ANALYSIS OF INDIVIDUAL DIFFERENCES IN MULTIDIMENSIONAL SCALING VIA AN N-WAY GENERALIZATION OF "ECKART-YOUNG" DECOMPOSITION

J. Douglas Carroll and Jih-Jie Chang

BELL TELEPHONE LABORATORIES

MURRAY HILL, NEW JERSEY

An individual differences model for multidimensional scaling is outlined in which individuals are assumed differentially to weight the several dimensions of a common "psychological space". A corresponding method of analyzing similarities data is proposed, involving a generalization of "Eckart-Young analysis" to decomposition of three-way (or higher-way) tables. In the present case this decomposition is applied to a derived three-way table of scalar products between stimuli for individuals. This analysis yields a stimulus by dimensions coordinate matrix and a subjects by dimensions matrix of weights. This method is illustrated with data on auditory stimuli and on perception of nations.

There has been an interest for some time in the question of dealing with individual differences among subjects in making similarity judgments on which a multidimensional scaling of stimuli is to be based. Kruskal [1968] and McGee [1968] have both incorporated different ways of dealing with individual differences into their scaling procedures. Tucker and Messick [1963] proposed an approach, which they called "Points of view analysis," which is probably the most widely used method for dealing with such individual differences. In this method, intercorrelations are first computed between subjects (based on their similarity judgments) and the resulting correlation matrix is factor analyzed to produce a subject space. One then looks for clusters of subjects in this subject space, and if such clusters are found, proceeds in one way or another to define "idealized" subjects corresponding to clusters. (The "idealized subject" for a given cluster may be defined, for example, by finding the pattern of similarity judgments corresponding to a hypothetical subject at the cluster centroid, by choosing the actual subject closest to that centroid, or, most simply, by averaging the similarity judgments for subjects in the given cluster.) The similarities for these "idealized subjects" are then, individually and independently, subjected to multidimensional scaling.

This approach has been criticized by a number of people, most recently by Ross [1966] (see Cliff, 1968, for a reply to Ross's criticism and a further discussion of the "idealized individuals" interpretation of "Points of view

analysis"). Perhaps the most cogent criticism is that the method is little more powerful than doing separate scalings on the individual subjects—and it makes no explicit assumptions about possible or probable communality of the dimensional structures for different real or idealized individuals. It would be very surprising if the various configurations had no structure in common. Rather, one might suspect that, for example, one or two dimensions are the same in two different configurations while a third is different, or that the same dimensions are present, but they have different relative *saliences*, or importances, for different people. Both of these possibilities, as well as the extreme case in which the configurations have nothing at all in common, are included in our model which, we might say, represents a different point of view about "Points of view."

The Model

We assume a set of r dimensions or "factors" underlying the n stimuli. These are assumed to be common to all individuals. We shall use x_{jt} to represent the value of the jth stimulus on the tth dimension (so j ranges from 1 to n and t from 1 to r). We assume the similarity judgments for each subject to be related in a simple way to a kind of modified Euclidean distance in this space. In particular, for now, we shall assume that

$$(1) \qquad\qquad s_{jk}^{(i)} = L(d_{jk}^{(i)})$$

where $s_{jk}^{(i)}$ is the similarity of the jth and kth stimuli for the ith individual $(i = 1, 2 \cdots m)$, and L is a linear function (with negative slope). The "modified" Euclidean distance for the ith subject is given by:

$$(2) \qquad\qquad d_{jk}^{(i)} = \sqrt{\sum_{t=1}^{r} w_{it}(x_{jt} - x_{kt})^2}$$

This formula differs from the usual Euclidean distance formula only in the presence of the weights w_{it} , which represent the saliences or importances mentioned above. Another way of looking at this formula is to say that the $d_{jk}^{(i)}$'s are ordinary Euclidean distances computed in a space whose coordinates are:

$$(3) \qquad\qquad y_{jt}^{(i)} = w_{it}^{1/2} x_{jt}$$

that is, in a space that is like the x space except that the configuration has has been expanded or contracted (differentially) in directions corresponding to the coordinate axes. This is the kind of transformation that would, for example, convert circles into ellipses with major and minor axes parallel to the coordinate axes, or spheres into (parallel) ellipsoids in three dimensions.

This model is sufficiently general to include all the models discussed above. Two completely different spaces could be accommodated, for example, by assuming a "common" space combining all the dimensions of the two separate spaces (the direct sum, in technical terms). The dimensionality

of this "super-space" would be the sum of the two dimensionalities. Then, by assuming that one group of subjects attaches zero weights or saliences to the dimensions of the first space, while a second group attaches zero weights to those of the second, this model becomes equivalent to the "two-separate spaces" model. Of course, the more interesting and exciting case is that in which some, but not complete, communality exists. The model is by no means completely general, however. In particular it does not allow differential rotation as well as differential weighting of dimensions. Possible methods of analysis in terms of such a more general model will be discussed in a later section. Suffice it to say that, in our viewpoint, it is actually *easier*, in principle, to develop a method of analysis in terms of the more general model, but that we feel this model to be of less interest than the present, more restricted one. The fact that the present model allows stretching and shrinking transformations only along the fixed coordinate axes means that these axes have a much stronger meaning than is the case in the usual scaling situation involving Euclidean metric. This means that the method to be proposed here should, and *will*, determine these axes uniquely. If the model we postulate is true, then these uniquely determined coordinate axes will correspond to meaningful psychological dimensions in a very strong sense (as postulated, they correspond to fundamental sensory, perceptual or judgmental processes that vary in salience, or strength of effect on perception, across individuals). This argument has been developed in more detail by Horan [1969].

The model may not hold in every case, but if it does we gain a unique and hopefully psychologically meaningful orientation of axes, thus obviating the rotational problem and defining much stronger scales of measurement than is usual in multidimensional scaling. One example will be presented in a later section to support the argument that this in fact does happen. Many more cases have now been collected [Wish, 1970a, b; Wish, 1971; Wish, Deutsch and Biener, 1971; Carroll, 1971; Carroll and Chang, 1970; Carroll and Wish, 1970; Bricker and Pruzansky, 1970; Rao, 1970; Green and Rao, 1970; Green and Carmone, 1971] that lend credence to this notion. In essentially every case the dimensions have proved to be interpretable directly as they are derived from this analysis (*i.e.*, without rotation). In cases where a set of *a priori* physical or theoretical dimensions were known, the recovered (unrotated) dimensions have always (to date) corresponded to them in an essentially one to one fashion. We therefore argue that it is appropriate to analyze data in terms of this very strong and specific model, and that only if this model fails to fit the data adequately should one have recourse to a more general model.

The Method of Analysis

The first step in the method of analysis is to convert the similarities into distance estimates. Under the linear assumptions we have made, this

can be done by using one of the standard procedures described in Torgerson [1958]. We then use the equations also described in Torgerson [1958, pp. 254–259] to convert the distance estimates for each subject into scalar products of vectors (to get the matrix of scalar products, we simply double center the matrix whose entries are $-1/2\ d_{jk}^2$). This gives us numbers $b_{jk}^{(i)}$, which, in the present case, can be regarded as scalar products between the vectors $y_{jt}^{(i)}$; *i.e.* (ignoring error terms):

$$(4) \qquad\qquad b_{jk}^{(i)} = \sum_{t=1}^{r} y_{jt}^{(i)} y_{kt}^{(i)} = \sum_{t=1}^{r} w_{it} x_{jt} x_{kt} .$$

The last expression results from substituting the expression in (3) for $y_{jt}^{(i)}$.

We solve for (least squares) estimates of these parameters by a procedure we have called "canonical decomposition of N-way tables". This method will be described in general terms at a later point. For now let us see how the method is applied to the specific problem of solving for parameters of the model in equation (4). Let us, first, replace (4) with (4-a) below:

$$(4\text{-}a) \qquad\qquad z_{ijk} \cong \sum_{t=1}^{r} w_{it} x_{jt}^{(L)} x_{kt}^{(R)}$$

where "$\cong$" implies, in the present context, that we seek a least squares solution for the parameters on the right.

We have, simply, made the substitution $z_{ijk} \equiv b_{jk}^{(i)}$, and put superscripts (L) and (R) on the x's (to distinguish the x on the left from the one on the right). We will let W, X_L and X_R represent the corresponding $(m \times r)$, $(n \times r)$ and $(n \times r)$ matrices. Suppose, now, that we are given initial estimates of X_L and X_R (these may, and generally would be, exactly the same), and want to derive a least squares estimate for W.

Letting $s = n(j - 1) + k$, so that s varies from 1 to n^2, define $g_{st} \equiv x_{jt}^{(L)} \cdot x_{kt}^{(R)}$, and $z_{is}^* \equiv z_{ijk}$. Then (4-a) can be written in the equivalent form:

$$(5) \qquad\qquad z_{is}^* \cong \sum_{t=1}^{r} w_{it} g_{st}$$

Given the equation in this form, it is immediately apparent that a least squares solution is available for the w's (holding the x's, and thus the g's, fixed). To see this more clearly, write (5) in the matrix form:

$$(6) \qquad\qquad Z^* \cong W G^T$$

where Z^* is the $m \times n^2$ matrix with entries z_{is}^* , and G is the $n^2 \times r$ matrix with entries g_{st} . (The columns of G can also be defined as the Kronecker product of corresponding column vectors of X_L and X_R .) The least squares solution for W is, by well known results:

$$(7) \qquad\qquad \hat{W} = Z^* G(G^T G)^{-1}$$

(that is, we define $\hat{W}$ by postmultiplying Z^* by the right pseudoinverse of G^T).

