possible treatments leads to recommendations for improved procedures in future investigations of similarity between persons.

The type of research on which our results bear can be illustrated by reference to several recent studies. One is the effort by Kelly and Fiske (22) to validate certain predictions made in the VA study of clinical psychologists. They compared profiles of assessors' ratings with profiles of criterion ratings. Many studies concerned with classifying patients on the basis of Wechsler-Bellevue profiles have studied the similarity of patterns of scores, and Barnette (1) has compared psychometric profiles of occupational groups.

---

*This study was made in connection with Contract N6ori-07135 between the University of Illinois and the Office of Naval Research, Human Relations Branch. Technical Report #2, April 1952.

Other investigators have been interested in the possibilities of "inverse factor analysis", as introduced by Burt and developed by Stephenson. In Stephenson's hands, the so-called Q-technique (34) has been applied widely to the study of similarity between persons, and to the identification of types of persons. Fiedler and others (17, 18) have used the method not only to compare one person to another, but also to compare various perceptions by the same person. An example is the experiment in which A describes himself along many dimensions, A predicts how B will describe himself, and then B describes himself. Three comparisons are possible, which might be said to indicate the "real similarity" of A and B, A's "assumed similarity" to B, and A's "insight" into B. In addition to the foregoing studies of the equivalence of one person's responses to another's, the statistical devices we consider are relevant to studies of stimulus equivalence. Osgood and Suci (26, 27) for example, is presently employing methods like those we discuss to study semantic problems by demonstrating which words elicit similar association patterns under controlled conditions. As another example, we find that sociometric data may be treated so as to indicate the extent to which two group members see the group in the same way, or so as to indicate the extent to which the two persons are perceived in the same way by the group. That is, we can study the persons as social perceivers and also as perceived objects. The formulas we discuss are relevant to all the foregoing types of investigation.

Despite the rather large number of studies which employ statistical measures of similarity, there has been no comprehensive analysis of the possible alternative procedures. In the present paper we state a general model which clarifies the problem of determining similarity of two score-sets. Within that model we compare the many formulas employed to date, and advance some proposals of our own. In our examination of procedures, we find that some popular methods, such as the procedure of correlating profiles, have serious limitations. The

Other investigators have been interested in the possibilities of [illegible] analysis, as [illegible] by [illegible] and [illegible] of [illegible]. In [illegible], the [illegible] (...) [illegible] in the [illegible] of [illegible] and in the [illegible] of [illegible] of [illegible], [illegible] and [illegible] (11, 12) have used the [illegible] in [illegible] and [illegible] by [illegible] to [illegible] a [illegible]. For example [illegible] (1) [illegible] a [illegible], [illegible] a [illegible] and [illegible], and then [illegible]. These comparisons are possible, which might [illegible] of A and B, [illegible] to [illegible] and [illegible]. In addition to the [illegible] studies of [illegible] to [illegible], the [illegible] of [illegible]. [illegible] and [illegible] (...) for example, [illegible] in [illegible] which [illegible] under [illegible]. It [illegible] that [illegible] the group in the same way, [illegible] and [illegible] in the [illegible] the group. That is, [illegible] and [illegible]. The [illegible] to [illegible] of [illegible].

Besides the [illegible] large number of studies which employ statistical measures of similarity, there has been [illegible] of the [illegible] procedures. [illegible] problem of [illegible] of two [illegible], although [illegible] and [illegible] of [illegible]. In our [illegible] of [illegible], [illegible] methods, such as the [illegible] of [illegible]. The methods of Stephenson and DuMas, in particular, magnify the errors of measurement for some if not most persons, and therefore are likely not to detect some significant relationships.

Techniques to describe similarity between persons are needed for investigating questions such as the following:

1. How similar are Persons 1 and 2?
2. How similar is Person 1 to Group Y?
3. How homogeneous are the members of Group Y?
4. How similar is Group Y to Group Z?
5. How much more homogeneous is Group Y than Group Z? Than combined sample?

Comparable questions may be asked in experimental studies regarding the two or more measures for the same person.

While an index capable of describing the degree of similarity between score sets is necessary for many of the investigations now being pursued, it is often equally or more important to test hypotheses such as, "Group Y and Group Z can be regarded as samples from the same population". The problems of inferential statistics relevant to similarity measures have been thoroughly studied by Fisher, Hotelling, and the Calcutta school. The necessary significance tests and distribution functions are available for normally-distributed variables, and have recently been summarized in a most helpful review by Hodges (20). We shall not discuss the inferential problems, being concerned in our treatment solely with the descriptive formulas for reporting degree of similarity.

A General Model and Notation

A profile or pattern pertaining to a person consists of a set of scores. We shall use the following notation:

methods of Cronbach and Gleser, in particular, measure the degree of correspondence of pairs of test score patterns, and illustrate the [illegible] which [illegible] some [illegible] relationships.

Measures of profile similarity between persons are needed for investigating questions such as the following:

1. How similar are persons A and B?
2. How similar is Person A to Group B?
3. How homogeneous are the members of Group C?
4. How similar is Group A to Group B?
5. How much more homogeneous is Group X than Group Y? (That [illegible] combined sample)

Comparable questions may be asked of measurements obtained concerning the two [illegible] for the same persons.

While an index capable of determining the degree of similarity between profiles is necessary for many of the investigations now being proposed, it is often equally or more important to test hypotheses such as "Group X and Group Y are not [illegible] the same population." The problems of "inferential statistics" relevant to similarity measures have been thoroughly studied by Rulon, Tiedeman, and the Harvard school. The necessary significance tests and discriminant functions are available for normally-distributed variables, and have recently been summarized in a most valuable review by Rulon ([illegible]). We shall not discuss the inferential problems, being concerned in the present study with the descriptive formulas for computing degrees of similarity.

A General Model and Notation

A profile or pattern pertaining to a person consists of a set of scores, obtained on the following measures:

j	any of the variates a, b, c, ... which are k in number
i	any one of the persons 1, 2, ... N
X, Y, Z	classes of persons
x_{ji}	the score of person i on variate j

Considering only two persons, we have the set of x_{j1} (x_{a1}, x_{b1}, ... x_{k1}) for person 1, and the set of x_{j2} for person 2. Without placing any restriction upon our data, we may regard the x_{j1} as the coordinates of a point P_1 in k-dimensional space. The x_{j2} define a point P_2. When the variates are independent they are properly represented by orthogonal axes, whereas correlated variables are more appropriately represented by oblique axes. As two profiles become more similar, the points representing them fall closer together. Accordingly, we define the dissimilarity of two score-sets as the linear distance between the corresponding points.

The formulas to be presented in this section apply to score-sets of many types; viz., responses to a series of items, raw scores on a set of tests, profiles of deviation scores, individuals' ratings of a group of stimuli on a subjective scale, or responses to a Stephenson forced-sort procedure. We shall later discuss the fact that in some of the above, and also in the treatment implied by conventional measures of correlation between persons, points are limited to certain subdivisions of the k-space. The formulas given here are as appropriate for these restricted score-sets as for the unrestricted case.

If we assume the axes to be orthogonal, the distance D between any two points may be easily obtained from its square by use of the generalized Pythagorean rule,

$$D_{12}^2 = \sum_{j=1}^{k} (x_{j1} - x_{j2})^2. \tag{1}$$

In subsequent formulas, we shall often use the symbol Δx_j to refer to the quantity in parentheses. The persons involved in the difference will be obvious from the context. This formula defines the basic measure under consideration in this paper.

We shall show that all the formulas presently used in psychological research involving similarity may be expressed in terms of formula (1) under certain stated restrictions. This points to the fact that in current practice correlations among the variates are generally ignored. We shall discuss later how formula (1) compares to the true measure of distance when score-sets are correlated. However, it can be noted here that when intercorrelations are uniformly low, no serious distortion in the ordering of similarities among a group of individuals is incurred by the use of the orthogonal measure. On the other hand, when one wishes to take the correlations among the variates into account, the advisable procedure is to transform the variates into an uncorrelated set, in which case (1) is fully appropriate. Formula (1) and its derivatives therefore promise to be suited for most psychological investigations of profile similarity.

With any one set of tests, the two most similar persons will have the smallest separation, D, and also the smallest D^2. If the two persons have identical score-sets, D^2 equals zero. If scores on any variate can range from $-\infty$ to ∞, as would be the case with normally distributed variates, D and D^2 can increase without limit. However, the large values have only an infinitesimal probability. D and D^2 result in identical ordering of individuals with respect to dissimilarity.

A particularly interesting distance in k space is that from the centroid of a population to any particular point P_i. This distance, which we shall call the eccentricity of an individual (E_i), is obtained by the formula:

$$E_i = \sqrt{\sum_{j=1}^{k} (x_{ji} - \bar{x}_{j.})^2} \quad . \qquad (2)$$

The expected value of E^2 for the population, that is, the dispersion of all the points about the centroid, is given by

$$\overline{E_i^2} = \sum_j \sigma_j^2 \ , \qquad (3)$$

where σ_j is the standard deviation of variable j in the population.

Since D^2 depends on the number of variates in the set and on the size of their units, it is frequently unsuitable for making comparisons from one score-set to another. A measure of distance in standard units is therefore required. The dispersion of the points in a population provides a "yardstick" for this purpose, since the expected value of D^2 is just twice the dispersion of the population. That is,

$$\overline{D^2_{ii'}} = \overline{2E^2_i} = 2\sum_j \sigma^2_j . \qquad (4)$$

The standard index, which we call S, is defined by the equation:

$$S^2_{12} = \frac{D^2_{12}}{\overline{D^2_{ii'}}} = \frac{D^2_{12}}{2\,\overline{E^2_i}} . \qquad (5)$$

When the measures used in a pattern of scores have been standardized on a large sample, as is true, for example for the Bellevue-Wechsler subtest scores, then the standard deviations of such a reference group may be used to determine $\overline{E^2}$. If, however, the only data available are those for a relatively small sample, then the best estimate of σ_j^2 for the population is

$$\text{est } \sigma_j^2 = \frac{N}{N-1} V_j , \qquad (6)$$

where V_j is the obtained variance in the sample.

$$\text{est } \overline{E^2} = \frac{N}{N-1} \sum_j V_j . \qquad (7)$$

This is the value used in obtaining the standard index S_{12}^2.

S^2, like D^2, can range from zero, for identical score-sets, to infinity. In the population used for reference, the mean of all $S^2_{ii'}$ is 1. The large values of S^2 are decreasingly frequent, and for most types of distribution the probability

of large values is infinitesimal.

<u>Correlated variates and the oblique model.</u> The formulas presented above ignore any correlation among the variates. The variates have been represented in an orthogonal model, with each axis perpendicular to the rest. If correlation were taken into account, the variates would be treated as a set of obliquely inclined axes, such that the angle between axes would be small for highly correlated variates. Such an oblique model is used in mathematical statistics, because it takes into account all the available information. This oblique model underlies the development of the discriminant function, the Hotelling T test, and the generalized distance measure of Mahalanobis (see Hodges (20), Rao (30). We therefore examine how a distance measure based on the more comprehensive oblique model differs from that obtained through the orthogonal model.

The problem which confronted Mahalanobis, and others who have used his technique, was that of determining the distance between two groups. This measure has been used particularly in anthropological research, where the purpose is to study the similarity of racial and tribal groups on physical measurements. His formula is usually written in the following form (using the block $\mathbb{D}$ to distinguish this measure from our D):

$$\mathbb{D}^2 = \Sigma \Sigma\, \alpha^{ij}\, d_i\, d_j$$

To avoid confusion, we can rewrite this in a notation consistent with ours:

$$\mathbb{D}^2 = \sum_j \sum_{j'} \alpha^{jj'} \Delta x_j \Delta x_{j'} \qquad (8)$$

Here $\alpha^{jj'}$ is the jj' element of the inverse of the combined within-group covariance matrix $\alpha_{jj'}$. As Mahalanobis develops the problem, he deals with differences between group means, but formula (8) can also be interpreted as related to the difference between individuals. The Mahalanobis formula gives the same result as would be obtained if the original variates were standardized, and then axes

[illegible]

[illegible]

[illegible]

$$[illegible] \qquad (1)$$

[illegible]

were rotated to any orthogonal set of variates so that formula (1) could be applied. In fact, since the computations required by (8) are impractical for more than a few variates, the usual method of dealing with an oblique space is to make such a transformation and apply (1). Rao (30) suggests one transformation (out of many possible), which is relatively easy to apply. Suppose we begin with variates a, b, c,... expressed in standard measure, and seek an orthogonal set a_o, b_o, c_o,... These equations may be used:

$$a_o = a$$

$$b_o = \frac{b - ar_{ab}}{\sqrt{1 - r^2_{ab}}}$$

$$c_o = \frac{c - ar_{ac} - b_o r_{cb_o}}{\sqrt{1 - r^2_{ac} - r^2_{cb_o}}}$$

etc.

This transformation defines b_o as the portion of b not predicted from a, and c_o as the residual in c not predicted from a and b_o. Then D^2 determined from a_o, b_o, ... is identical to $\mathbb{D}^2$ determined from a, b, ...