Now, having solved for W we may get a better estimate of X_L (say) by similar means. In this case, let $u \equiv m(i - 1) + k$, (so that u varies from 1 to mn) and define $h_{ut} \equiv w_{it} \cdot x_{kt}^{(R)}$ and $z_{ju}^{**} \equiv z_{ijk}$. Now we may rewrite equation (4-a) as

$$(8) \qquad\qquad z_{ju}^{**} \cong \sum_{t=1}^{r} x_{jt}^{(L)} h_{ut}$$

or, in matrix form as

$$(9) \qquad\qquad Z^{**} \cong X_L H^T$$

from which the least squares estimate for X_L is

$$(10) \qquad\qquad \hat{X}_L = Z^{**} H (H^T H)^{-1}$$

Given the new values for both W and X_L we may, by a similar procedure, solve for better estimates of X_R . We then return to solve again for a still better estimate of W, then of X_L and of X_R , continuing this procedure iteratively until the process converges. Note that, at each step of this iterative procedure we are reducing (or, certainly, not increasing) the total error sum of squares. At each stage we are seeking out the values of one of the three sets of parameters yielding a precise minimum sum of squares for the given (fixed) values of the other two sets. This can be regarded as a generalization of standard "relaxation" procedures for optimizing a function of many variables, in which the optimum is sought out with respect to one variable, holding all the other fixed, then with a second variable, holding the first and the remaining variables fixed, and so on, iteratively, until convergence is achieved. The difference in the present procedure is that whole subsets of variables are treated in this way. It is also important to note that we can go to the precise minimum for the given subset at each stage in the process.

The reader may also note that, during the iterative process there is no constraint making $X_L = X_R$. In fact these two matrices may be quite different in early stages of the process. However, the basic symmetry of the data (the $b_{jk}^{(i)}$'s or z_{ijk}'s) reflected in the fact that $z_{ijk} = z_{ikj}$ for all i, j and k, guarantees that, when the process finally converges, X_L and X_R will be, if not precisely equal, equivalent in a slightly more general sense. That is to say, they will be related by a diagonal transformation, as follows:

$$(11) \qquad\qquad \begin{cases} X_L = C X_R \\ X_R = C^{-1} X_L \end{cases}$$

where C is an $r \times r$ diagonal matrix with nonzero entries on the diagonal. From inspection of (4-a) it should be evident that any two of the three matrices, W, X_L and X_R can be premultiplied by arbitrary non-singular diagonal

matrices, and the third multiplied by the product of the inverses of these two diagonal matrices, without changing the right side of that equation. (In non-matrix notation, we can multiply, say, $x_{it}^{(L)}$ by $c_t^{(L)}$ and $x_{kt}^{(R)}$ by $c_t^{(R)}$, and compensate by multiplying w_{it} by $1/c_t^{(L)} \cdot c_t^{(R)}$). This means that we can expect our solution to be defined only up to this class of transformations. In particular it means we can easily make X_L and X_R *precisely* equal by an appropriate choice of diagonal matrices. (In practice, we actually, as a final step, simply set X_L equal to X_R and recompute W. This assures that any discrepancies between X_L and X_R due to lack of perfect convergence will be corrected. X_L is set equal to X_R rather than the converse because X_R is the last computed, and should, therefore, be closer to the optimal value). These arbitrary diagonal transformations will be involved also in the next section on normalization of the final solution.

It might be supposed, at this point, that we could have easily constrained the two matrices to be the same at *every* stage of the algorithm by recomputing only one at each iterative cycle rather than both, and simply forcing the other to take on the same set of values. While such an algorithm may work, and might even prove to be more efficient than the present one in some cases, it has been rejected at present because its properties are not so well understood. In particular, while it is easy to see that the method we have proposed will converge at least to a local minimum (and, we believe, to the global minimum in "almost all" cases), no such statement can be made at present for this alternative algorithm.

Normalization

There are two different normalization questions that should be considered. The first relates to normalization of the initial *data*, and is basically a question of relative *weighting* of the data for different subjects; that is, with how much each subject should influence the analysis. The second question has to do with normalization of the final solution. This does not concern the actual solution obtained, but only the most appropriate or informative way to present that solution.

The first problem, that of initial data normalization, reduces, in the present case, to the relative scaling of the scalar products matrix for each subject. The sum of squares of the scalar products provides an appropriate measure of the relative scaling of these matrices. Since the mean of the scalar products is necessarily zero for each subject (by virtue of the way these are defined) the sums of squares will be *proportional* to the variances. The most natural normalization procedure, in our view, is to *equate* these variances. Accordingly, we have normalized the data to be presented below by scaling each matrix so that its sum of squares equals one. In an earlier analysis of the same data, in which these variances were not equated, one subject so dominated the analysis that one of the dimensions was evidently determined

on the basis of that subject's data alone. This normalization not only equalizes the subjects influence on the analysis, but also simplifies interpretation of the subject space. This will be seen more clearly subsequently. We recommend this normalization procedure in general, unless there is a compelling reason for weighting subjects differentially.

The need for a normalization of the final solution becomes evident upon inspection of (2) and (4). In both of these equations it can be seen that it is possible to multiply the x_i's by any arbitrary (nonzero) constant, and compensate by multiplying the w_i's by the *reciprocal* of the *square* of that constant. This scaling problem can be seen in a more general form in (4-a), as discussed above, and will later be seen to apply more generally to the models defined in (36) and (45). In the present application to "individual differences" scaling analysis, this means that we may scale the dimensions of the stimulus space in an arbitrary way, and compensate by introducing the appropriate scaling of the subject space.

While a number of options are open, the most natural procedure, in our view, is to normalize the stimulus space so that the variances of projections of stimuli on the several coordinate axes are equal (to one, in the present case). When the appropriate companion normalization is applied to the weight space, this leads to certain interpretive niceties. It means that the square of the Euclidean distance of a subject's point from the origin can be interpreted (approximately) as total variance accounted for in the *scalar products* data for that subject. We say "approximately" because the exact "variance accounted for" will depend, as well, on the correlations among stimulus dimensions. If these dimensions are orthogonal, in the sense that their correlations are zero, the square of the Euclidean distance (of the subject point from the origin) will provide a direct measure of variance accounted for. If the initial data are normalized in the way discussed earlier, a subject's point will be of unit distance from the origin only if all the variance for that subject is accounted for. In case the dimensions are correlated, however, the converse is not true. That is to say, the distances of a subject's point from the origin may be less than one, but all his variance may nonetheless be accounted for.

Analysis of Illustrative Sets of Data

P. D. Bricker and S. Pruzansky (see Bricker and Pruzansky, 1970) have kindly made available data collected at Bell Telephone Laboratories on 24 auditory tones. The stimuli and data were very similar to a set described earlier by Bricker, Pruzansky, and McDermott, 1968. The tones were generated by varying three physical properties, modulation waveform (sine wave vs. square wave), modulation percentage, and modulation frequency, in a 2 × 3 × 4 factorial design. Dissimilarity data were obtained from 20 subjects, who were asked to rate the degree of difference of each pair on a

scale ranging from 0 (for a maximally similar, or indistinguishable pair) to 9 (for a maximally dissimilar pair). The dissimilarities were assumed linear with distances. To convert them into (ratio scale) distance estimates additive constants were estimated (separately for each subject) by the "one dimensional subspace" method described in Torgerson [1958]. Scalar products matrices were computed from these estimated distances, normalized to unit sum of squares, and subjected to the analysis described above. This analysis accounted for about 59.8% of the variance in the individuals' scalar products matrices. This seemed reasonable, in view of the fact that these were individual data, and thus susceptible to large error variances.

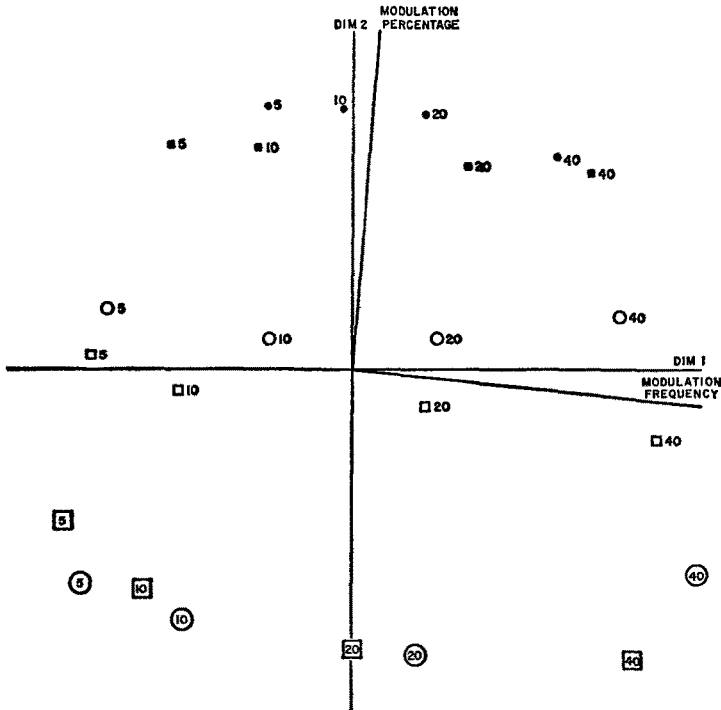The "group" stimulus space obtained via this three-way analysis is

FIGURE 1

The one-two plane of the group stimulus space for the 24 tones (data due to Bricker, *et al.*). Circles stand for sinusoidal modulation waveform, and squares for square or rectangular modulation waveform. Small closed figures indicate 3% modulation percentage, intermediate open figures indicate 10% modulation percentage, and the largest figures (with numbers inside) indicate 25% modulation percentage. The number either beside or inside the figure encodes modulation frequency (in hertz). Projections into this one-two plane of the straight lines optimally corresponding to modulation frequency and modulation percentage are also shown.
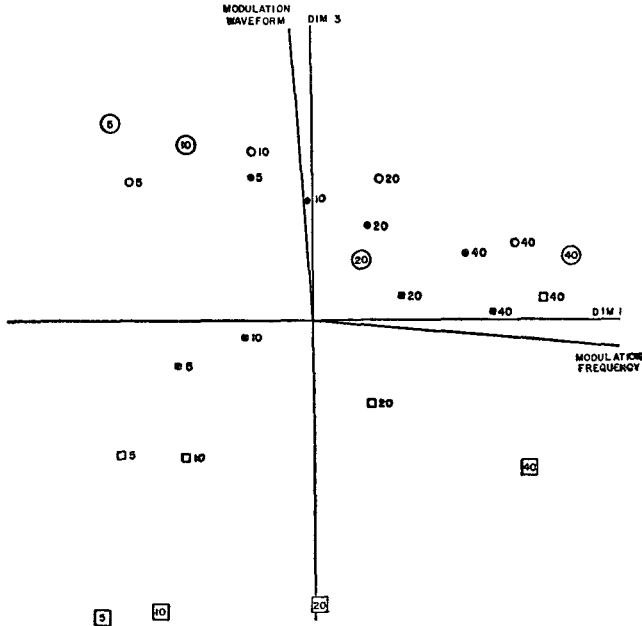
FIGURE 2

The one-three plane of the group stimulus space for the 24 tones (from Bricker, *et al.*).
Projections of straight lines optimally corresponding to modulation percentage and modulation waveform are also shown.

shown in Figures 1, 2 and 3, showing the planes defined by dimensions one and two, one and three, and dimensions two and three, respectively. In these figures circles represent the sinusoidal waveforms while the squares represent the square (or rectangular) waveforms. Small filled figures correspond to the smallest modulation percentage (3%), intermediate open figures to the intermediate modulation percentage (10%) and the largest figures (with numbers inside) to the highest modulation percentage (25%). Finally, the number either beside or within the figure representing a stimulus encodes its modulation frequency (in cycles per second, or "hertz").

The coordinate axes of the stimulus space in this analysis are not arbitrary, as is usually the case in multidimensional scaling involving the Euclidean metric. This is true because of the differential weights or saliences for different individuals. The class of permissible transformations of the stimulus space induced by differential stretching or shrinking of axes is *not* invariant under orthogonal rotation. For this reason it makes sense to interpret this space in terms of the particular axes obtained in this analysis. Since, in the present case, an *a priori* set of "fundamental" physical dimensions exists, one might expect these axes to correspond in a one-to-one manner to the known
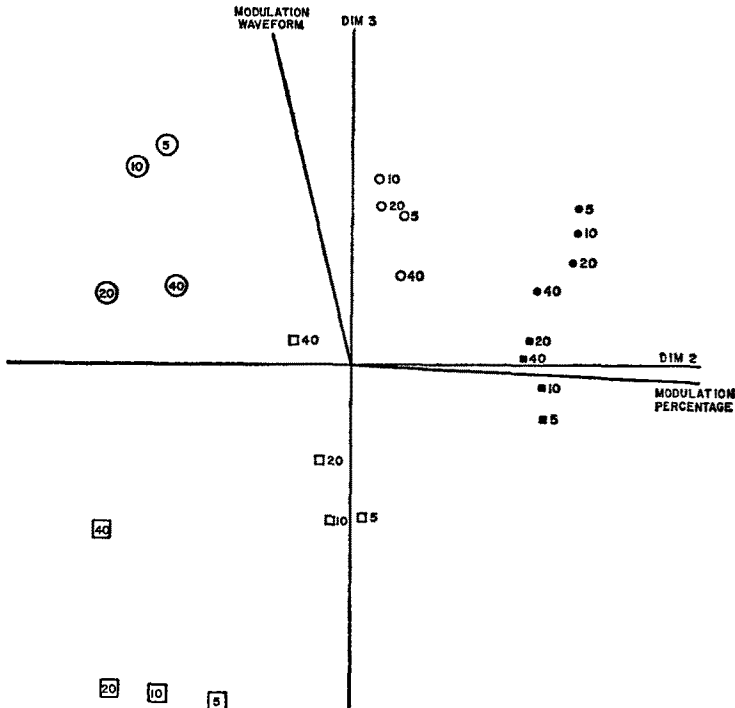
FIGURE 3

The two-three plane of the group stimulus space for the 24 tones (from Bricker, *et al.*). Projections of straight lines optimally corresponding to modulation percentage and modulation waveform are also shown.

physical dimensions. To test this notion we used a non-linear regression technique to locate directions in the stimulus space optimally corresponding to the three physical dimensions. As postulated, these directions seemed to correspond very directly to the coordinate axes as determined by the individual differences analysis. Dimension one corresponded very closely to modulation frequency, dimension two to modulation percentage and dimension three to modulation waveform (it should be noted here that the ordering of the dimensions corresponds approximately to relative "variance accounted for"). The cosine of the angle between the line corresponding to a physical property and the axis most nearly corresponding to that property provides an appropriate measure of the extent of this correspondence. These cosines were .990, .995 and .968 for modulation frequency, percentage and waveform respectively. The correlation ratio, η^2, was computed as a measure of correspondence between projections of stimulus points on these best fitted lines and the physical properties. These η^2's were .945, .970, and .678 respectively. Since

the nonlinear regression procedure used in the present case is equivalent to maximizing η^2, which is in turn equivalent to a linear discriminant analysis, we may use standard methods to test the significance of these η^2's. Applying the appropriate test from the theory of linear discriminant functions reveals that all of these fits are significant beyond the .001 level. A procedure based on linear correlation was also used to locate property axes, and the resulting directions were virtually indistinguishable from those based on nonlinear correlation. The resulting linear correlations were .956, .969 and .824 respectively, all significant beyond the .001 level, by the standard test for multiple correlations.
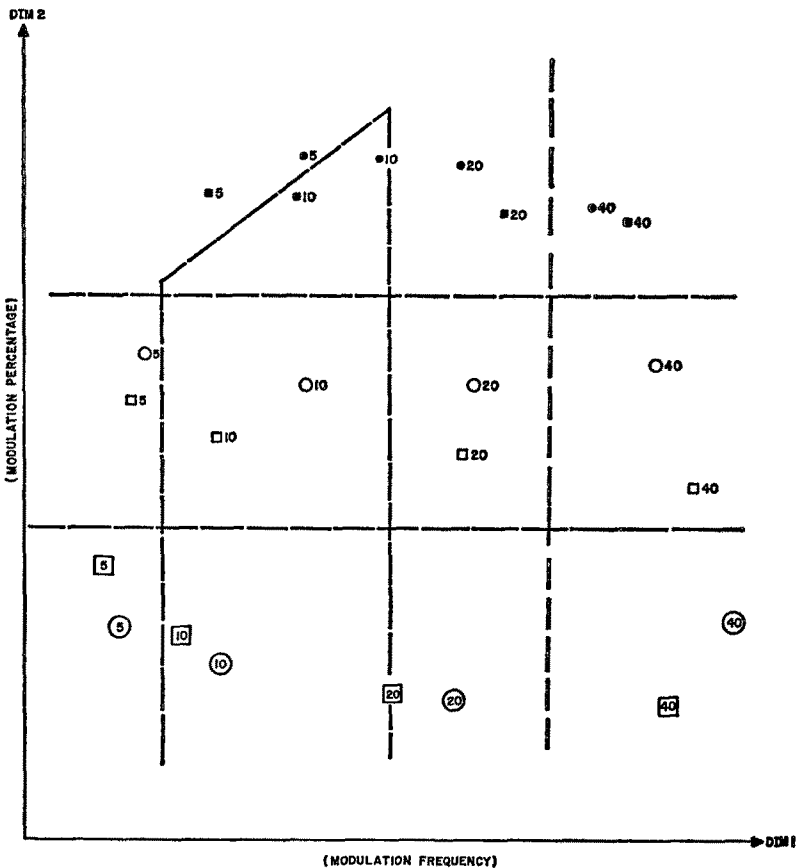

FIGURE 4

The one-two plane of the group stimulus space for the Bricker, *et al.* tone data, with dotted lines drawn to divide plane into regions corresponding to distinct levels of modulation frequency and modulation percentage.

In Figures 1 through 3 the projections of these property lines are shown in the planes formed by the three coordinate axes. In each plane we show only the projections of the two lines most closely corresponding to the two coordinate axes defining that plane. These figures provide a visual impression of the degree of correspondence between physical dimensions and the particular "psychological coordinates" determined by this analysis.