We may note several important properties of $\mathbb{D}^2$, or of D^2 obtained from standardized and transformed variates. This measure has a known distribution and thus forms the basis for testing the significance of the difference between groups. It may also be used to determine whether additional variables add significantly to the discrimination between groups. Moreover, $\mathbb{D}^2$ is closely related to Fisher's discriminant function, and particularly to the proportion of individuals classified into the wrong group by the most efficient possible discriminant function; it is therefore a measure of the efficiency of classification. One of the striking features of $\mathbb{D}^2$ is that all orthogonal components in a set of variables

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

have equal weight in the measure. The consequences of this if D^2 is used as an index of similarity between individuals require discussion.

In any set of variates, some variance is likely to be due to general qualities or factors represented in several variates, some is due to common factors present in only a few variates, and some is due to factors found only in a single variate. This unique variance may be due to real traits specific to a single test, or it may represent error variance. Ordinarily, when we wish to investigate similarity in a domain, we are concerned with general qualities found among a population of variables, rather than with characteristics defined by a single sample of items. We would like the similarity index obtained to be reliable from one sample of items to another, so that the same pairs of people will be reported as similar on both occasions. This problem is of greater importance in psychological work, where the number of variates is unlimited and some correlations between them are low, than in anthropological work, where variates are accurately measured and highly intercorrelated, and the total domain under study is relatively restricted. Stability of the similarity index from one set of variates to another demands more consideration in measuring similarity of individuals than in measuring similarity of groups. Unreliable factors will not discriminate appreciably between groups and therefore will not influence D^2 between groups.

Now the Mahalanobis measure, which is designed primarily to capitalize on separation of groups in any reliably measured factor, assigns equal weights to all factors, whether they be general or unique. In a set of physical measurements, it would assign equal weight to such factors as height, breadth with height constant, and so on. If D^2 were applied to measuring the distance between two individuals on the Wechsler-Bellevue test, one factor might be general ability, and a second factor might be an element common among the verbal tests. If there were ten scores in the profile, however, there would be eight other independent factors extracted

and assigned equal weight. Most of these would be specific to particular tests, and many of them would be primarily loaded with error of measurement. Hence D^2 would assign as much weight to differences in these unimportant and perhaps meaningless factors as to the general factor. This means that for particular pairs of persons D^2 will be unreliable from trial to trial and from one set of tests to another set chosen from the same general domain.

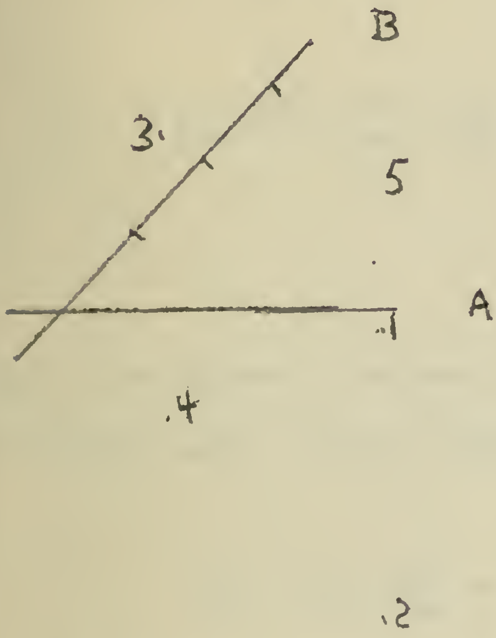
A satisfactory solution which also takes into account the correlation among the original variates is to assign weights to the transformed variates deliberately. Each a_o, b_o, c_o,..., can be assigned a weight according to its apparent importance, before formula (1) is applied. This would be especially feasible if the orthogonal variates were based on a factor analysis, so that the investigator knows which scores represent important general qualities, and which are unimportant residuals. If the investigator knows that the Wechsler profile contains only four factors he wishes to weight in the similarity index, he can assign zero weight to the unimportant and unreliable factors. It is certainly troublesome, however, to transform variates, especially if factor scores must be estimated. Generally, a wiser procedure is for the investigator to make his initial measurements on a set of variates which are nearly uncorrelated, and each of which is important and reliably measured. This requires care in the original planning of an investigation, but once such a set is employed, formula (1) applies directly, and the similarity measures obtained would be generally stable if a second set of instruments measuring the same factors were applied to the same people.

We may next examine what happens if D^2 is applied directly to a set of original standardized correlated variates, with the correlation not being taken into account. This in effect takes axes which are oblique to each other and stretches the space to place them perpendicular (Fig. 1). If we express the resulting measure in

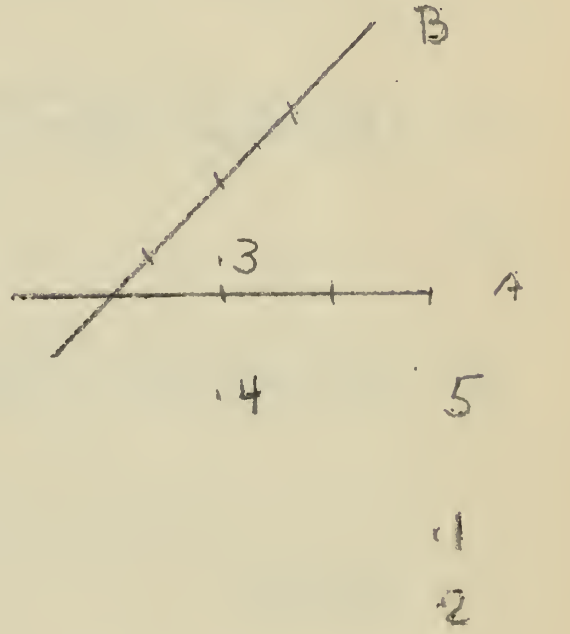
Points	1	2	3	4	5
A	3	3	1	1	3
B	2	0	2	0	3

$\sigma_A = 2 \quad \sigma_B = 4$

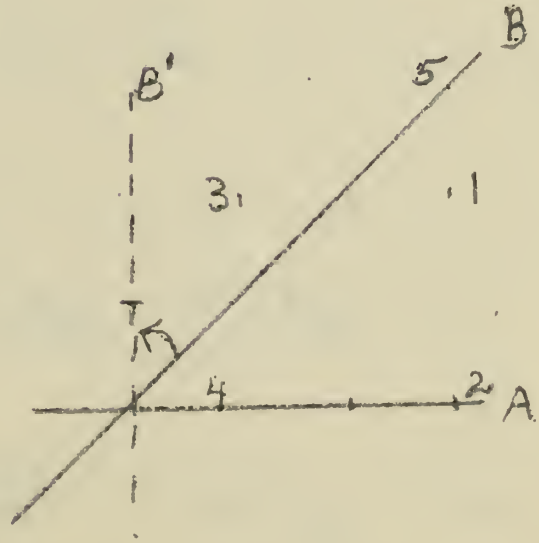
Points	1	2	3	4	5
a	1.5	1.5	0.5	0.5	1.5
b	0.5	0.0	0.5	0.0	0.75



Oblique representation of
original data



Variates standardized
Distances shown are proportional
to the ID measure



Variates treated as if orthogonal
Distances shown are proportional to the D measure

Figure 1. Effect on separation of persons when data are treated
by formulas for ID or D

terms of the variates a_o, b_o, etc. considered above,

$$D^2 = \Delta^2 a + (1 - r_{ab}^2)\Delta^2 b_o + r_{ab}^2 \Delta^2 a + 2 r_{ab}\sqrt{1 - r_{ab}^2}\,\Delta a \Delta b_o$$

$$+ (1 - r_{ac}^2 - r_{cb_o}^2)\Delta^2 c_o + r_{ac}^2 \Delta^2 a + r_{cb_o}^2 \Delta^2 b_o + 2 r_{ac} r_{cb_o} \Delta a \Delta b_o$$

$$+ 2 r_{ac}\sqrt{1 - r_{ac}^2 - r_{cb_o}^2}\,\Delta a \Delta c_o + 2 r_{cb_o}\sqrt{1 - r_{ab}^2 - r_{cb_o}^2}\,\Delta b_o \Delta c_o + \dots \tag{10}$$

Here it is apparent that some factors are weighted more heavily than others. If we collect terms, we find coefficients as follows:

$$\Delta^2 a: 1 + r_{ab}^2 + r_{ac}^2 + \dots \quad \text{(k terms)}$$

$$\Delta^2 b_o: (1 - r_{ab}^2) + r_{cb_o}^2 + r_{db_o}^2 + \dots \quad \text{(k - 1 terms)}$$

$$\Delta^2 c_o: (1 - r_{ab}^2 - r_{cb_o}^2) + r_{dc_o}^2 + \dots \quad \text{(k - 2 terms)}$$

etc.

$$\Delta a \Delta b_o: 2 r_{ab}\sqrt{1 - r_{ab}^2} + 2 r_{ac} r_{cb_o} + 2 r_{ad} r_{db_o} + \dots \quad \text{(k - 1 terms)}$$

$$\Delta b_o \Delta c_o: 2 r_{cb_o}\sqrt{1 - r_{ac}^2 - r_{cb_o}^2} + 2 r_{bd} r_{dc_o} + \dots \quad \text{(k - 2 terms)} \tag{11}$$

etc.

Since by definition a, b_o, c_o, ... are orthogonal, any cross-product terms summed over all pairs of persons in a sizeable sample approaches zero. Hence in considering the weights of the various factors, <u>on the average</u>, we may disregard these terms. Then the coefficients of the square terms indicate the weights of the various factors. If the original variates are left with unequal variance, those variances also would affect the weights. (For example, the terms in $\Delta^2 a$ would be $\sigma_A^2 + \sigma_B^2 r_{ab}^2 + \sigma_C^2 r_{ac}^2 + \ldots$).

It is immediately evident that a factor which appears in several of the original variates receives greater weight in D than a factor which appears in just a few variates. In particular, a unique factor receives relatively little weight, and for that reason D will be more stable than $\mathbb{D}$ from one trial to another or from one set of tests to another.

For any particular pair of individuals it is impossible to evaluate the exact weight of the various factors resulting from the use of D. This is due to the fact that the cross-product terms contribute to the weight of each factor. Since the product of differences may be either positive or negative, the factors make a different contribution to D in each pair of persons. When a large number of variables are involved, these terms will tend to cancel out so that their sum approaches zero. Even so, it cannot be assumed for any given pair of persons that the cross-product terms are negligible. Of course, the more nearly the original variates are uncorrelated the less the cross-products influence the resulting D measure.

Figure 1 sketches the transformations involved in using $\mathbb{D}$ and D as measures of dissimilarity, for two variates with substantial correlation ($r_{AB} = .70$). It is evident that both procedures alter the distance between points, unless we begin with standardized variates (in which case $\mathbb{D}$ preserves the distances unaltered) or begin with orthogonal variates (in which case D preserves the distances

unaltered). In the figure we note, for example, that originally 1 and 4 are closer together than 1 and 2; but for both D and $\mathbb{D}$ measures, 1 and 4 are found to be farther apart than 1 and 2.

The conclusions from our examination of the relationships between $\mathbb{D}$ and D are as follows:

1. While $\mathbb{D}$ is the appropriate statistic to use in testing hypotheses for significance, it is not a desirable descriptive measure of similarity for psychological work because it places excessive weight on unimportant residual factors. It is relatively unsatisfactory for exploratory studies seeking to chart similarity relations in order to formulate hypotheses.

2. D has the advantage over $\mathbb{D}$ that it will tend to be more stable from one sample of variates to another, but the presence of cross-product terms in D makes its psychological composition uncertain for any given pair. This same uncertainty of factorial makeup applies to any other distance measure using an orthogonal model when variates are correlated.

3. If the investigator chooses his variables so that each one is important and so that the set is relatively uncorrelated, then D is quite satisfactory as a descriptive index. D will be stable from one set of reliably-measured variates to another, provided the sets are "parallel" in content, i.e., designed to measure the same factors. (If variates are largely uncorrelated save for a single general factor, an index D_W which we shall introduce later provides for an altered weighting of such a general factor.)

<u>Pearson's CRL.</u> A precursor of the Mahalanobis measure was Pearson's coefficient of racial likeness (28), which was likewise intended to measure distances between groups. In its original form, CRL was essentially the same as our D^2, save that each variate was expressed in standard form, and that a multiplier involving the number of cases per group was included. A modified form of CRL which allows

for correlation among variates was also developed, but was not used because of the computational labor required by it. Except for the factor representing number of cases, it is essentially the same as Mahalanobis' measure.

Many of those who tried to use Pearson's index in anthropological research were dissatisfied with it. The criticisms arising out of its sensitivity to differences in number of cases from group to group are irrelevant to our search for measures of similarity between individuals. Morand, in discussion of a paper by Rao (30), notes that the form of CRL which ignores correlations has given unreasonable results in some anthropological research, notably when the index is determined for groups which are intuitively or theoretically quite dissimilar. This appears, from the context, to be a consequence of the high weight CRL (like D) assigns to any general factor having large loading among the variates. High correlations are usual among anthropometric measures. A solution to this difficulty appears to be an altered weighting for the first component, such as our D_W (see below) provides.

Choice of Scale for the measurement of dissimilarity - Although we have defined D and S as measures of dissimilarity, formulas have been presented in terms of D^2 and S^2. It is evident that for the purpose of description either the linear measures or their squares could be used. Both CRL and D^2 are expressed in terms of the square of the distance. It should be noted, however, that these measures were developed to test whether groups differ significantly with respect to the linear distance between them.

Other metrics might also be chosen as measures of similarity. Cattell proposes a transformation of S^2 such that the values will range from 1 to -1. The usual product-moment correlation between persons may be obtained from D^2 by a transformation of the form $1 - cD^2$. The choice of an appropriate scale depends upon both theoretical and practical considerations.

One desired property of a descriptive index is that it convey to the reader a sense of the magnitude of the quality assessed. D, which is interpretable as for correlation among variables was then developed, but was not used because of the computational labor required by it. Except for the [illegible] [illegible] [illegible] scores, it is essentially the same as [illegible] [illegible].

Many of those who tried to use Pearson's index in anthropological studies were dissatisfied with it. The criticism arising out of its sensitivity to a decrease in number of cases from group to group are irrelevant to [illegible] [illegible] measure of similarity between individuals. [illegible], in [illegible] [illegible] [illegible] [illegible] (19[illegible]), notes that the form of CRL which ignores correlations [illegible] [illegible] [illegible] [illegible] results in some anthropological research, [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] groups which are intuitively or theoretically quite dissimilar. [illegible] [illegible] from the context, to be a consequence of the high weight [illegible] [illegible] [illegible] the general factor having large loading among the variables. [illegible] [illegible] are usual among anthropometric measures. A solution to this difficulty [illegible] is [illegible] weighting for the first component, [illegible] [illegible] [illegible] [illegible] [illegible]

<u>Choice of units for the measurement of dissimilarity</u>. – Although [illegible] [illegible] [illegible] D and D^2 as measures of dissimilarity, formulas have been presented in terms of D^2 and d^2. It is evident that for the purpose of presentation either the [illegible] measures or their squares could be used. Both [illegible] [illegible] [illegible] [illegible] [illegible] of the square of the distance. It should be noted, however, that [illegible] [illegible] were developed to test whether groups differ significantly with respect to the [illegible] distance between them.

Other metrics might also be chosen as measures of [illegible], [illegible] [illegible] a transformation of D^2 such that the values will range from 1 to -1. The usual product-moment correlation between persons may be [illegible] [illegible] [illegible] transformation of the form $1 - cD^2$. The choice of [illegible] [illegible] [illegible] [illegible] both theoretical and practical considerations.

One desired property of a descriptive index is that [illegible] [illegible] [illegible] the [illegible] [illegible] of the [illegible] in the [illegible] [illegible], D, which [illegible] [illegible] linear distance, is particularly suitable in this regard whereas D^2 has no such physical representation. In addition, since distance from a given reference point is measured on an absolute scale, one can choose any particular individual, i, as a reference point and place all other individuals on a scale relevant to him. Then distance has the property that an individual 6 units away from i is 3 units further or twice as far as one 3 units away from i, etc. Such a concept is operationally useful in the investigation of dissimilarity.

A second desirable property is that errors of measurement should be independent of true score and equal on the average for all pairs of persons. This assumption is reasonably accurate, ordinarily, for measures of an individual's score on any variate. But when this assumption holds for the single measures, the distance measure between the score sets of any two persons does not have this property. Pairs with large D tend to have larger absolute error than pairs with small D; this inequality of errors is considerably greater when dissimilarity is measured in D^2 or S^2. A closely related reason for preferring D to D^2 is that in determining averages, variances, and the like, the square measure gives far greater weight to dissimilar than to similar pairs.

A third desirable property is ease of computation. In this, we find that D^2 and S^2 have considerable advantage. A number of simple formulas are available for determining such results as the average D^2 for all pairs in a group, without computing each D^2. No such simple formulas are available for D.

Before reaching a conclusion as to the best scale to employ, let us consider Cattell's transformation and indicate why we have not chosen to use it. This index may be expressed in terms of our measures, although Cattell explicitly restricts the formula to uncorrelated variates expressed in standard deviation units with equal variance. In his notation

$$r_p = \frac{2k - \Sigma d^2}{2k + \Sigma d^2} \qquad (12)$$

Here k is not the number of dimensions, as in our notation, but the median chi-square corresponding to the given number of variates. Σd^2 is the same as our $\Sigma \Delta^2 X_j$, except for Cattell's restrictions upon his variates. If we let K be written for his 2k, we can put r_p in a form for comparison with our S, using k now for number of variates.

$$r_p = \frac{K - 2kS^2}{K + 2kS^2} \quad . \tag{13}$$

r_p ranges from 1 to -1 when S ranges from 0 to ∞. With a large number of variates, $r_p = 0$ corresponds to S = 1. Hence Cattell's index is directly related to ours, but he has compressed the large distances into a small range on the r_p scale. His index is also opposite in direction to ours. Cattell's formulation is dictated by a desire to have a correlation-like index, symmetrically distributed about zero and ranging from 1.00 to -1.00.

A correlation-like index does not appear as advantageous as a distance measure. The reasons are as follows:

1. If our data permit people to be located anywhere in a k-space, no matter how far apart P_1 is from P_2 there is no theoretical reason why there cannot be a P_3 such that $P_1P_3 > P_1P_2$. We see no reason why the measure of separation should be forced to converge toward a limit. "Complete dissimilarity of persons" is an indefinable concept, unless there is some largest and smallest value for each variate. We do not usually expect such limits for traits.

2. The demand for a symmetric measure seems unnecessary; on the contrary, one might anticipate that in a multivariate normal distribution of persons there will be many very similar pairs, and relatively few pairs who are far from each other.

3. Formulas such as we have for mean D^2, etc., are not possible with r_p.

4. r_p lacks the usual properties and advantages of correlation.

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

In view of its operational properties and its more uniform error for all pairs of persons, we conclude that the linear measure D (or its standard form S) is the best descriptive index of dissimilarity between persons. It will be noted, however, that D, D^2, r_p, and many other transformations give identical results so far as the ordering of dissimilarity is concerned. The investigator who prefers to stop with D^2 or S^2 rather than take the square root should use non-parametric statistical methods to analyze the results. This applies also to Q correlations, if those are obtained, and DuMas' r_{ps}. Non-parametric procedures include computation of medians, chi-square, and many significance tests which have recently been reviewed by Moses (24).

For correlating the dissimilarity measure with a criterion, it is advisable to express the measure in terms of D or S and apply the product-moment formula. It is not wise to use such procedures as computation of means, product-moment correlation, analysis of variance, and the t-test with D^2 or S^2, if the distribution is skewed appreciably. Skewness will be small and no serious error will be introduced if all S^2 are fairly near to 1.00, but this will ordinarily occur only for restricted types of data. While we shall present short-cut formulas based on mean D^2, these formulas should ordinarily be used only in rough comparisons, where the saving of time they afford compensates for the fact that they assign large weight to the most distant pairs and emphasize errors of measurement for such pairs. These formulas also are of some use for checking computations.

Reduction of Data in Profiles using Derived Scores. If raw scores or standard scores are entered in the formulas we have been discussing, we examine all the information about individual differences which the data provide. This procedure has been recommended by Cattell and by DuMas (12), but has rarely been followed in psychological work. Instead, the more common practice is to study similarity of profile "shape", disregarding differences in the overall level of scores for the

person, or to study similarity of shape after equating profiles in terms of both level and variability within the person (as in Q technique). In effect, investigators who choose procedures of this sort are studying profiles of scores restricted to spaces of k-1 or k-2 dimensions, respectively. Such restriction may or may not be wise in a particular investigation. The ensuing section analyzes the various methods by which investigators reduce the data given them, so as to give a clearer picture of the special effects of each method.

Each derived score treatment "projects" the scores into a more restricted space. A summary of information about the various treatments, which we shall develop gradually, is presented in Table 1. Figure 2 provides a series of sketches to illustrate the discussion.

If we begin with raw scores for each person on k variates, each person can be represented by a point in a k-dimensional space, as sketched in the first panel of Figure 1. Two specific persons, P and Q, are located. The point O is the centroid, whose coordinates are $\overline{x_{1.}}$, $\overline{x_{2.}}$, ... The point C is the origin.

We shall now define certain terms necessary to our later explanation.

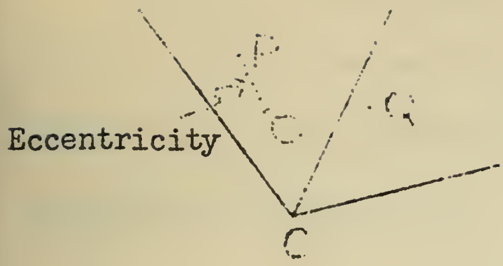
Elevation is the mean of all scores for a person $(\overline{x_{.i}})$.

Eccentricity is the square root of the sum of squares of the individual's deviation scores from the group means $\bar{x}_{j.}$ Geometrically, it is the person's distance from the centroid, as shown in the figure. If E_i represents the eccentricity of person i,

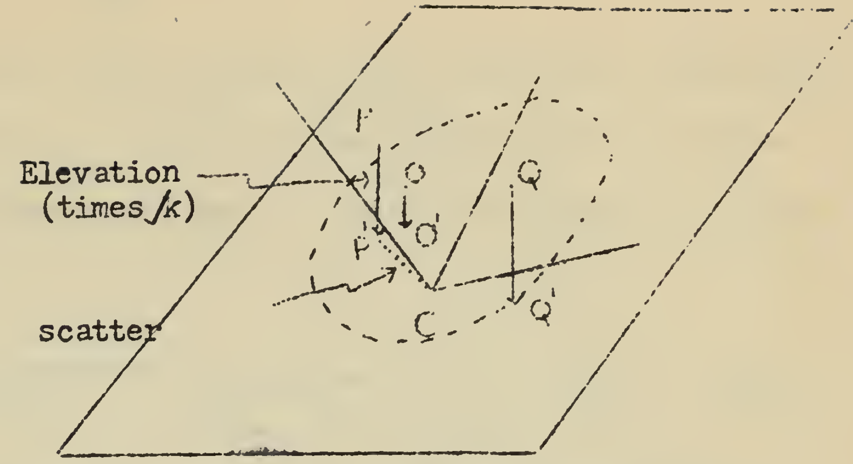
$$E_i = \sqrt{\sum_j (x_{ji} - x_{j.})^2} \quad . \tag{14}$$

Table 1. Summary of Systems for Reducing Profiles

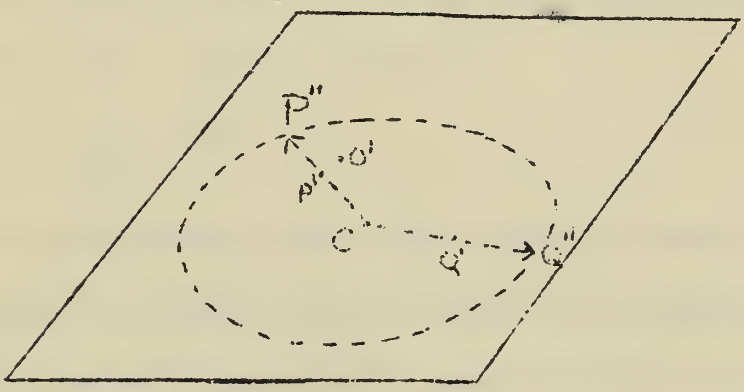
Score sets restricted to	Symbol	Scores used in profile	Relation defining high similarity	Differences which lower similarity	Differences not influencing similarity	Procedure having this character
Unrestricted k space	D, S	Raw or standard	Identical score sets	Elevation Scatter or eccentricity Shape	----------	D^2 (Mahalanobis) D or S; CRL r_p (Cattell)
k - 1 hyperplane	D'	Deviations about the person's mean	Parallel profiles	Scatter Shape	Elevation	Pattern tabulation (9) Analysis of Wechsler profile (3) Analysis of Q covariance (5)
k - 1 hypersphere, center at absolute origin	$D^{\cdot}$	Extended vector coordinates	Proportional profiles of raw measures	Elevation Shape	Size	Similarity as measured in Geometry
k - 1 hypersphere, center at centroid of k space	$D^{:}$	Extended vector coordinates	Proportional profiles of deviations about group means	Elevation Shape	Eccentricity	
k - 2 hypersphere, center at point in k-1 plane where all j equal	D"	Deviations about own mean, standardized for each person	Proportional profiles deviations about own mean	Shape	Elevation Scatter Eccentricity	Correlation (Q, rho, Tau) Ranking of variates Forced distribution (Stephenson) r_{ps} (DuMas)



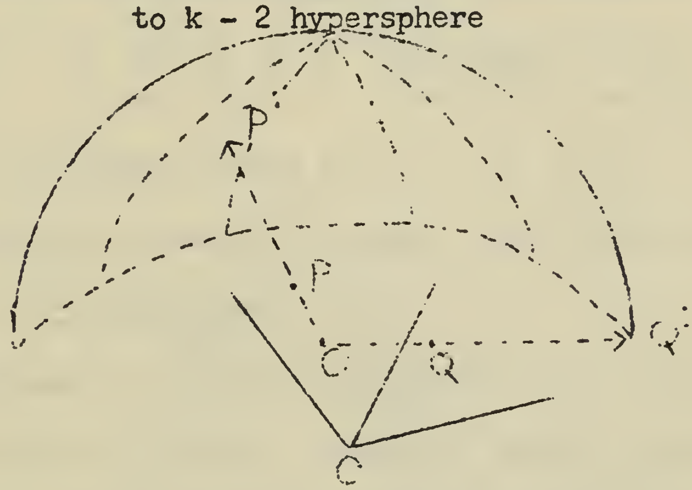
D : Data in k space



D' : Data projected on k - 1 hyperplane



D" : Data projected from k - 1 hyperplane
to k - 2 hypersphere



D: : Data projected from k space
to k - 1 hypersphere

Figure 2. Projections implied by various distance measures

Scatter is the square root of the sum of squares of the individual's deviation scores about his own mean $\overline{x}_{.i}$. That is, it is $\sqrt{k}$ times the standard deviation within the profile. Using E_i' to represent scatter, and primes to represent scores expressed in deviate form,

$$E_i' = \sqrt{\sum_j (x_{ji} - \overline{x}_{.i})^2} = \sqrt{\sum x'^2_{ji}} \qquad (15)$$

Shape is the residual information in the score-set after equating profiles for both elevation and scatter.