In Figures 4, 5 and 6, we show these same three planes in a somewhat different way. In these figures, an attempt has been made to divide the stimulus space up into regions corresponding to the various levels of the three physical variables. This division has been done, as far as possible, by lines
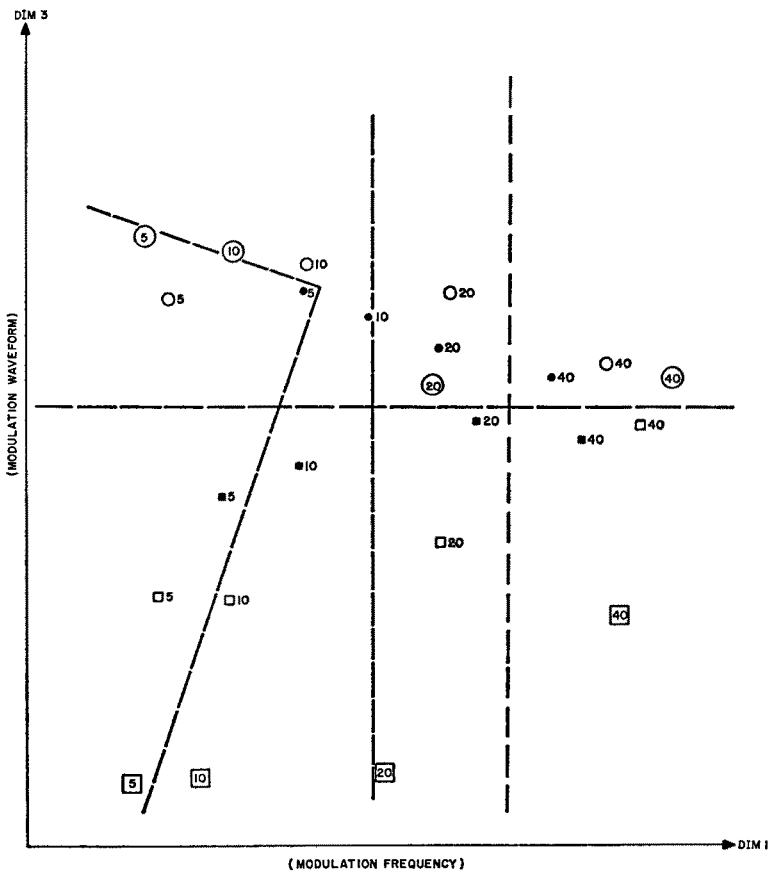

FIGURE 5

The one-three plane of the group stimulus space for the Bricker, *et al.* tone data, with dotted lines drawn to divide plane into regions corresponding to distinct levels of modulation frequency and modulation waveform.
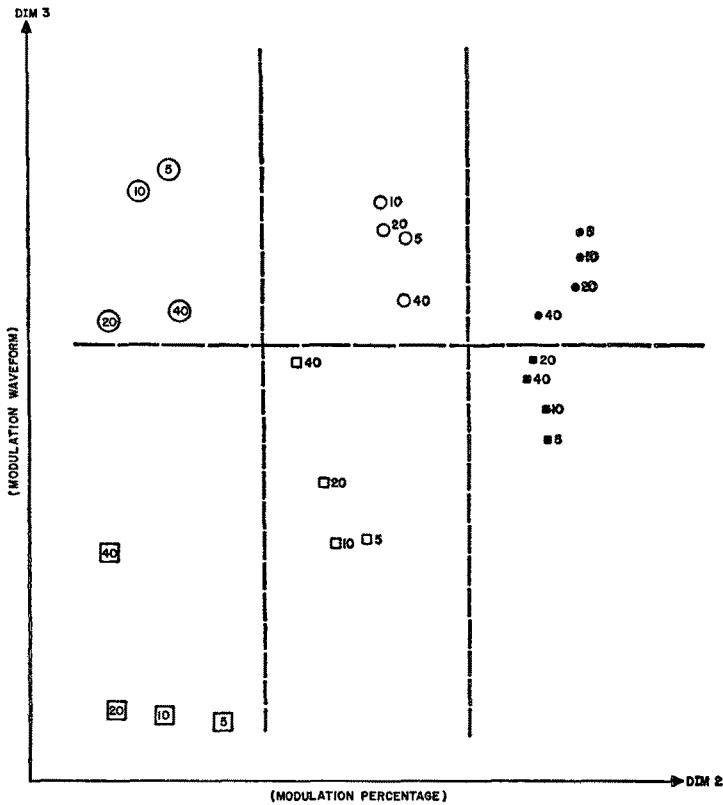
FIGURE 6

The two-three plane of the group stimulus space for the Bricker, *et al.* tone data, with dotted lines drawn to divide plane into regions corresponding to distinct levels of modulation percentage and modulation waveform.

perpendicular to the psychological axis most closely corresponding to the given physical dimension. The fact that we have been very nearly successful at this is indicative of the degree of correspondence between the unrotated dimensions from the individual differences scaling analysis and the physical dimensions. The only failure, as can be seen, involves the two 5 hertz tones at the lowest modulation percentage. This seems to be due to a mild interaction between modulation percentage and modulation frequency, that has the character of making modulation frequency somewhat less salient at this lowest modulation percentage than at the higher modulation percentage. Even these stimuli can be accommodated by allowing a boundary defined by two straight line segments rather than a single straight line perpendicular to the modulation frequency axis.

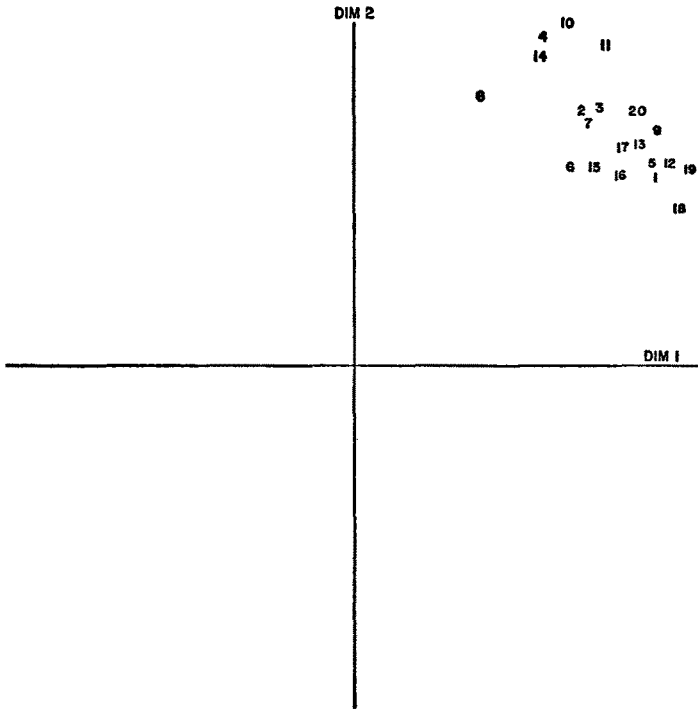Figures 7, 8 and 9 show the planes formed by the three corresponding

FIGURE 7
The one-two plane of the subject space for the Bricker, *et al.* tone data.

dimensions of the subject space. Recall that the coordinates of the point for a given subject in this space correspond to the *weights* of the various dimensions (in the stimulus space) for that subject. These weights *should* be positive or zero, since a negative weight means that the "distances" are not Euclidean distances at all (and, indeed, may not even satisfy the triangle inequality or other metric axioms, and may be imaginary rather than real numbers in some cases). In the canonical decomposition solution obtained here, however, no explicit positivity constraint is applied to the weights. However, as frequently happens with this method (this will be discussed in more detail later), all the weights *are* positive, as is evident in the fact that all the points lie in the positive octant of the three dimensional subject space.

In interpreting the subject space, it is important to keep in mind that in this space the origin is *not* arbitrary, but has a fixed meaning. In Stevens' terms, the weights are measured on a *ratio* rather than simply an interval scale. As noted earlier, distance from the origin corresponds, at least roughly, to variance accounted for, so that if a subject's point is precisely at the origin no variance at all is accounted for in the data for that subject. *Direction*

from the origin relates to the pattern of the data for that subject, or, more precisely, to the pattern of his "perceptual space" as solved for in this analysis. Two subjects who lie on the same straight line issuing from the origin would have configurations identical except for a single overall scale factor, so that the two configurations differ only by a similarity transformation.

One subject's being closer to the origin on that line would indicate simply that less of the variance in his data is accounted for by that common configuration (either because his data are noisier or because additional dimensions are needed to fully account for this subject's data). For some purposes, then, it may be more useful to describe the subject space in polar rather than rectangular coordinates.

It is of interest to note that the magnitude of individual differences for these data is relatively small, as attested to by the small spread of the subjects within the positive octant of the subject space. There seem to be almost no differences in the weights for dimensions 3, the principal difference being in the relative weights for dimensions 1 and 2 (which correspond, essentially, to modulation frequency and modulation percentage, respec-
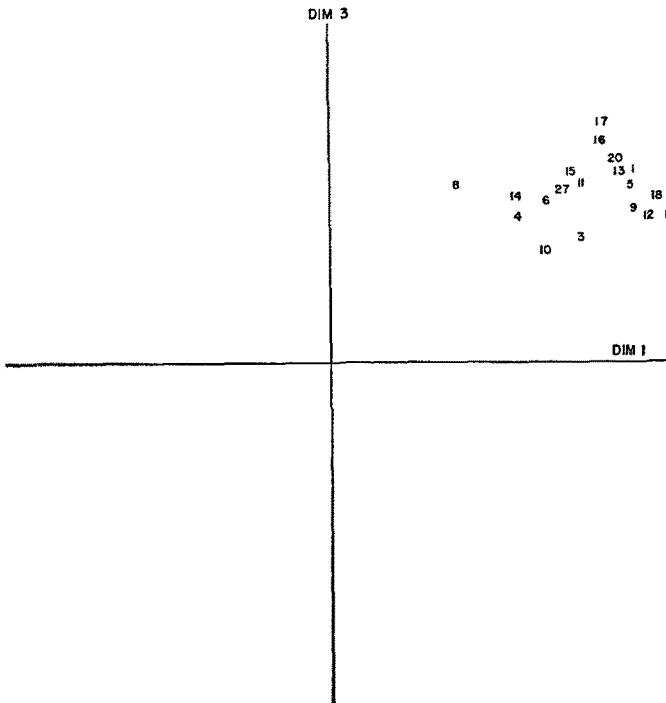

FIGURE 8
The one-three plane of the subject space for the Bricker, et al. tone data.
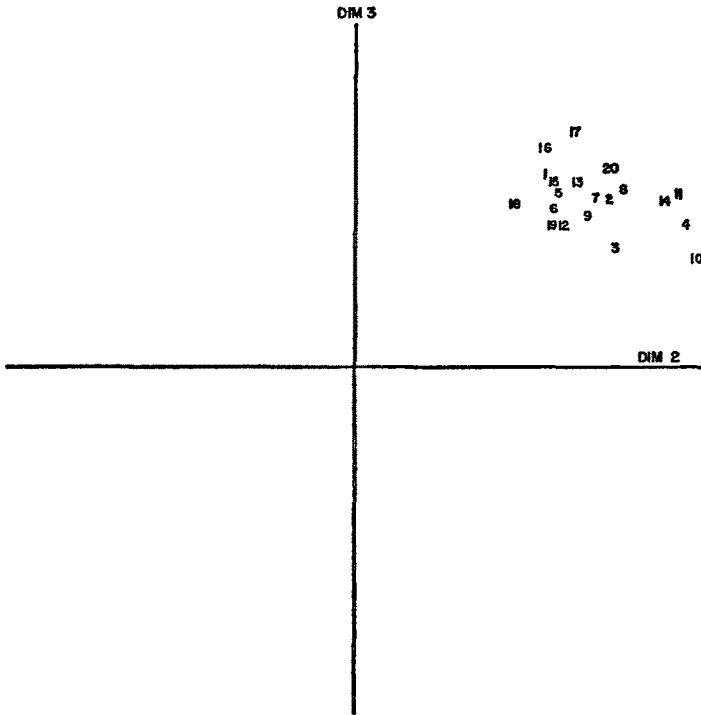
FIGURE 9

The two-three plane of the subject space for the Bricker, *et al.* tone data.

tively). Subjects 8 and 19 provide a good contrast in this respect. Subject 8 weights dimension 2 considerably more than dimension 1, while subject 19 shows the opposite tendency. Figure 10 contrasts the "perceptual spaces" for these two subjects, by showing plots of the planes formed by the three coordinate axes for each, transformed by the appropriate weighting factors (the square roots of the w's). In this figure no attempt has been made to make the scales between subjects comparable, but the scale factors have been maintained *within* the plots for each subject. It is evident from this figure that subject 8's "space" gives much more weight to dimension two, while the "space" for subject 19 weights dimension one more heavily.

A second set of data which illustrates very nicely the power of the subject space in discriminating among subjects was collected by Wish (in press) who has generously allowed its use here. The stimulus and subject spaces resulting from analysis of these data are shown in Figures 11 through 14. Since the description of subjects, data collection and basis for interpretation of dimensions are described by Wish (in press) we shall not describe these. The reader may confirm for himself the adequacy of labeling of the three

dimensions, called (by Wish) communist-noncommunist, economically de-
veloped-underdeveloped, and East–West, respectively. These are the *un-
rotated* dimensions from the individual differences scaling analysis. Wish
categorized his subjects as "dove", "moderate" or "hawk" (with respect
to attitudes on the Vietnam war) on the basis of an independent question-
naire. The straight line drawn in the one-two plane of the subject space
shows that the "doves" and "hawks" can be distinguished very well on
the basis of the dimensions of this space. We can characterize this distinction
by saying that the "doves" appear to give much more weight to economic
development than do "hawks", with the "moderates" falling, generally,
somewhere between (as evidenced by the fact that "moderates" fall on both
sides of the line). This example shows that the subject space provides a
useful paramaterization of the subjects, which may relate very systematically
to other relevant subject attributes.

Relation to Other Work on Individual Differences in Scaling

The work of Kruskal [1968] and McGee [1968] on individual differences
in scaling has already been mentioned. These approaches are both quite
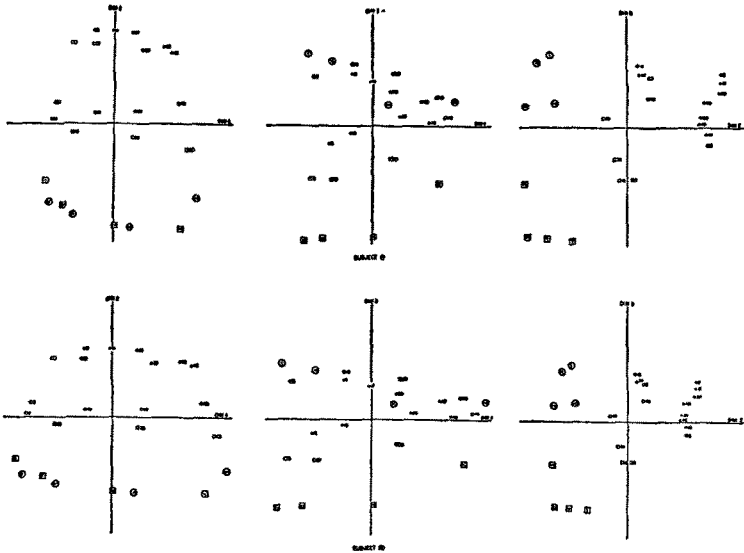

FIGURE 10

A comparsion of the "psychological space" of the tone stimuli for two different
subjects, as determined by applying the appropriate scale factors (square roots of the
w's) for these two subjects to the group stimulus space. The scale for within subjects
plots have been equated, but the scales *between* subjects are not comparable (*i.e.*, while
the *relative* "potencies" of the dimensions can be compared between subjects, the *absolute*
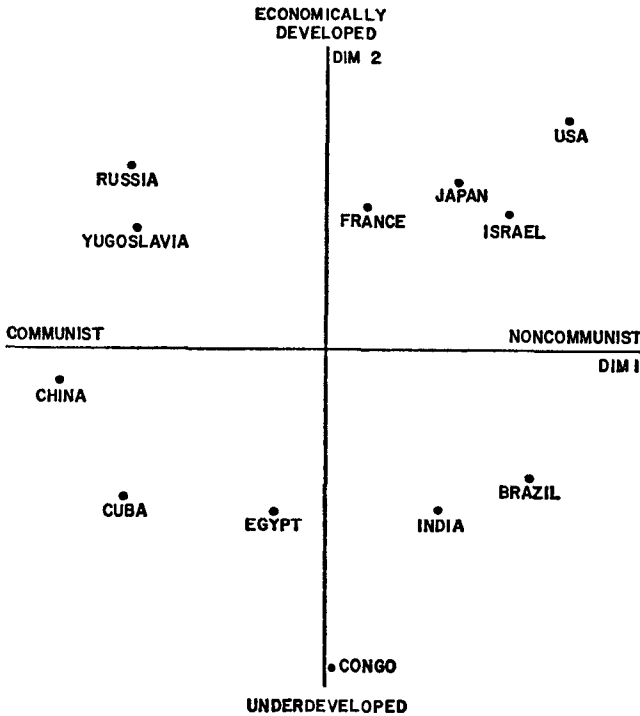sizes of the two configurations are arbitrary).

FIGURE 11

The one-two plane of the group stimulus space for 12 nations (data due to Wish). Dimensions one and two were interpreted by Wish as political alignment (communist-noncommunist) and economic development (economically developed-underdeveloped) respectively.

different in philosophy from the approach discussed here. Kruskal has two approaches. The first assumes each subject to have a different monotone function (relating distances to similarity or dissimilarity judgments) but constrains them to have *identically* the same configuration (no degrees of freedom for weighting of dimensions or the like are allowed). The second assumes all subjects to have the same monotone function, but allows each his own idiosyncratic configuration. These two represent two extremes of a continuum (or, perhaps, of two continua) of which there are, of course, many intermediate points. McGee's approach covers at least some of these intermediate points. McGee allows for either the case in which each subject has his own monotone function, or all are constrained to have the same. He then introduces a parameter that monitors the degree to which the configurations for different subjects are constrained to be similar. At one extreme, these configurations must be identical; at the other there is no constraint at all on how similar they must be. At intermediate values of this parameter, they

must be "intermediately" similar. McGee's approach, however, says nothing explicitly about how these configurations may depart from identity (the criterion of departure is simply a "sum of squared coordiante differences" criterion, which monitors degree, but not direction of departure from identity).