When we change scores to deviations about the person's mean, we develop new scores x'_{ji}, which are subject to the linear constraint

$$\sum_j x'_{ji} = 0.$$

This removes from the scores any information about the person's average. Even though there are k scores still, there are only k-1 degrees of freedom. Two people whose scores in each j are separated by a constant amount have the same profile of x'_j. For example, suppose score sets in k space are as follows:

For person 1:	2	-2	0	3	2	(Elevation is 1)
For person 2:	0	-4	-2	1	0	(Elevation is -1) ,

Then the deviation score-set for either person is 1 -3 -1 2 1 .
For these people, D^2 in k space is 20, but D'^2 is zero. We shall use the prime to refer to measures in the k - 1 hyperplane.

When a profile is subjected to one linear constraint, we have in effect, projected the points into a space of one less dimension, which we refer to as a k - 1 (dimensional) hyperplane. We place our hyperplane through the origin, perpendicular to the direction representing the elevation factor. Each point P projects into a new point (P') as the figure shows. The distance PP' is the elevation, and CP' is the scatter.

It should be clear that k and k-1 spaces yield different information about the similarity of persons, although measures of similarity in both spaces may be of value. Comparison of deviation scores is most frequently found in psychology in studies of Wechsler-Bellevue profiles, where attempts are made to interpret the shape and variability within a subject's profile. Burt also deals with such deviation scores when he employs covariances rather than correlations between persons to obtain a matrix which he then factors into types (5).

Before discussing further the measurement of dissimilarity in k-1 hyperplane, let us consider how such score-sets may be constrained to lie on a k-2 (dimensional) hypersphere. A hypersphere is the locus in space of points all of which have the same distance from some center. This geometric property is imposed whenever all score-sets are subject to the constraint that the sum of squares for each set is a constant. But it may easily be seen that this is precisely the type of constraint which is imposed by standardizing a set of scores; i.e., dividing by their standard deviation. Dividing by the scatter of the profile has a similar effect. If we divide each deviation score for an individual by his scatter, this results in a score set (x''_{ji}) for which

$$\sum_j x''^2_{ji} = \sum_j \frac{x'^2_{ji}}{\sqrt{\sum_j x'^2_{ji}}^{\,2}} = 1 \qquad (16)$$

Since, whenever scores are constrained as in a set of x''_{ji}, the sum of squares is a constant, differences in scatter among persons have been eliminated from consideration, just as differences in elevation are eliminated when scores are expressed as deviations from the person's mean. Conversion of score-sets from deviation scores to sets of x''_j has projected points from the k - 1 hyperplane into a k - 2 hypersphere with unit radius. This is sketched in the third panel of Figure 2. (Because our sketch is based on a set of only three variates, k - 2 is only one,

and the sphere in this instance is reduced to a circle.) We define the measure of dissimilarity (D") on the k - 2 hypersphere as the distance between score-sets having unit-scatter. We might have divided scores by their standard deviations, which would have placed all points on a sphere of radius $\sqrt{k}$. Distances on this sphere would be a constant multiple of corresponding distances on the unit sphere.

Eliminating differences in scatter from consideration is widespread in present statistical studies of profiles in psychology. Sometimes this is done consciously, as when Stephenson asks subjects to sort descriptive statements into piles with a fixed number of statements per pile, so that the resulting scores for each person have the same standard deviation. More commonly, standardization is introduced through a correlation formula. The product-moment formula, for example, divides cross-products by the product of the standard deviations, and thus standardizes. Other formulas such as rho, Tau, and r_{ps} have the same effect. Our diagram shows how points P and Q, which appeared reasonably near each other in k space because they are quite similar in elevation, are found to be fairly distant from each other when measured in k - 1 hyperplane; and diametrically placed, i.e., virtually as dissimilar as possible, in k - 2 hypersphere. Differences of this sort make it imperative for the investigator to decide on a rational basis which type of score-set is to be his basis for studying the relation between persons.

The k-1 hyperspheres, which have not been used in psychological work, have properties of considerable interest. Such a distribution of points is obtained by dividing the original set of scores for each person by the square root of the sum of squares. If the original variates are measured in meaningful units with an absolute zero, then the square root of the sum of squares, which represents the distance of a point from the origin, might be considered to be a measure of overall "size." Division by this measure extends all points to unit distance from the origin. Two score sets which are in the same proportion, such as (4, 8, 2) and (2, 4, 1) lie on the same vector and thus project to the same point on the hypersphere.

[illegible]

[illegible]

[illegible]

Thus proportional score sets have zero dissimilarity as measured on this hypersphere, just as geometric figures for which corresponding sides are in proportion are termed "similar". Dissimilarity may be measured in terms of the distance between the points on the unit hypersphere ($D^{\bullet}$) or in terms of the cosine of the angle between the vectors.

More appropriate to psychological data is projection onto the hypersphere with the centroid of the population as center. This projection is achieved by dividing each score by the eccentricity. Thus differences in eccentricity are removed from consideration. All persons who deviate in the same direction from the group average are projected into the same point and thus are considered to be the same "type". The measure of separation on this hypersphere is designated $D^{:}$.

Relations between measures of dissimilarity for original and derived score-sets. Formula (1) for D, and (5) for S, are equally correct whether data occupy the k space or are confined to a smaller space by one or more constraints. It is of value to compare the indices by examining the effect of treating the same set of data successively in the various spaces. We begin with the relation between D and D'.

$$D'^2_{12} = \sum_j \Delta^2 x'_j = \sum_j \Delta^2 x_j - k\,\Delta^2 \bar{x}. \qquad (17)$$

The first member on the right-hand side is D^2 and the second component is proportional to the difference in elevation, $\Delta^2 x.$; i.e.,

$$D'^2_{12} = D^2_{12} - k\,\Delta^2 \bar{x}. \qquad (18)$$

On the average over all pairs,

$$\overline{D'^2_{ii'}} = 2 \sum_j \sigma^2_j - 2k\,\sigma^2_{\bar{x}.} \qquad (19)$$

Here, $\sigma^2_{\bar{x}.}$ is the variance of elevation scores, over the population of persons.

$$S'^2 = \frac{D^2 - k\,\Delta^2 \bar{x}.}{2(\Sigma\,\sigma^2 - k\sigma^2_{\bar{x}}\ )} \qquad (20)$$

These relationships between measures of dissimilarity in k and k - 1 space suggest the possibility of constructing a new measure of dissimilarity in which elevation is given any desired weight w. Such a measure would allow for weighting the elevation and shape factors to predict a particular criterion, if one is available. It also permits reducing the excessive weight the elevation factor receives when variates are substantially correlated, as for the investigations where Morand found difficulties with CRL. Suppose we denote the new measure of distance by D_w.

$$D_w^2 = D'^2 + w k \Delta^2 \bar{x}. \tag{21}$$

and

$$S_w^2 = \frac{D'^2 + w k \Delta^2 \bar{x}.}{2(\Sigma \sigma_j^2 + w k \sigma_{\bar{x}.}^2)} = \frac{D^2 - k(1-w) \Delta^2 \bar{x}.}{2(\Sigma \sigma_j^2 - k(1-w) \sigma_{\bar{x}}^2)} \tag{22}$$

When w is zero, D_w reflects differences in shape and scatter only; as w approaches 1, D_w approaches D. <u>Because of its flexibility, formula (22) (or its numerator alone) appears to be the most suitable basis for determining similarity of persons.</u> We shall discuss below some reasons for this recommendation.

The relationship of the measure of distance in k-space to that on the k-1 hypersphere may be derived from the law of cosines. For the hypersphere with center at the centroid and radius equal to unity,

$$D_{12}^{:2} = \frac{D_{12}^2 - (E_1 - E_2)^2}{E_1 E_2} = \frac{D_{12}^2 - \Delta^2 E}{E_1 E_2}, \tag{23}$$

where ΔE is the difference in eccentricity of the two individuals. Since the distance measure is defined for a unit hypersphere, the values of $D^{:}$ have a possible range of 0 to 2, regardless of the number of variates involved. Thus, when score-sets are divided by their sum of squares the measure of distance is comparable from one set of variates to another and there is no need of further standardization by a measure such as $S^{:}$.

These relationships between [illegible] at [illegible] $T = 1$ and $T = 2$ which suggest the possibility [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]

[illegible]

and

[illegible]

[illegible] [illegible] [illegible] [illegible]

The [illegible] [illegible] [illegible] [illegible]

[illegible]

[illegible] [illegible] [illegible] [illegible] [illegible]

The relationship between D' (in the k-1 hyperplane) and D" (k-2 hypersphere) is analogous to that which holds between D and D'.

$$D''^2 = \frac{D'^2 - (E'_1 - E'_2)^2}{E'_1 E'_2} = \frac{D'^2 - \Delta^2 E'}{E'_1 E'_2}, \qquad (24)$$

where E' is the measure of scatter.

D" may be written in terms of the D measure from k-space as

$$D''^2 = \frac{D^2 - k\,\Delta^2 x_{j.} - \Delta^2 E'}{E'_1 E'_2} \qquad (25)$$

This formula shows clearly what types of differences between individuals, represented in the original data and in D^2, are discarded when we employ only k-2 space information. One of the subtracted terms represents differences in elevation; the other represents differences in scatter.

Here, again, since we have defined D" as a measurement on the unit sphere, the values range from 0 to 2 ($0 \leq D''^2 \leq 4$). The expected value of D''^2 is

$$\overline{D''^2_{ij}} = 2(1 - k\,\sigma^2_{\bar{x}''_{j.}}) \quad , \qquad (26)$$

where $\sigma^2_{\bar{x}''_{j.}}$ is the variance, over variates, of the means of the scores after division by the scatter. It may be noted that the average value is 2, if and only if all variates (in derived score units) have equal means over all persons. This is of particular interest because of the close relationship of D''^2 to the correlation measure frequently used to show similarity between persons, which we now discuss.

Comparison of measures in k-2 space. Let us specifically consider the relationship between Q, the correlation between persons, and D", It is easily shown that

$$D''^2 = 2(1 - Q) \qquad (27)$$

Thus it is evident that our formulation in terms of D''^2 encompasses any results obtained by product-moment correlation among profiles, including those from Stephenson's forced-sort method. In particular, any distortions imposed by use of an orthogonal model for correlated variates will affect studies using correlation between persons.

It may be noted that average Q for a population is zero when $\overline{D''^2} = 2$. But we have seen that this is true only when all variates have equal means. Thus, if items of unequal popularity are chosen for the sample of traits, the expected value of Q is greater than 0. Inclusion of some items on which members of the sample tend to agree will increase the correlation between individuals. Some implications of this will be discussed later.

The three prominent correlational procedures using ranking are rho, Tau (23), and DuMas' r_{ps} (13). Rank-correlations are sometimes used in the belief that assumptions regarding the test score metric are thereby avoided.. This is not the case for rho. When each score is assigned a rank, the separation between two adjacent ranks is fixed, over the whole range. The result is that all profiles are forced into the same rectangular distribution, just as Stephenson's forced-choice sorting forces all profiles into the same normal distribution. Such forced distributions appear to be fully justified only if the investigator regards a particular distribution as most likely to represent the nature of his profiles. Usually rho and product-moment correlations give about the same results for a particular set of data.

Kendall's Tau gives values substantially lower than rho. It is a rank correlation based on the direction of differences between all possible pairs of variates. Tau is preferred to rho in some studies because its sampling distribution is known. If $Tau_{12} > Tau_{13}$, then $rho_{12} > rho_{13}$, in almost all pairs of cases. That is to say, Tau is very nearly a function of rho. Analysis by Tau will therefore yield conclusions very like those from rho, and both of these will be reasonably close to

Thus it is evident that [illegible] obtained by [illegible] among problems, [illegible] method. In particular, any distortions [illegible] will affect [illegible] between problems.

It can be shown that [illegible] for a population [illegible] but we have seen that this is true only when all [illegible] have equal means. Thus if items of unequal popularity are chosen for the sample of tests, the expected value of [illegible] is greater than [illegible]. Implications of this will be discussed later.