The Tucker-Messick procedure, which has already been touched on, also makes no explicit assumption about communality of dimensions among different subjects. We shall discuss this in more detail at a later point. For the moment, let us consider the work of two other investigators, both of whom have dealt with essentially the same model as the present authors.

Horan [1969] is the first author to publicly propose the model we have assumed here. Horan devised a method to solve for what we call the "group stimulus space" (he calls it the "normal attribute space") under the assumptions of our model. Horan's method is based on the observation that, if the model stated in our equation (2) is correct then,

$$(12) \qquad \sqrt{\frac{1}{m} \sum_{i=1}^{m} [d_{jk}^{(i)}]^2} = \sqrt{\sum_{t=1}^{r} \left(\frac{1}{m} \sum_{i=1}^{m} w_{it}\right)(x_{jt} - x_{kt})^2}$$

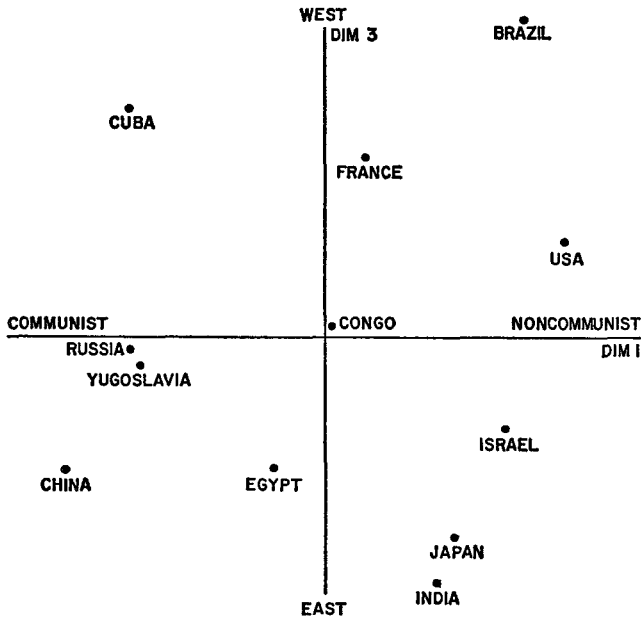so that the *root mean square* of the distances (over individuals) will be ordinary

FIGURE 12

The one-three plane of the group stimulus space for the Wish data on 12 nations. Wish interpreted dimension three as a geography dimension (East-West).
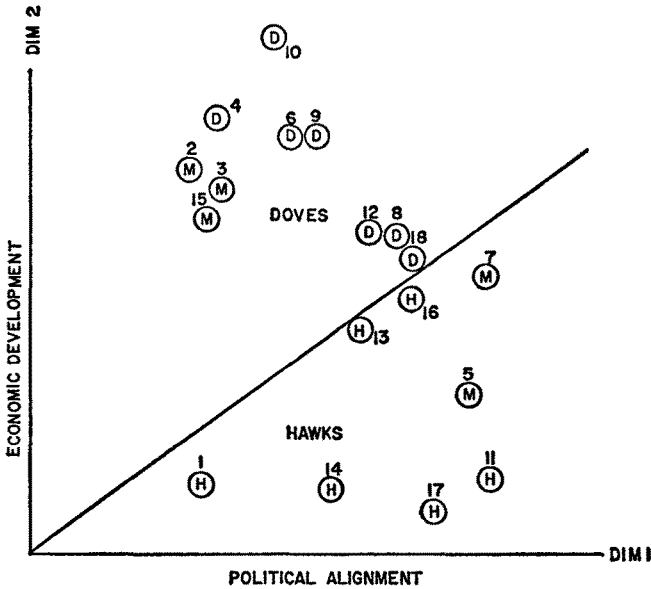
FIGURE 13

The one-two plane of the subject space for the Wish nation data. D, H and M stand for "dove", "hawk" and "moderate" (as determined by subjects' self report) vis à vis attitudes on Vietnam war. 45 degree line divides "doves" from "hawks", with "moderates" on both sides.

Euclidean distances in a space with coordinates y_{it}^* given by

$$(13) \qquad\qquad y_{it}^* = w_{.t}^{1/2} x_{it}$$

where

$$(14) \qquad\qquad w_{.t} = \frac{1}{m} \sum_{i=1}^{m} w_{it}$$

Thus, Horan shows, if the data are sufficiently strong to estimate ratio scaled distances, averaging the data via root mean squares will produce distances between points in a space which includes all the requisite dimensions. The individual spaces will then be related to this "common space" by at most a linear transformation.

The problem with this, from our point of view, is that there is nothing in Horan's averaging procedure to guarantee that the "common space" as derived from it will be described in terms of the correct orientation of axes. Since his procedure reduces all the distances to a common set of Euclidean distances, and then applies a scaling procedure to produce a space from these distances, the rotationally invariant property of Euclidean distances

means that no unique orientation of axes will be defined. While one *can* uniquely define the orientation of axes by fiat (say, by rotating to principal axes), there is nothing at all to guarantee that such an orientation will correspond to the "correct" orientation. Our procedure, we claim, does guarantee this unique orientation and, furthermore, solves at the same time for the matrix of weights.

The second approach is one due to Bloxom [1968], who derived the equations for solving numerically (via a gradient method optimization procedure) for the parameters of the model expressed in (1) through (4). At the time of the memorandum referred to above Bloxom had not actually programmed the procedure he proposed, but he now indicates (personal communication) that this has been accomplished. It has not yet been possible to compare the two procedures for relative efficiency, freedom from local minima, etc., but Bloxom has indicated (personal communication) that his procedure is somewhat slow in converging. His method has the advantage of being capable of solving for the "group stimulus space" (presumably with the *correct* orientation of axes) without simultaneously solving for subject weights (these can be solved for at a later time, if desired). Our method has the advantage, we believe, of generalizing very easily and straightforwardly
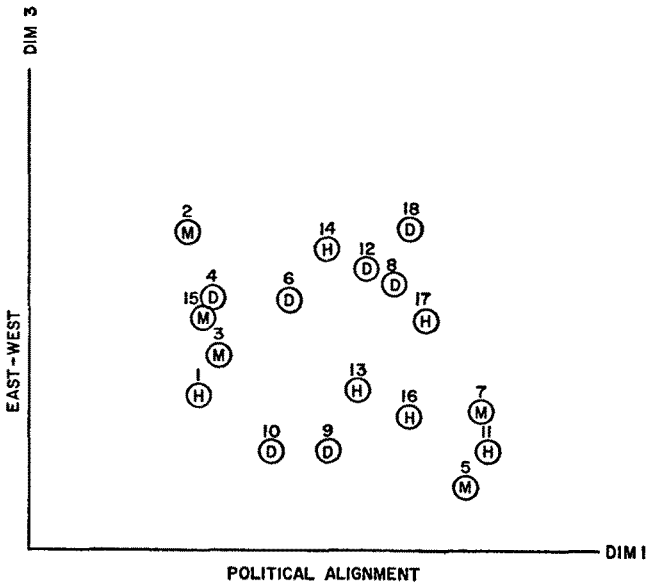

FIGURE 14

The one-three plane of the subject space for Wish nation data. Coding of subjects is as in Figure 13.

to the nonsymmetric and to the higher-way case. We believe it will also prove to be more efficient numerically.

The final procedure to which we shall compare this method in some detail is that of Tucker and Messick [1963]. They introduce an equation [their equation (16)] which has some resemblance to our equation (6). Rewriting this equation in notation closer to our own, this could be expressed as

$$(15) \qquad\qquad D \cong D^*V$$

where D is an $[n(n - 1)/2] \times m$ matrix whose elements are the distances, $d_{jk}^{(i)}$ (with the $n(n - 1)/2$ distance elements for each subject arrayed into an $n(n - 1)/2$ column vector), while D^* is an $[n(n - 1)/2] \times r$ matrix with elements d_{st}^* (for $s = 1, 2, \cdots n(n - 1)/2$) and V is an $r \times m$ matrix which bears a strong resemblance to the transpose of our matrix, W, of subject weights. The columns of D^* (after appropriate rotation) are supposed to correspond to patterns of distances or dissimilarities typical of "ideal subjects", while V represents a set of linear combinations transforming these "ideal subject distances" into the actually observed distances given in D. Tucker and Messick's equation (16) is actually more correctly stated, in our notation, as:

$$(16) \qquad\qquad \hat{D}_p = CV$$

where $\hat{D}_p$ is the best p dimensional approximation (in a least squares sense) to the originally given matrix D. (To make the translation from the above equation to Tucker and Messick's equation (16), we need only make the substitutions $\hat{D}_p = \hat{X}_p$, $C = Z$ and $V = B$). As Ross [1966] has pointed out, it does not make analytic sense to suppose that arbitrary linear combinations of Euclidean distances will yield Euclidean distances. It turns out, however, that had Tucker and Messick originally conceived their method in terms of *squared* Euclidean distances, no such theoretical problem would have arisen. To see this in terms of our model, as stated in (2), we may write the equivalent matrix equation:

$$(17) \qquad\qquad D^{[2]} = \Delta^{[2]} W^T$$

where $D^{[2]}$ is the $[n(n - 1)/2] \times m$ matrix whose entries are $[d_{jk}^{(i)}]^2$ (appropriately arrayed into a two-way matrix), while $\Delta^{[2]}$ is the $[n(n - 1)/2] \times r$ matrix whose general entry is given by:

$$(18) \qquad\qquad \delta_{(jk)t}^2 = (x_{jt} - x_{kt})^2$$

with the parentheses about the (jk) subscript pair indicating, again, the appropriate dispersal of the half matrix (for $j < k$) into an $n(n - 1)/2$ array. W has exactly the same meaning as in our earlier equations. This shows that if, in fact, the model we propose in equations (1) through (4) holds, the "subject space" determined from this modified Tucker–Messick analysis

ought to bear a close resemblance to our subject space. All that can be said in practice is that the two should be related by some affine transformation. At a minimum, a rotation would be required to bring them into correspondence. In fact, it is possible for the modified Tucker–Messick "subject space" to have fewer dimensions than our subject space. That is, p may be less than r. The two should, if the model we assume holds, have the same *rank*, however. This apparent anomaly is explained by the fact that our "subject space" is not constrained to be of full rank (that is, the dimensions of the subject space from our analysis may be linearly dependent). It is not possible, however, for both the "subject space" and the "group stimulus space" from our analysis to be of less than full rank. If this model holds and yet the "modified" Tucker–Messick "subject space" is of lower dimension than ours, it means that it is not possible to find a dimension preserving transformation that would make the $D^{[2]}$ matrix of the form described in (17) and (18). Rather, some or all of the columns of $\Delta^{[2]}$ would have to be (positive) linear combinations of the $[\delta_{jk}^{(i)}]^2$'s. As such they would themselves correspond to squared (weighted) Euclidean distances defined on the X-space, or to squared (unweighted) distances defined on an appropriate transformation of the X-space.

It should also be clear that if the D matrix that appears in (15) were factored nonmetrically, by, say, the Shepard–Kruskal nonmetric factor analysis (see Shepard, 1966) the result would be essentially the same as a metric factoring of the D^2 matrix (if the distances are, in fact, measured up to a ratio scale). This nonmetric analysis would also compensate for possible nonlinearities in the relation between similarities or dissimilarities and distances. Since, however, Shepard [1966] has shown via a Monte Carlo study that a metric factoring reveals essentially the same structure as the nonmetric if the correct dimensionality is known (differing from the nonmetric analysis primarily in the introduction of spurious extra dimensions), one might expect that it would not make too much difference whether D or $D^{[2]}$ is factored, as long as the dimensionality is correctly assessed. From this reasoning it would seem to follow that a strong relation should hold between at least a subspace of the "subject space" from the (unmodified) Tucker–Messick analysis and our own subject space. This conjecture was confirmed empirically by Green and Morris, [1969] who found very high canonical correlations between a three dimensional "subject space" from our analysis and the first three dimensions of the "subject space" from a Tucker–Messick points of view analysis. (Interestingly enough, the Tucker–Messick space will usually tend to be of higher dimension, although it could, theoretically, be of lower dimension. This appears to be primarily due to the extra dimensions that are needed to account for nonlinearities in the function relating similarities to squared distances.)

A more general reformulation of the Tucker–Messick approach in terms of squared distances is possible in the following form. Let us suppose there

exists a "group stimulus space" of r dimensions, and that the similarity judgments of individual subjects are (linearly) related to an arbitrary linear combination of these dimensions. That is to say, we assume, again, that

$$(19) \qquad s_{jk}^{(i)} = L(d_{jk}^{(i)})$$

where

$$(20) \qquad d_{jk}^{(i)} = \sqrt{\sum_{i=1}^{r} (y_{ji}^{(i)} - y_{ki}^{(i)})^2}$$

and

$$(21) \qquad y_{ji}^{(i)} = \sum_{i'=1}^{r} \beta_{ii'}^{(i)} x_{ji'}$$

(of course, the transformation matrix $\beta_i = \|\beta_{ii'}^{(i)}\|$ may be singular, and therefore rank reducing).

It might be noted at this point that the model expressed in (19), (20) and (21) is very general indeed. In effect, all it specifies in addition to the linearity assumption of equation (19) is that distances in the stimulus space of every subject are Euclidean. This is so because we can always define a Euclidean space sufficiently general that any one of a finite number of other Euclidean spaces can be defined as linear transformations of that space. One way to define this general space would be to define the direct sum of all the individual spaces, rotate to principal axes in this direct sum space, and eliminate axes corresponding to zero eigenvectors (as a way of eliminating redundant dimensions). Again, the linear assumption of (19) could be dropped if the factoring to be discussed below were done nonmetrically, and even the Euclidean assumption is not too critical, since *any* metric could be incorporated by allowing some of the components in the transformation matrices to be imaginary (which simply means that some or all of the R_i matrices defined below may not be positive definite or semi-definite; that is, they may have some negative eigenvalues).

If this very general model (as it always will in "sufficiently high" dimensionality), holds, it follows that:

$$(22) \qquad (d_{jk}^{(i)})^2 = \sum_{i=1}^{r} (y_{ji}^{(i)} - y_{ki}^{(i)})^2$$

$$= \sum_{i} \left[\sum_{i'} \beta_{ii'}^{(i)} x_{ji'} - \sum_{i'} \beta_{ii'}^{(i)} x_{ki'} \right]^2$$

$$= \sum_{i} \left[\sum_{i'} \beta_{ii'}^{(i)} (x_{ji'} - x_{ki'}) \right]^2$$

$$= \sum_{i} \sum_{i'} \sum_{i''} \beta_{ii'}^{(i)} \beta_{ii''}^{(i)} (x_{ji'} - x_{ki'})(x_{ji''} - x_{ki''})$$

$$= \sum_{t'} \sum_{t''} (x_{jt'} - x_{kt'})(x_{jt''} - x_{kt''}) \sum_{t} \beta_{tt'}^{(i)} \beta_{tt''}^{i}$$

$$= \sum_{t'} \sum_{t''} r_{t't''}^{(i)}(x_{jt'} - x_{kt'})(x_{jt''} - x_{kt''})$$

where

(23) $$r_{t't''}^{(i)} \equiv \sum_{t} \beta_{tt'}^{(i)} \beta_{tt''}^{(i)}.$$

If we, now, let $s = (j, k)$ range from 1 to $n(n - 1)/2$ and (noting that $r_{t't''}^{(i)} = r_{t''t'}^{(i)}$) let $u = (t't'')$ range from 1 to $r(r + 1)/2$ (note that diagonal elements must be included here), and define

(24) $$\delta_{su} \equiv \delta_{(jk)(t't'')} = (x_{jt'} - x_{kt'})(x_{jt''} - x_{kt''})$$

and

(25) $$r_{(t't'')i}^{*} \equiv r_{ui}^{*} = r_{t't''}^{i}$$

then (22) can be expressed as

(26) $$(d_{jk}^{(i)})^2 = \sum_{u} \delta_{su} r_{ui}^{*}$$

or, in matrix form, as

(27) $$D^{[2]} = \Delta(R^*)^T$$

where $D^{[2]}$ has the same meaning as in equation (17), Δ is the $[n(n - 1)/2] \times [r(r + 1)/2]$ matrix with general entry $\delta_{(jk)(t't'')}$ and R^* has general entry $r_{(t't'')i}^{*}$, as defined above. We might note that, if the more general model in (19) through (21) holds, the columns of the Δ matrix do not correspond at all to distances, in any space, but rather to products of (signed) first differences *between* pairs of dimensions. It is no wonder, then, that in many situations the results of the "points of view" analysis are not meaningfully interpretable, for many of the "distance profiles" obtained from it may correspond to just such nondistance like entities.

It should be noted, here, that we are making the fairly strong assumption that the data provide distances measured on a ratio scale, and that it is their *squares* that are entered into the Tucker–Messick analysis (rather than their first powers). It seems that, even when the analysis is done in this "correct" manner anomalies may arise. When these assumptions do not hold, other kinds of distortions can occur. Some of these may be represented by the factors in the "subject space" reflecting differences across subjects in the monotone transformation applied to the posited underlying distances.

Finally, it would appear to be relevant to generalize Horan's procedure to the case where the more general model expressed in (19) through (21) holds. If we start with (22) for $(d_{jk}^{(i)})^2$, sum over i and divide by m (the number of subjects), we get

$$(28) \quad (d_{jk}^{(\cdot)})^2 \equiv \frac{1}{m} \sum_{i=1}^{m} (d_{jk}^{(i)})^2 = \sum_{t'} \sum_{t''} (\frac{1}{m} \sum_i r_{t't''}^{(i)})(x_{jt'} - x_{kt'})(x_{jt''} - x_{kt''})$$

$$= \sum_{t} \sum_{t''} r_{t't''}^{(\cdot)}(x_{jt'} - x_{kt'})(x_{jt''} - x_{kt''})$$

where $r_{t't''}^{(\cdot)}$ is the average (over subjects) of the $r_{t't''}^{(i)}$'s.