The [illegible] [illegible] the same [illegible] distribution, [illegible] normal distribution. [illegible] [illegible] give about the same results [illegible].

[illegible] It is a [illegible] of differences between all possible pairs of [illegible] [illegible] [illegible] [illegible] results from D''^2 and Q. One difficulty with Tau is that the number of comparisons which must be made increase rapidly with the number of variates.

The third coefficient, proposed by DuMas*, is simple to compute. Kelly and Fiske drew our attention to the fact that r_{ps} is an approximation of sorts to Tau. Whereas Tau calls for considering all possible pairs of variates, r_{ps} uses only the adjacent variates. I.e., if a profile is written in a certain order (Computational, Scientific, Mechanical, ...), r_{ps} would consider the direction of difference between Computational and Scientific, and Scientific and Mechanical, but would not use the difference of Computational and Mechanical. Rearranging the profile in different order would change the correlation, for different pairs would now be used. If the arrangement of variates is a random selection out of all possible orders, or if the variates are uncorrelated, r_{ps} is an estimate of Tau. If there is any rationale underlying the arrangement, r_{ps} is peculiarly biased. Consider a Wechsler profile of five verbal and five performance scores. These will conventionally be listed in that order. For this profile, Tau would be based on 45 pairs of scores: 10 verbal with verbal, 10 performance with performance, and 25 verbal with performance. r_{ps} would use only nine pairs: 4 V V, 4 P P, 1 V P. In this example, r_{ps} is determined almost wholly by the smallest differences in scores, which are least reliable. r_{ps} would therefore be lower than Tau for Wechsler profiles, and possibly by a large amount. Because it uses relatively little information (here, 9 pairs out of 45), r_{ps} is expected to be inexact even when it is unbiased.

---

*Incidentally, DuMas (13) suggests chi-square as the preferred method of estimating similarity where a more precise approach is required. This suggestion is unsound, since profile entries are scores rather than frequencies and chi-square cannot be used with such data.

Basic Decisions in Profile Comparison

The comparison of two profiles will give different results, depending upon the investigator's choices at several points in the planning of the investigation. During the development of techniques of Q correlation, there has been some confusion and dispute regarding these matters, but at this point, Burt and Stephenson, at least, seem to be in agreement on the principles underlying the method. Many of the issues have been discussed with exceptional soundness by Burt, in *The Factors of the Mind*, Chapters VI and XI (30). Any one who proposes to study relations between persons by Q correlation or other measures should examine Chapter VI with care. Although Burt discusses specifically the use of Q correlation in factor analysis, the same questions regarding metric and domain apply to any descriptive studies of relation between persons.

The investigator must define a trait-domain within which similarity is to be investigated. There is a certain amount of loose thinking regarding the concept of similarity of persons which occasionally leads investigators to regard their studies as an attempt to determine which persons are *generally* similar. Such views are encouraged by occasional references to Q-technique as a method for studying "the whole personality". Actually, the investigator, either by plan or by the necessary limitations of any instrument can study only a relatively limited segment of the person, and it will be noted that Stephenson himself now places great emphasis on the proper definition of the segment of personality to be investigated.

The investigator defines the domain where he is seeking to investigate similarity by four choices:

1. He chooses the set of variates.
2. He chooses a metric for each variate.
3. He assigns equal or differential weights to each variate.
4. He decides to study similarity in k space or in some restricted portion of the k space.

[illegible] in Profile Comparisons

The computation of the profiles will give different results, depending upon the investigator's choices at several points in the planning of the investigation. During the development of techniques of Q correlation, there has been some confusion and [illegible] regarding these matters, but at this point, [illegible] and [illegible], at least, [illegible] in agreement on the principles underlying the methods. Some of the issues have been discussed with [illegible] soundness by Burt, in the Factors of the Mind (Chapters IV and VI [illegible]). [illegible] for use with [illegible] of Q correlation [illegible] Chapter VI [illegible]. Although [illegible] of Q correlation [illegible], the same decisions regarding [illegible] apply to the [illegible] of relation between persons.

The investigator must define a [illegible] which similarity is to be [illegible]. There is a [illegible] amount of [illegible] regarding the concept of similarity of [illegible] investigators [illegible] are generally similar. [illegible] to [illegible] as a method for studying "[illegible] personality"; [illegible] the investigator [illegible] of the [illegible] and [illegible] [illegible] will be [illegible] of the concept of personality [illegible].

The investigator defines the [illegible] by [illegible] choices:

1. To choose the set of variables.
2. To convert a [illegible] variables.
3. To [illegible] of differential weights to these variables.
4. To [illegible] each [illegible] of the [illegible] [illegible].

The investigator can make each of these decisions quite arbitrarily, but he is more likely to arrive at useful and scientifically meaningful results if he has a carefully-considered reason for each decision. Different decisions will be arrived at in different problems. In the use of objectively measured variates, such as are used in anthropometric studies, the appropriate decision may be different from the decision reached in designing a study of subjective estimates of personality. This is a departure from Stephenson's view, since he always employs variates restricted to k - 2 space.

Choice of variates. Similarity is always similarity in some respects. If we know that two people are quite similar in ten different characteristics, we cannot infer that, on some other set of characteristics uncorrelated with the first ones, they will be any more similar than randomly selected people. Since the number of characteristics which might be the object of study is essentially unlimited, it is reasonable to expect that people who are similar in one respect will be quite dissimilar in some other domains of behavior. The domain to be studied will have to be selected with care.* One group of qualities especially promising for investigations of similarity are pervasive and general variables which affect performance in many situations; examples are general mental ability, cultural background, and some of the commonly identified personality traits. Another type of variate which may be profitably used is the more specific qualities which seem likely to be associated with some criterion performance with which the experimenter is concerned. For example, in study of performance of a military group, it might be appropriate

---

*This comment also applies to measures of "empathy" or "diagnostic accuracy." There is little evidence that the person who is able to judge one quality is also a superior judge of other qualities. (8)

to determine the similarity of members' attitudes toward being in military service, or attitudes toward military discipline. Having defined a domain of traits in which he is interested, the investigator might, in theory, draw a random sample of traits to be measured. He obtains much greater control over his investigation if he uses a planned or stratified sample in which deliberately chosen characteristics are measured as reliably as possible. Such a procedure is exemplified in Stephenson's recent use of a "factorial design" for selecting variates (34).

In general, the more frequently a quality is represented in the set of variates, the more weight it has in the similarity measure. If items are grouped into uncorrelated subtests, each of which has known variance, the variance of the subtest indicates its relative weight in the total. It is therefore appropriate to include a greater number of items dealing with qualities which seem especially important for the investigation.

Choices regarding metric. All psychometric studies must make some assumption regarding the metric or scale units in which the variates are measured. Except in very limited problems, psychologists and educators have lacked scales with equal units, or scales in which a unit on one scale is exactly comparable to a unit on the other scale, representing equal amounts of the properties being measured. In fact, it is doubtful whether comparability between scales can ever be established save by arbitrary assumption. Yet, any study of similarity of persons demands assumptions of comparable units along a scale and between scales.

The investigator must choose for each variate a scale such that he regards one unit as representing the same amount of the property at all points of the scale. If the investigator does not regard this assumption as valid for some scale, he should transform the units to a scale he regards as more nearly linear with respect to the property measured. In most psychological studies, so much error is present to obscure relationships that failure to obtain a scale of equal intervals will have little effect upon the conclusions. When studies turn to more precisely measured variables to determine the similarity of positions. Considerable research among the military [illegible] of structural [illegible] military occupations, [illegible] a number of [illegible] [illegible] of [illegible], the [illegible]; however, [illegible] [illegible] be [illegible]. [illegible] [illegible] [illegible] [illegible] [illegible] cannot [illegible] of a [illegible] (11).

In general, the more frequently a position is encountered in the set of [illegible], the more weight to give to the [illegible]. If [illegible] [illegible] content [illegible] relative weight to the [illegible]. It [illegible] include a greater number of items [illegible] methods for the [illegible].

[illegible] regarding the [illegible] very [illegible], psychologists, and [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] personnel. [illegible] the degree of similarity [illegible] [illegible]

The [illegible] such as [illegible] the [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] variables than present psychological tests afford, this question becomes more crucial.

The second assumption is peculiar to profile analysis: the investigator must assume that one unit represents the same degree of similarity on all variates. Comparability of metrics is unlikely to be testable in most problems. The assumption enters research on similarity because a two-point difference in Block Design, for instance, has the same effect on the index of profile similarity as does a two-point difference in Arithmetic. The use of standard scores is only a device to improve on manifestly non-comparable raw score units; the new units may also lack perfect comparability, and to that degree studies of profile similarity contain error. The investigator may modify the units to make them more comparable to one another, in whatever respect concerns him. If he regards one measure as more important than another in an overall estimate of similarity, he may deliberately assign larger units to that variate.

When an investigation deals with objectively measured variates, such as physical measures or test scores, the metric is altered by standardizing, weighting, and other transformations. We can illustrate choice of a metric by referring to the Rorschach M score. When an investigator uses raw score units, he is counting the difference from zero M to 3 M as equal to a difference from 30 M to 33 M. If he normalizes the score, he will weight the former difference more heavily because the raw-score distribution is skewed. Normalizing would be advisable if, on psychological grounds, the investigator regards the difference from 0 to 3 as more important than the second difference. No general recommendation can be made as to whether a variate distribution should be normalized or not.

The variance of each characteristic over persons, and therefore its influence on the D measure, will be determined by the choice of units. If variates are expressed in standard units, each variate is assigned equal weight. Now sometimes this is quite appropriate; it is common in identifying physical types to express length of

nose and length of limb in standard units so that they make equal contributions. Conceivably, in a study of resemblance in appearance, such equal variation would be inappropriate. Does a person who departs from the average by one standard deviation in length of eye-brow seem as distinctive as one who departs by one s. d. in length of nose? The selection of weights is ordinarily arbitrary, and equal weights (i.e., standardization of variates) is often the best arbitrary choice. If a criterion is available, optimal predictive weights may be selected. The discriminant function is a device for weighting variates to maximize separation between criterion groups.

When the measurements are subjective, the choice of metric presents further difficulties. Subjective ratings are used in studies of esthetic preferences, in self-ratings, or in ratings of other persons. One may standardize the ratings assigned on each variate, but this assumes that the stimuli judged are equally variable on each quality. Perhaps it is more reasonable to suppose, for example, that pupils vary much more in sociability than in obedience. Ratings of different qualities can sometimes be made more comparable by defining the points along the rating scale explicitly. Sometimes the ratings by a person can be expressed in terms of his j.n.d. Sometimes one can accept the rating scale as a scale of equal-appearing intervals. After comparable subjective judgments on the several variates are obtained, differential weights according to supposed importance may be assigned if desired.

Inclusion of elevation in the difference measure. The domain is further defined by the decision to use k, k - 1, or k - 2 space. Elevation is defined by the average of a person's scores. It has an obvious meaning in the Wechsler test, where elevation is essentially an overall measure of ability. In a Rorschach score-set, elevation represents responsiveness, being highly correlated with total R. Holzinger (21) has demonstrated that the average of scores is heavily loaded with the first principal component of the scores, i.e., with the general factor or other frequently represented factor. Thus if scores are correlated, elevation represents the common thread among them. On the other hand, if scores have low correlations, the elevation [illegible] that they were equal [illegible]. Conceivably, [illegible] of [illegible] appropriate. [illegible] in terms of [illegible] of [illegible]. The [illegible] is [illegible], and [illegible] variables, [illegible].

When the [illegible] and [illegible], the [illegible] [illegible], [illegible] [illegible]. [illegible] [illegible].

Isolation of [illegible]. The [illegible] of [illegible] [illegible]. This is [illegible] the [illegible] [illegible]. On the other hand, if [illegible] the [illegible], the [illegible] score represents a mixture of factors and has no interpretable significance. A ratio based on the sum of interitem covariances has been suggested (11,19) as an index which generally reflects the extent to which elevation represents common factors. If this ratio is large, one can regard elevation as saturated with some psychological quality. If this ratio is small, however, elevation lacks psychological meaning. In fact, the elevation component may be purely arbitrary if scores are uncorrelated. For instance, many personality profiles could be scored as logically if the direction of some variables were reversed, submission being counted instead of dominance, for example. This would lower the elevation (average on all traits) for very dominant persons, and raise it for submissive ones. Such reversals do not alter D in k space, but they do affect D' and D". Stephenson attempts to avoid this problem when he obtains self-descriptions from a set of variates which includes a statement and another nearly opposite in sense. He might have one submissive statement, and an opposite dominant one. For such a balanced set of variates, the elevation should be near zero for each person, and any non-zero elevation score could be safely disregarded as due to inconsistency of response.

Elevation can be considered a meaningful score rather than an arbitrary composite only when variates are generally correlated, so that the "positive" direction of each can be determined operationally. When the elevation score is interpretable, one can decide whether differences in elevation should affect D. Sometimes it is wise to include elevation in the difference measure and sometimes it is unwanted. Often the elevation score represents a response set (10) such as tendency to say Like to interest items in general, or to say Yes in checking descriptions of symptoms. Investigators differ in their judgment as to whether such variables are due to transient verbal sets or are important aspects of behavior, and, indeed, response sets seem to involve both qualities. If the investigator wishes to include elevation, whatever it measures, in determining differences between persons, he should use the full k-space data. He may be well advised, however, to use a special [illegible] of [illegible] [illegible]. A [illegible] in the [illegible] ([illegible]) [illegible] [illegible] the extent to which [illegible] [illegible] from [illegible] [illegible], is [illegible], however. [illegible] [illegible]. In fact, [illegible] [illegible]. For [illegible] [illegible] of the direction of some variables [illegible], *[illegible]* being [illegible] instead of *[illegible]* [illegible]. [illegible] [illegible] [illegible] [illegible]

[illegible] [illegible] [illegible]

[illegible]

weight for elevation, following our equation (21), since without this precaution the general factor tends to have a disproportionate influence. The investigator may instead decide to extract the common factor and study similarity in elevation separately from similarity in profile shape. If he decides to discard elevation or to study it separately, he will then go on to compute distances in k - 1 space (or, for reasons to be discussed, in k - 2 space).

Cattell and DuMas have argued that elimination of elevation is always questionable. For many studies, it is surely valueless to say that two people are similar in profile shape but not in elevation. For example, "Vocabulary higher than Digit Span" means something qualitatively different for a college graduate with IQ 120 from what it means for a ten-year old of IQ 100 or an adult of IQ 80.

The elimination of elevation, moreover, eliminates what is often the more reliable information in the score sets, and differences from test to test within a profile of deviate scores may be extremely unreliable and therefore a poor basis for investigations. This difficulty is especially to be expected when variates are highly correlated.

Our own conclusion is that:

1. Elevation should be included in the distance measure with a deliberately chosen weight if to do so makes similarity a more interpretable property.
2. Elevation should be eliminated from the distance measure only when the investigator decides that the average is saturated with a quality he desires to exclude from the domain in which similarity is measured.

It is of interest to note that Ebel (14), working on the related problem of reliability of ratings (which deals with similarity of score-sets) arrives at a similar recommendation. In his problem, the mean level of ratings assigned by each rater is comparable to our "elevation", and he lists practical considerations which make it wise at some times, and unwise at others, to consider differences in level in assessing agreement between sets of ratings.

<u>Transformation to k - 1 sphere.</u> The projection which eliminates differences in eccentricity from consideration and places all points on a k - 1 hypersphere may or may not have practical value. In studies where configurations having geometric similarity are thought of as representing similar types, $D^{\cdot}$ appears to deserve consideration. Such a problem is likely to be encountered in work with body-types or other physical measurement, where concern is literally with shape rather than with size.

Measurement of shape by projecting onto a sphere with the population centroid as center to obtain $D^{:}$ likewise has possible interest. Unlike measurement of shape by D' in the k - 1 plane, $D^{:}$ is invariant no matter which end of a dimension is taken as the positive direction. We may think of a person as having a factor specification equation, just as a test can be specified in terms of reference factors. The specification for the person tells what factors account for his deviation from the mean. Since $D^{:}$ treats as identical people who have the same factorial specification, no matter how far they deviate, it may be the appropriate measure for some type-theories. The limitations upon interpretation of $D^{:}$, however, include the serious difficulties which we discuss below in connection with k - 2 space.

<u>Considerations in using k - 2 sphere.</u> The treatments in k - 2 space will be discussed at length, because such procedures are especially common. Projection onto the k - 2 sphere treats as identical those profiles which are proportional when expressed as deviations from the person's elevation. For example, D'' would be 0, and Q would be 1, for this pair of score-sets:

3 1 0 4 (Elevation = 2; deviation profile is 1 -1 -2 2)

1 -3 -5 3 (Elevation = -1; deviation profile is 2 -2 -4 4)

Those profiles having small scatter are magnified in projection onto the sphere, (or we might say that those having large scatter are diminished proportionately). Figure 3 draws attention to some consequences.

We note that differences between persons near the center of the sphere are much magnified. The small D'_{12} becomes a large D''_{12}, but D''_{34} is little greater than D'_{34}. Points 1 and 2 represent persons with flat profiles. People who would be judged quite similar in k or k - 1 plane are sometimes reported as markedly dissimilar in the k - 2 measure.

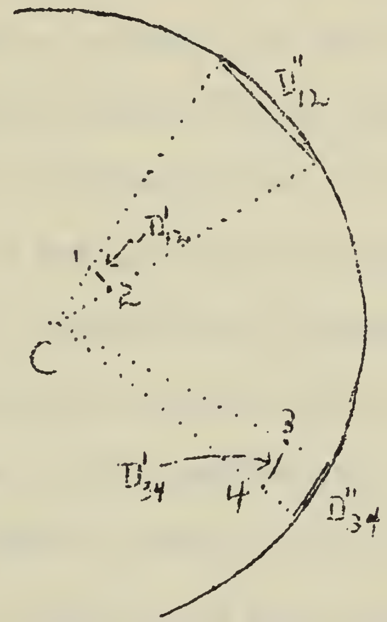
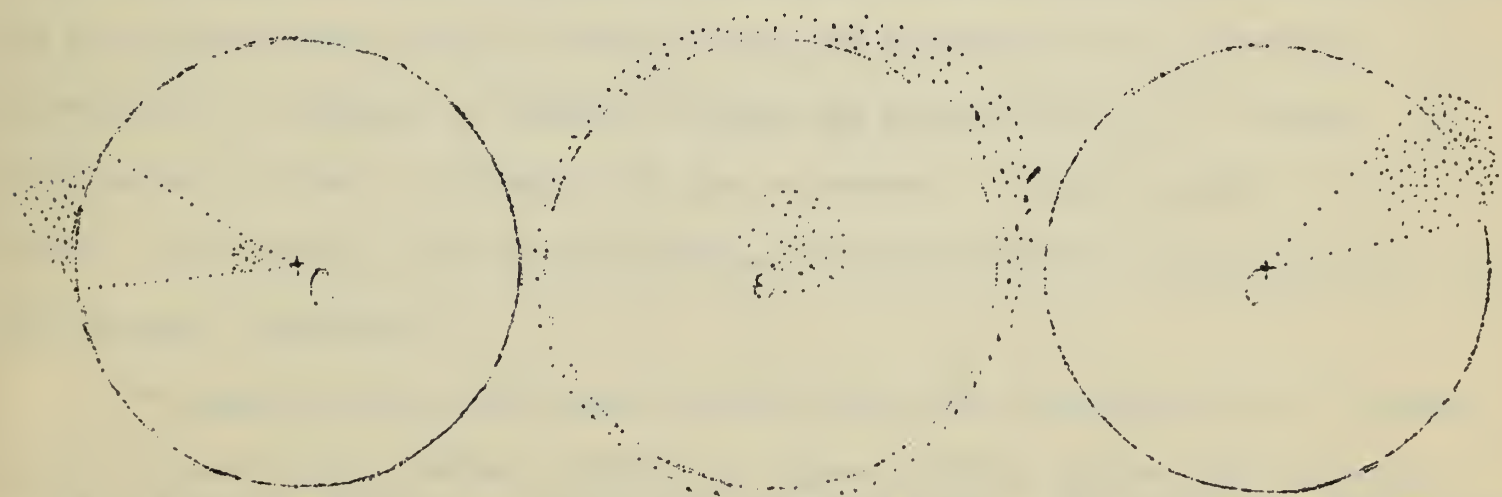


Figure 3. Magnification of distances in projection onto sphere

Figure 4 indicates the effect of the projection when error of measurement is involved. Each sketch shows a set of obtained measurements such as might be obtained



Low scatter, low error Low scatter, moderate error High scatter, moderate error

Figure 4. Effect of error and scatter on the projection onto a sphere

on repeated testing of one person, assuming that his error variance over trials is equal for all variates, and that errors are independent in k space. We show three cases; low scatter low error, low scatter moderate error, and high scatter moderate error. The second circle makes it clear that if a profile has small scatter, even a small amount of error may cause a drastic variation in the person's position in k - 2 space. A person for whom the variates are truly equal would fall at C in the k - 1 plane. On different trials he would have an equal probability of falling anywhere on the sphere, and might at different times take diametrically opposite positions. The implication is that the position of some persons in k - 2 space will be far more variable than others, and that such methods as D", Q, rho, and Tau will give unreliable similarity measures for persons with rather flat profiles. This is an expression, in other terms, of the sometimes-neglected principle that differences between two variates within a profile cannot be interpreted with confidence unless the original variates are reliable and not saturated with a common factor (25). If the conventional assumption that error of measurement is equal for all persons is approximately true for the original variates, and if flat profiles in k space can be expected, the assumption of equal error is not even approximately true for measures of people's positions in k - 2 space.

Stanley (33) has provided some data which confirm our analysis. He administered the Allport-Vernon Study of Values twice, and correlated the two profiles. This correlation is a measure of distance between the two profiles in k - 2 space. For each person, he had a correlation and also a measure of scatter within the profile; these two correlated .38 over all persons, the greater scatter being associated with the greater reliability.

The question must now be raised whether the study of profiles in k - 2 space, or more specifically, whether correlation between profiles in the usual manner,is a justifiable line of investigation. If the removal of the first factor and the magnification of error variance when scatter is equated are both disadvantageous, is a

procedure which involves both of these worth further consideration? Knowing, however, that Stephenson and others using his techniques have obtained significant results, we cannot dismiss the method until we determine why the faults we suspect have not interfered too drastically with their investigations. The explanation seems to be that there are conditions where k - 2 space data give useful and not-unduly inaccurate results.

Consider first the question of removal of the first principal component, as is done when deviate scores are obtained. This projects a distribution of points in k space into a k - 1 space, and in so doing removes the variance due to elevation. The same elimination of elevation is accomplished by the forced-sort technique. An essential condition for the resulting data to be useful is that the position in k - 1 space must be determined with substantial reliability. Under what circumstances can we expect reliability after the first component is removed? If the variates are nearly uncorrelated, each variate contributes to the total dispersion of persons approximately in proportion to V_j, and the elevation score removed constitutes one kth of the total variance. The component removed from D^2 will on the average be only one kth of the total, and D'^2 will be quite similar to D^2. Now this is what happens in the type of Q-sort Stephenson originally proposed, where variates were sampledfrom a heterogeneous collection. If a set of variates involves about fifty factors, all more or less equally weighted, removal of one factor is not expected to alter distances between persons enough to cloud results. As more correlated variates are used, extracting the elevation factor does discard more of the possibly-important variance, and the residual information will be more unreliable as a result.

The second question relates to the effect on reliability of projection from k - 1 hyperplane to k - 2 hypersphere. This projection leads to substantial magnification of error if a profile is flat. Recalling that C represents the center of the sphere, and is the point corresponding to a flat k - 1 profile, and that O' is the centroid in k - 1 space, we can expect few flat profiles if the dispersion of persons $\overline{O'P'}^2$ is much smaller than CO'^2. This demands that O' be some distance from C or

in other words, that the means for the k variates not be equal. (CO'^2 is the sum of squares of these means). The more persons fall close to C, the more will magnification of errors for them obscure results obtained with k - 2 measures. When many variates are used in the profile, as in Stephenson's forced-sort method, there is a good chance that some of the means will be unequal, and flat profiles then are less common. It is to be noted that when original scores are expressed as deviations around the group mean there will be many flat profiles; such scores are badly suited to k - 2 space procedures. In general, the essential condition is that flat profiles in k - 1 space be rare or absent.

While use of variates with unequal means will reduce the number of flat profiles, this has the disadvantage that correlations then tend to become larger and more uniform, so that one obtains less information about differences between persons. In the extreme, if items differ widely in popularity, most persons will rank them in the same order and almost all Q correlations will be 1.00.

Similarity between individuals or within a group can apparently be given no psychological interpretation unless it is measured in a domain in which at least some pairs of people are dissimilar. The similarity index obtained for any set of items depends to a major degree on the discriminating power of the items. This means that the absolute magnitude of the Q correlations cannot be directly interpreted and may have no practical significance to investigations of similarity. Only where it is demonstrated that a difference between groups or between pairs of individuals in magnitude of correlation exists is it possible to offer an interpretation. Fiedler (16), for example, asked therapists of several schools to rate statements describing a therapeutic relationship in order to determine if they differ along school lines in their concept of an ideal therapeutic relationship. The correlations between ratings were positive and large (median .64). One would be tempted to interpret such correlations as indicating a high degree of similarity among the therapists regardless of school. But it is also possible that the statements used represented such markedly desirable and undesirable qualities that high agreement could be found in almost any

In other words, [illegible]
squares of linear means). [illegible]
cation of errors for [illegible] results obtained with k = 2 [illegible]
variates are used [illegible]
would create that some of [illegible]
cannot. It is [illegible] scores and expected [illegible]
around the [illegible]
to k = 2 [illegible]
if k = 1 [illegible]

While [illegible] will [illegible] the [illegible]
[illegible], that has the [illegible] correlations [illegible]
more uniform, so that [illegible] between persons
in the extremes, [illegible] persons [illegible]
the same [illegible] correlations will be [illegible].

Similarity between [illegible]
psychological interpretation unless it [illegible]
pairs of people [illegible] are [illegible]. The similarity index [illegible]
items depends [illegible] discriminating [illegible]
that the absolute [illegible] correlations cannot be [illegible]
and have no [illegible] significance in investigations of similarity. [illegible]
demonstrated that a difference between groups of persons [illegible]
ratios
conditions of [illegible] it is possible to offer an interpretation. [illegible]
for example, found [illegible] schools in rate [illegible]
their pupils relationships [illegible] if they differ [illegible]
twenty percent of an item [illegible]. The correlations between [illegible]
[illegible] and [illegible] would be [illegible]
[illegible] a high degree of [illegible] among the [illegible]
school. But it [illegible]
[illegible]

sample of persons acquainted with therapy. Undoubtedly statements about more debatable qualities in the therapeutic relationship could be found which would result in much lower correlations among therapists of different schools. However, Fiedler goes on to extract the valuable information that the expert therapists correlate higher with one another, regardless of school, than they do with non-expert therapists of the same school. This difference supports his major conclusion, since it indicates that the choice of items had not completely pre-determined the correlations. It is disturbing to realize, however, that choice of more obviously desirable and undesirable statements might have resulted in higher correlations in both groups, so that the differences he found would have been obscured. This demonstrates that while Q correlations can be used to show the relative similarity of two pairs of persons, or persons in two groups, little meaning can be attached to the size of a Q correlation *per se*.

It is not surprising that most profile studies today utilize comparisons in k - 2 space, since the problems have been conceived in terms of correlation as used to study relationships between tests. However, it is questionable whether that model is a particularly good one. For the determination of similarity between two tests, it is reasonable to eliminate the mean and variance from consideration. As Thomson (35) and Burt have pointed out, the test mean represents its general level of difficulty for the population, while the variance is a function of the units used. Both of these values are usually quite arbitrary, depending on the choice and number of items, and since we are mainly interested in the underlying relationship between tests, these values are equated. However, in dealing with similarity of individuals, it is necessary to consider rather carefully what is involved when individuals are equated for level and scatter.

To illustrate the interpretation that can be made for measures in k or k - 1 space, which measures in k - 2 space do not allow, we refer to a study by Bendig (2). He asked professors of psychology to rank 15 professional journals in terms of their importance for study by graduate students. These ranks were correlated and factor

analyzed, leading to the conclusion that there were three bi-polar types of persons, described in terms of (a) interest in experimental approach to psychopathology, (b) interest in statistical and psychometric theory, (c) interest in theory construction in clinical area. Suppose Bendig had asked the judges to rate the journals on some objectively-defined scale ranging, for example, from "Knowledge of most contents of this journal should be required on comprehensive examinations" to "Reading this journal will not be worthwhile for any student". Then the elevation factor (tendency to give many journals high ratings) would reveal something about the judge's view of graduate training. A judge who wants students to read many journals differs from a judge who rates only a few high, even though he gives the same rank order to the journals. Moreover, the variability of the ratings by a judge would indicate his tendency to differentiate within the field of psychology, regarding some areas as worthwhile and some as trivial. A judge with a flat profile would be reporting that he is equally sympathetic to a wide range of psychological interest. A judge with a wide variation of ratings indicates a stronger differentiation. Two judges who ranked the journals in the same order, but who differed in the scatter of their ratings, would be expected to allow quite different latitude for students in training. At one point, Bendig characterizes his subjects as arranged from a "theoretical-experimental-statistical" pole to a "practical-nonexperimental-intuitive" value orientation. Possibly, rather than this typology, a k - 1 or k space measure would reveal that the judges could be better grouped in terms of specialized versus catholic values.

Combining our two conditions, it appears that measures in k - 2 space can give useful information only if the dispersion of persons in k - 1 space and also the scatter for nearly all persons are large relative to the error dispersion. Data in k - 1 space are required to determine whether these conditions are met. Then one can determine whether profiles in k - 1 space are reliable (15), and whether there are many flat profiles. Moreover, one can if he wishes eliminate the people with flat profiles from the study. The forced-sort does not collect k - 1 data and one has analyzed, leading to the conclusion that there were three bi-polar types of persons, describable in terms of (a) interest in experimental approach to psychotherapy, (b) interest in statistical and psychometric theory, (c) interest in theory and [illegible] in clinical areas. Suppose [illegible] had asked the judges to rate the journals on some objectively-defined scale ranging, for example, from "Knowledge of most contents of this journal should be required on comprehensive examinations" to "Reading this journal will not be worthwhile for any student". Then the elevation factor (tendency to give many journals high ratings) would reveal something about the judge's views of graduate training. A judge who wants students to read many journals differs from a judge who wants only a few read, even though he gives the same rank order to the journals. Moreover, the variability of the ratings by a judge would indicate his tendency to differentiate within the field of psychology, regarding some areas as worthwhile and some as trivial. A judge with a flat profile would be reporting that he is equally sympathetic to a wide range of psychological interests. A judge with a wide variation of ratings indicates a stronger differentiation. Two judges who ranked the journals in the same order, but who differed in the scatter of their ratings would be expected to allow quite different latitude for students in training. At one point, [illegible] characterizes his subjects as arranged from a "theoretical-experimental-statistical' pole to a 'practical-nonexperimental-intuitive" value orientation. Possibly, rather than this typology, a k - 1 or k space analysis would reveal that the judges could be better grouped in terms of specialized versus catholic values.

Combining our two conditions, it appears that measures in k - 2 space can give useful information only if the dispersion of persons in k - 1 space and along the elevation [illegible] for nearly all persons are large relative to the error dispersion. Data in k - 1 space are required to determine whether these conditions are met. Then one can determine whether profiles in k - 1 space are reliable (15), and whether there are many flat profiles. Moreover, one can if he wishes eliminate the people with flat no basis for testing whether profiles are reliably located. It seems quite important for those studying similarity to investigate reliability directly by obtaining two estimates for each profile. Reliability of k - 2 space measures has ordinarily not been examined in past investigations of similarity.

While we have discussed the conditions under which measures which force equal scatter on all persons can be made maximally useful, we do not recommend such procedures. Our consideration of all possibilities leads us to suggest that the method most generally advisable is the measure of equation (21) where k - 1 plane data are combined with the measure of elevation using a deliberately chosen weight (which may be zero) for elevation. (When the weight is unity, this measure is the same as D in k space). Excepting treatment of physiological and anthropometric measures, we know of no psychological or educational problem where "correcting" profiles for scatter is advantageous.

In those studies where k - 2 space measures have been used in the past, properly interpreted positive results need not be discounted. The faults to which we have drawn attention operate to obscure true relations and to make the measurement technique insensitive. This would make non-significant results, or low Q-correlations, likely in some instances where a better technique would find more relationship. We know of no biassing factor or systematic error in these procedures, however, which would have introduced significant apparent relations where none should be found.

The specialized problem of comparing a person's profile with his estimated profile introduces an interesting minor question. Several such studies are listed in a recent paper by Brown (4). The usual method is to administer (say) the Kuder Preference Record, and then to require the person to rank his interest in the categories. The profile from the test is rank-correlated with the estimated profile. But this is not precisely the question that should be asked. If one were to predict the interests of the average man, they would not all be equal; on the contrary, some categories are generally more popular. The estimated profile, obtained by the usual directions, is a k - 2 space profile based on the estimated strength of interests

to scale for [illegible] may [illegible]. It seems quite im-
portant for those [illegible] to investigate [illegible] ing the [illegible] for [illegible], particularly if [illegible] has [illegible] not been [illegible] of validity.

While we [illegible] which these might [illegible] in the [illegible]. The [illegible], that the writer [illegible] are concerned with [illegible] might yield not the usual [illegible]. ([illegible] weight [illegible] same way in K space). [illegible] of [illegible] and [illegible], we know of no psychological [illegible] where "[illegible]" [illegible] the [illegible] in circumstances.

In those [illegible] where [illegible] have been used in the past, [illegible] interpretation [illegible]. The [illegible] to which we [illegible] relations [illegible] the [illegible] technique [illegible]. [illegible] non-significant results, as the [illegible], likely [illegible] a [illegible] more [illegible]. [illegible] of [illegible] in [illegible], however, which would have [illegible] significant [illegible] should be found.

The [illegible] a [illegible] profile with [illegible] file [illegible] questions [illegible] are [illegible] in a recent paper by [illegible]. The usual method [illegible] the [illegible] and [illegible] the person to [illegible] his interest on the categories. The profile from the [illegible] with the [illegible] profile. The [illegible] is [illegible] the question [illegible] should be [illegible]. If [illegible] to [illegible] the contents of the [illegible] equally on the [illegible], then [illegible] are [illegible]. The [illegible] ([illegible]), [illegible] in the usual [illegible], is [illegible] the estimated [illegible] of [illegible] relative to each other. The test profile is a k (or k - 1) profile based on the estimated size of the deviations of the person's interests from the interests of the norm group. These two profiles should not normally be highly correlated, because the average popularity of the categories has been equalized in the test profile. To determine if people can estimate their own test profiles, the experiment can be re-designed to make the estimate more like the test in logical structure. Perhaps the easiest technique would be to ask the person to guess his percentile standing on each category. A D measure (or D_w) based on this profile would take into account elevation and scatter, and would correctly compare profiles expressed in terms of derived scores.

Short-Cut Formulas Based on Mean D^2

One use of measures of similarity is to compare any two persons. In research, however, the questions more often relate to the similarity of two groups, or the homogeneity of some particular group. If questions could be answered without computing the measure of similarity for each pair of persons involved, it would be possible to obtain the answers much more rapidly. We have discovered several formulas based on D^2 which relate to such inquiries. Unfortunately, however, they are based on the average of D squared for a set of pairs, and there seem to be no similarly helpful approaches for obtaining the average of D directly. We have indicated earlier the difficulties which make D^2 inappropriate as an interval scale to measure distance. The following formulas are presented for three purposes. They may be employed as a first rapid way of answering questions about groups, provided the investigator recognizes that different results might be obtained if mean D or median D had been determined instead of mean D^2. A second value of the formulas is that they provide insight into the nature of distance measures. Factors which increase mean D^2 will also, in general, increase mean D and median D, even though not in the same amount. The third use is for checking computations.

Average similarity within group. It was previously noted that in any group,

$$\overline{D^2_{ii'}} = 2 \frac{N}{N-1} \sum_j V_j \; , \qquad (28)$$

where i and i' vary over all persons in the sample. This index expresses the average similarity of a group, i.e., its homogeneity, except that it gives greater weight to large distances than would a linear measure of distance. This formula might be used to compare the homogeneity of one group with that of another, as in an inspection of a grouping of persons into postulated "types".

If O_Y is the centroid of the sample in the space under analysis (whether it is the center of the reference class or not), and i varies over the sample,

$$\overline{E^2_i} \text{ or } \overline{O_YP_i^{\;2}} = \frac{1}{2} \overline{D^2_{ii'}} \; . \qquad (29)$$

This is the mean second moment of persons about the centroid, and is analogous to a variance for the distribution. It is not mathematically a variance, however, since the mean E is greater than zero. We have referred to $\overline{E^2}$ as a measure of dispersion. It will be noted that if points are distributed on a hypersphere, O_Y lies within the hypersphere, and no one can fall at the centroid of the sample.

Formulas comparable to the above can easily be written for $\overline{S^2}$, and for measures in which weights are assigned to the variates.

Similarity of person to group. For a single person, it may be interesting to know his average distance from all other members in a group. If i is a member of Group Y,

$$\overline{D^2_{ii'}} = \frac{N}{N-1} (O_YP_i^{\;2} + \sum_j V_j) = \frac{N}{N-1} O_YP_i^{\;2} + \frac{1}{2} \overline{D^2_Y}$$

$$(i = 1, i' = 2, 3, \ldots N) \qquad (30)$$

Here $O_YP_i^{\;2}$ is $\sum_j (x_{ji} - \bar{x}_{j\cdot(Y)})^2$.

[illegible]

[illegible]

$$[illegible] \qquad (\ldots)$$

[illegible]

[illegible]

$$[illegible] \qquad (\ldots)$$

[illegible] the particles.

[illegible]

[illegible]

$$[illegible]$$

[illegible]

where [illegible]

$\overline{D^2_Y}$ is the average similarity within Group Y.

If i is not a member of Group Y,

$$\overline{D^2_{ii'}} = O_Y P^2_i + \frac{1}{2} \overline{D^2_Y} \qquad \text{(i not in Y; i' = 1, 2, ... N)} \qquad (31)$$

The difference between (30) and (31) is due to the inclusion of i in Y in the first case. As N increases, (30) approaches (31). As before, one must bear in mind that we have averaged the squared distances.

Distance between groups. The measure of similarity between two groups might be found in O_YO_Z, the separation of their centroids. This is the measure most used in comparison of groups to test the null hypothesis, but we sometimes desire to determine instead the average D^2 between members of the two groups. It permits us to ask whether a group resembles another group as closely as members within the group resemble each other. For this we have

$$\overline{D^2_{ii'}} = \overline{O_Y P^2_i} + \overline{O_Z P^2_{i'}} + O_Y O_Z{}^2 \qquad (i = 1, 2, \ldots N_Y;\ i' = 1, 2, \ldots N_Z) \qquad (32)$$

Here we see the average cross-similarity as made up of three components: squared distance between group means, dispersion within the first group, and dispersion within the second group.

The formula can be rewritten as follows, if $\sigma^2_{j(Y)}$ is the variance of j for the population Y represents, etc.:

$$\overline{D^2_{ii'}} = \sum_j \left(\sigma^2_{j(Y)} + \sigma^2_{j(Z)} + (\bar{x}_{j.(Y)} - \bar{x}_{j.(Z)})^2\right) . \qquad (33)$$

There is one term for the variance within each group, and one which is twice the variance between groups.

Summary and Recommendations

Studies attempting to determine the similarity of persons have used a variety of statistical procedures. Some of these procedures are more advantageous than others, and we have attempted to analyze each procedure so that investigators can choose the method most likely to reveal the effects they seek to measure.

Each procedure for determining the similarity of two score-sets in effect determines the distance between two points in a space defined by the variates. The decisions facing the investigator, which determine what results he will obtain, are: choice of variates, choice of metric for each variate, assignment of weights to variates, and choice of index of similarity. The choices made define the domain within which similarity is to be determined.

For profile comparison it is necessary to express each variate in a scale of equal units, such that units on the several scales are comparable. Unless the investigator is satisfied to assume that his units do possess these qualities, he should transform his variates or assign differential weights to them in order to get units he regards as comparable.

An index may be based on either an orthogonal or an oblique model, the latter taking into account the correlation among variates. All indices of similarity in general use in psychological studies are based on the orthogonal model. We propose an index D of this type, and a standardized index S which has similar properties. An oblique model treated by Mahalanobis leads to the index $\mathbb{D}$, which is especially suited to classification problems where groups are defined a priori. It is closely related to the Hotelling T and the discriminant function. Our comparison shows that D gives more weight to common factors among the variates than $\mathbb{D}$. As a consequence, D is more reliable from trial to trial, and more stable from one sample of variates to another. If variates have little correlations, D approaches $\mathbb{D}$. In general, for descriptive purposes, the index D based on the orthogonal model seems superior to $\mathbb{D}$ because of its greater stability. D, however, when applied to correlated variates, has certain distorting properties which cause factors to have greater weight in some pairs of persons than in others, and its interpretation is unclear.

If it is desirable to take correlation into account, the practical procedure is to transform the correlated variates to an uncorrelated set, and apply the orthogonal model. If possible, it is wise to begin with a set of nearly uncorrelated variates, each reliably measured, or with a set of variates having only one general

Each procedure for determining the similarity of two score-profiles in effect determines the distance between two points in a space defined by the variates. The decisions facing the investigator, which determine what results he will obtain, are: choice of variates, choice of metric for each variate, assignment of weights to variates, and choice of index of similarity. The former ones define the domain within which similarity is to be determined.

For profile comparison it is necessary to express each variate in a scale of equal units, such that units on the several scales are comparable. Unless the investigator is satisfied to assume that his units do possess these qualities, he should transform his variates or assign differential weights to them in order to get units he regards as comparable.

An index may be based on either an orthogonal or an oblique model, the latter taking into account the correlation among variates. All indices of similarity in general use in psychological studies are based on the orthogonal model. We propose an index D of this type, and a standardized index S which has similar properties. An oblique model treated by Mahalanobis leads to the index Δ, which is especially suited to classification problems where groups are defined a priori. It is closely related to the Hotelling T and its discriminant function. Our comparison proves that Δ gives more weight to common factors among the variates than D. As a consequence, Δ is more reliable from trial to trial, and more stable from one sample of variates to another. If variates have little correlation, Δ approaches D. In general, for descriptive purposes, the index D based on the orthogonal model seems superior to Δ because of its greater stability. Δ, however, when applied to correlated variates, has certain distorting properties which cause factors to have greater weight in comparison of persons than in theory, and its interpretation is unclear.

If it is desirable to take correlations into account, the practical procedure is to transform the correlated variates to an uncorrelated set, and apply the orthogonal model. If possible, it is wise to begin with a set of nearly uncorrelated variables, each reliably measured, or with a set of variates having only one central ("elevation") factor.

The various orthogonal indices may be classified as follows:

- k — space measures, which reflect differences in profile shape, elevation, and scatter. These include the measure D or S which we describe, Cattell's r_p, and one form of Pearson's CRL.
- k - 1 — hyperplane measures, which remove differences in elevation from the data before comparison of profiles. The index D' is used for such data. A special index D_w is suggested which permits the investigator to reintroduce the elevation factor with any desired weight.
- k - 1 — hypersphere measures, which remove differences in eccentricity of profiles. Measures in this group are chiefly of theoretical interest.
- k - 2 — hypersphere measures, which remove differences in elevation and scatter from the profile. These include product-moment correlation, rho, Tau, r_{ps}, and correlation based on Stephenson's forced-sort procedure.

The investigator should eliminate elevation and scatter from his distance measure only if there is a psychological reason for regarding differences in these as unimportant. For most purposes, we regard the index D_w as best suited to similarity studies. When w is one, this becomes D. If D or D_w is used, the investigator treats as alike those people who have the same profile, but considers that profiles having different elevation or different scatter are as truly different as profiles having different high or low points. In contrast, measures in k - 1 space (based on deviations around the person's mean) and measures in k - 2 space (with scores standardized in each profile), discard some of the most reliable information in the score set. Profiles in k - 1 space are less reliably determined than k-space profiles. In going to k - 2 space, error is greatly magnified for persons with small scatter. Such magnified errors are likely to obscure true relationships.

Most investigations have been based on k - 2 space measures. We do not believe that such indices are generally the best for research on similarity. It is true that some studies have successfully discovered relationships with these measures. Measures in k - 2 space can be dependable when variates are reliably

measured, and where there are few flat profiles. Even in studies where k - 2 measures have been useful, a more powerful technique would be expected to produce the results with greater clarity. In studies which yielded no significant relations involving k - 2 similarity measures, an index such as D_W might have found relationships of importance.

In choosing between D, D', and D_W, the investigator must decide whether there is an interpretable elevation factor, and whether this factor should be allowed to influence his distance measure. If the variates do not have substantial positive intercorrelation, we recommend that D, computed on the original measures in k space, be used to determine dissimilarity. If the variates do generally measure a common factor, the investigator should consider the meaning of this factor and decide whether it is one he wishes to count. If he wishes to eliminate it from consideration because he regards it as irrelevant to his problem, he will use D' as his index. If he wishes to include the factor, he may choose an appropriate weight for it and use D_W. The advantage of D_W over D is that with substantially intercorrelated variates the elevation factor may receive greater weight in D than it should, relative to the weight given to the shape of the profile.

The distance index may be expressed in terms of D, D^2, or some transformation to another scale. It appears unwise to force D into a correlation-like index ranging from +1 to -1 as Cattell suggests. There is probably no limit on how dissimilar two people can be, save as one is imposed by the method of gathering data. Hence in k space or k - 1 hyperplane D can range from 0 (perfect similarity) to ∞. If similarity is reported as D^2, we have useful formulas for mean D^2 under various situations. D^2, however, seems to be less meaningful than D as a measure of distance, expecially as D is literally the distance between points in our geometric model. It is also more likely to have statistical properties which make it possible to utilize means, variances, and product-moment correlations. Thus we advise that D^2 be used only in preliminary studies where its simplicity is of value and where ordering of

[illegible]

[illegible] the shape of the crystal.

[illegible]

similarity is the major question. The use of Q as a measure should also be limited to these conditions.

This paper has given little attention to problems of reliability, but it is clear that measures of distance between points cannot be determined dependably if the locations of the points are undependable. Therefore, any steps the investigator takes to make his profiles more reliable are well worth while.

Profile research is necessarily faced with severe difficulties. The results of any investigation are influenced by numerous choices which must be made in part arbitrarily. Even when these decisions are made wisely, the difficulty of making reliable measurements on many variates at once is a severe one. We hope that in spite of these problems, the adoption of techniques of analysis which include as much information as the data permit, and which do not introduce additional errors of their own, will permit studies of similarity to advance psychological knowledge.

REFERENCES

1. Barnette, W. L. Occupational aptitude patterns of selected groups of counseled veterans. Psychol. Monogr., 1951, 65, No. 322.

2. Bendig, A. W. A Q-technique study of the professional interests of psychologists. J. Psychol., 1952, 33, 57-64.

3. Block, J., Levine, L., and McNemar, Q. Testing for the existence of psychometric patterns. J. abnorm. soc. Psychol., 1951, 46, 356-359.

4. Brown, M. N. Expressed and inventoried interests of veterans. J. appl. Psychol., 1951, 35, 401-402.

5. Burt, C. L. Correlations between persons. Brit. J. Psychol., 1937, 28, 59-96.

6. Burt, C. L. The factors of the mind. London: University of London Press, 1940.

7. Cattell, R. B. r_p and other coefficients of pattern similarity. Psychometrika, 1949, 14, 279-298.

8. Chowdhry, K., and Newcomb, T. M. The relative abilities of leaders and non-leaders to estimate opinions of their own groups. J. abnorm. soc. Psychol., 1952, 47, 51-57.

9. Cronbach, L. J. "Pattern tabulation": a statistical method for analysis of limited patterns of scores with particular reference to the Rorschach test. Educ. psychol. Measmt., 1949, 9, 149-172.

10. Cronbach, L. J. Further evidence on response sets and test design. Educ. psychol. Measmt., 1950, 10, 3-31.

11. Cronbach, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 297-334.

12. DuMas, F. M. A quick method for analyzing the similarity of profiles. J. clin. Psychol., 1946, 2, 80-83.

13. DuMas, F. M. On the interpretation of personality profiles. J. clin. Psychol., 1947, 3, 57-65.

14. Ebel, R. L. Estimation of the reliability of ratings. Psychometrika, 1951, 16, 407-424.

15. Edgerton, H. A., Bordin, E., and Molish, H. Some statistical aspects of profile records. J. educ. Psychol., 1941, 32, 185-196.

16. Fiedler, F. E. The concept of an ideal therapeutic relationship. J. consult. Psychol., 1950, 14, 239-245.

17. Fiedler, F. E., Blaisdell, F. J., and Warrington, W. G. The relationship between unconscious attitudes and sociometric choice. J. abnorm. soc. Psychol., in press.

[illegible]

1. [illegible]

2. [illegible]

3. [illegible]

4. [illegible]

5. [illegible]

6. [illegible]

7. [illegible]

8. [illegible]

9. [illegible]

10. [illegible]

11. [illegible]

12. [illegible]

13. [illegible]

14. [illegible]

15. [illegible]

16. [illegible]

17. [illegible]

18. Fiedler, F. E., and Senior, Kate. An exploratory study of unconscious feeling reactions in fifteen patient-therapist pairs. J. abnorm. soc. Psychol., in press.

19. Gleser, Goldine C., Loevinger, Jane, and DuBois, P. H. Resolution of a pool of items into relatively homogeneous subtests. Amer. Psychologist, 1951, 6, 401.

20. Hodges, J. L. Jr. Discriminatory Analysis: 1. Survey of discriminatory analysis. USAF School of Aviation Medicine. Randolph Field, Texas. 1950.

21. Holzinger, K. J. Factoring test scores and implications for the method of averages. Psychometrika, 1944, 9, 257-262.

22. Kelly, E. L., and Fiske, D. W. The prediction of performance in clinical psychology. Ann Arbor, University of Michigan Press, 1951.

23. Kendall, M. G. Rank correlation methods. London. Charles Griffin and Co., Ltd., 1948.

24. Moses, L. E. Non-parametric statistics for psychological research. Psychol. Bull. 1952, 49, 122-143.

25. Mosier, C. I. Batteries and Profiles in Lindquist, E. F. (ed.) Educational measurement. Washington Amer. Council on Educ., 1951, pp 764-808.

26. Osgood, C. E. and Suci, G. A difference method for analyzing the structure of intercorrelations among variables which is applicable to raw-score matrices. Psychol. Bull. in press.

27. Osgood, C. E. and Suci, G. The nature and measurement of meaning. Psychol. Bull, in press.

28. Pearson, K. On the coefficient of racial likeness. Biometrika, 1928, 18, 105-117.

29. Rabin, A. I., and Guertin, W. H. Research with the Wechsler-Bellevue test: 1945-1950. Psychol. Bull., 1951, 48, 211-248.

30. Rao, C. R. The utilization of multiple measurements in problems of biological classification. J. roy. stat. Soc., Sec. B, 1948, 10, 159-203.

31. Rao, C. R. Tests of significance in multivariate analysis. Biom. 1948, 35, 58-79.

32. Rao, C. R., and Slater, P. Multivariate analysis applied to differences between neurotic groups. Brit. J. Psychol., Stat. Sec., 1949, 2, 17-29.

33. Stanley, J. C. Insight into one's own values. J. educ. Psychol., 1951, 42, 399-408.

34. Stephenson, W. A statistical approach to typology; the study of trait-universes J. clin. Psychol., 1950, 6, 26-38.

35. Thomson, G. The factorial analysis of human ability, Fourth Edition, London: University of London Press, Ltd., 1950.