If we let $R_i \equiv \|r_{t't''}^{(i)}\|$, then $R_{\cdot} \equiv \|r_{t't''}^{(\cdot)}\|$ is just the average of these matrices. Since each R_i is a positive definite or semi-definite matrix, $R_{\cdot}$, being a positive linear combination of the R_i's must also be positive definite or semi-definite, and can therefore be decomposed into a product of the form

$$(29) \qquad\qquad\qquad R_{\cdot} = \beta^T \beta.$$

or, in summational form

$$(30) \qquad\qquad\qquad r_{t't''}^{(\cdot)} = \sum_t \beta_{tt'}^{(\cdot)} \beta_{tt''}^{(\cdot)}.$$

Tracing the development in (22) backwards, it follows that $(d_{jk}^{(\cdot)})^2$, as defined in (28), can be written as

$$(31) \qquad\qquad\qquad (d_{jk}^{(\cdot)})^2 = \sum_{t=1}^{r} (y_{jt}^{(\cdot)} - y_{kt}^{(\cdot)})^2$$

where

$$(32) \qquad\qquad\qquad y_{jt}^{(\cdot)} = \sum_{t'} \beta_{tt'}^{(\cdot)} x_{jt'}.$$

So that the root mean square distances are again, under these more general assumptions, Euclidean distances in a space containing all the requisite dimensions.

The elements of the individual transformation matrices (the β_i's) could then be determined by regression procedures. To see how the β_i's could be so defined, consider (27), where $D^{[2]}$ is the basic data matrix defined earlier, and Δ is defined from the matrix of coordinates of stimulus points as in (24). It is clear from this equation that, given $D^{[2]}$ and Δ we may solve for the least squares estimate of R^* yielding the equation:

$$(33) \qquad\qquad\qquad \hat{R}^* = (\Delta^T \Delta)^{-1} \Delta^T D^{[2]}$$

Having solved for $\hat{R}^*$, we may "unpack" its entries to form the square symmetric matrices $\hat{R}_i$ (the least squares estimates of the R_i's). Finally, each $\hat{R}_i$ can be factored into a product of the form:

$$(34) \qquad\qquad\qquad \hat{R}_i = \hat{\beta}_i^T \beta_i$$

This factoring could be accomplished by diagonalizing $\hat{R}_i$, and defining the rows of β_i to be the eigenvectors multiplied by the square roots of the corresponding eigenvalues. If $\hat{R}_i$ is positive definite or semi-definite $\hat{\beta}_i$ will

be a real matrix; otherwise it will contain some imaginary entries, as discussed above (of course, if the negative roots are small, they could be assumed to correspond to "true" values of zero, and so actually set equal to zero).

The matrix R_i could also be expressed as

$$(35) \qquad\qquad R_i = T_i^T W_i T_i$$

where T_i is an *orthogonal* transformation matrix and W_i is a diagonal matrix with a set of weights for dimensions along the diagonal. This shows more clearly the relation between this more general model and our earlier model. In effect this more general model allows an orthogonal rotation of the system of coordinate axes followed by differential weighting of dimensions as defined by the new coordinate axes, which allows circles or spheres to be transformed into ellipses or ellipsoids whose axes are not (necessarily) parallel to the (original) coordinate axes. In this case, $\hat{T}_i$, the "estimate" of T_i would be defined directly by the eigenvectors of $\hat{R}_i$, while the diagonals of W_i are given by the eigenvalues. This formulation makes clearer the meaning of negative eigenvalues.

The procedure outlined above provides a complete solution for the model defined in (19), (20) and (21). It is not, however, a least squares solution, nor is there any other well defined criterion of fit that it satisfies. However, it probably would be very close to the least squares solution, and should be quite adequate in practice. One could derive an exact least squares solution, incidentally, by use of a NILES estimation scheme similar to the one we have used to solve for the parameters of our earlier model.

It may have occurred to the reader that it might be possible to apply the procedure originally proposed by Horan [1969] to get a "group stimulus space" (or what he calls a "normal attribute space") under the more restrictive assumptions expressed in our equations (1) through (3), and then to use regression procedures to solve for the subject weights. This, however, is not the case. The reason for the unfeasilibity of this procedure is the lack of rotational invariance (as discussed earlier) of the dimensions postulated in this more restrictive model. Since the dimensions as derived from Horan's procedure are necessarily in an arbitrary orientation, it would not be appropriate to compute weights for such dimensions. Apart from the theoretical inelegance of this, this computation could not be expected to account for as much variance in the (scalar products) data as would the more appropriate analysis. Furthermore, the important advantage of unique orientation of axes would be lost.

Ironically, then, it is *easier* computationally to analyze data in terms of the more general and complex model of (19) through (21) than in terms of the simpler and more restrictive model of (1) through (3). We believe it to be advisable, however, to analyze data first in terms of the stronger model for, if this model is true, one is likely to learn much more about one's data. In

particular, a unique orientation of axes is obtained which, if the model is correct, will correspond to *psychologically meaningful* dimensions (and, indeed, our method will "seek out" this correct orientation of axes at the same time it is finding the right general "space"). Secondly, we believe a model of this sort to be more easily interpretable than one postulating different rotations as well as different weightings of dimensions. The principle of parsimony, after all, is not strictly a matter of counting the number of parameters in an equation, but also has something to do with the meaningfulness of these parameters to a human interpreter. We believe the model we have proposed here is parsimonious in this general sense. We also believe, as is argued quite effectively by Horan [1969], that this model makes a great deal of psychological sense and, in fact, holds in a wide variety of judgmental situations.

Canonical Decomposition of General 3-Way Tables

We shall now turn to a consideration of the general case of three mode data which we want to analyze in terms of a model of the form:

$$(36) \qquad\qquad z_{ijk} \cong \sum_{t=1}^{r} a_{it} b_{jt} c_{kt}$$

where i, j and k range from 1 to n_1 , n_2 and n_3 , respectively. It can easily be seen that the earlier model for individual differences in scaling is a special case of this. This can be seen especially clearly by comparing (36) to the earlier (4-a). If we let $n_2 = n_3$ and make the identifications $a_{it} = w_{it}$, $b_{jt} = x_{jt}^{(L)}$ and $c_{kt} = x_{kt}^{(R)}$ the translation is complete. As will be discussed later, the model in (36) is very similar to Tucker's three mode factor analysis model, but is restricted in two important ways. These are, first of all, that the dimensionality of all three "spaces" must be the same (while in Tucker's model all three may have different dimensionalities). The second important constraint is that the "core matrix" introduced by Tucker is restricted to be a kind of 3-way analogue of an identity matrix (to be precise, it is a three-way diagonal matrix which, by convention, can be constrained to have only ones on the diagonal).

The method of analysis for this model is, of course, very similar to the method described for the earlier special case. At the risk of redundancy, however, it will be described here in somewhat more general terms. We shall then show that it generalizes quite nicely and straightforwardly to the multiway case (*i.e.*, the case in which z may have more than three subscripts). This "canonical decomposition" procedure has, in fact, been programmed for up to seven-way tables.

Starting with some initial estimates of the b's and c's, we compute least squares estimates of the a's, holding the b's and c's fixed. This can be done by standard multiple regression procedures. The least squares estimate for the a's can be stated in matrix equations as:

(37)
$$A = PQ^{-1}$$

where A is the $n_1 \times r$ matrix whose general entry is a_{it}, while P is another $n_1 \times r$ matrix whose entries, p_{it}, are given by

(38)
$$p_{it} = \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} z_{ijk} b_{ji} c_{kt}$$

and Q is an $r \times r$ matrix whose general entry q_{st} is given by

(39)
$$q_{st} = \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} (b_{js} c_{ks})(b_{jt} c_{kt})$$

$$= \left(\sum_{j=1}^{n_2} b_{js} b_{jt} \right) \left(\sum_{k=1}^{n_3} c_{ks} c_{kt} \right)$$

To see that (37), (38) and (39) do, indeed, give the appropriate least squares estimates for the a's, we may rewrite (36) as:

(40)
$$z_{is}^{*} \cong \sum_{t=1}^{r} a_{it} g_{st}$$

where $s = j \cdot k$ (and so ranges from 1 to $n_2 \cdot n_3$), while $g_{st} = b_{jt} \cdot c_{kt}$. If we let Z^* represent the $n_1 \times (n_2 \cdot n_3)$ matrix with entries z_{is}^*, and G the $(n_2 \cdot n_3) \times r$ matrix with entries g_{st}, equation (40) can be written in matrix form as:

(41)
$$Z^* \cong AG^T$$

To determine the least squares estimate of A we postmultiply Z^* by the right pseudoinverse of G^T, yielding

(42)
$$\hat{A} = Z^* G(G^T G)^{-1}$$

This equation can be seen to be the same as equation (37), if we make the substitutions:

(43)
$$A = \hat{A}$$
$$P = Z^* G$$
$$Q = G^T G$$

It should be evident from inspection, and from the definitions of Z^* and G, that (38) and (39) are equivalent to the equations in (43) for defining the elements of the matrices P and Q.

Having solved for these least squares estimates of the a's holding the b's and c's fixed, we may then, in a similar manner, solve for better estimates of the b's holding the a's and c's fixed, and then, similarly, solve for the c's. When this is done, we may repeat the whole process to determine still better estimates of the a's, b's and c's, and continue this procedure, iteratively, until no further improvement is possible. While there is as yet no proof that

position of interaction numbers once additive effects have been subtracted).

This "NILES" estimation scheme seems to converge quite rapidly, even when completely arbitrary starting values are used, and to be relatively unsusceptible to problems of local minima. Further theoretical work would seem appropriate to explore the properties of such schemes, which could be applied to a wider variety of models, including Tucker's more general model for multi-way factor analysis (hopefully yielding an exact least-squares solution, rather than an approximate one) and, as we have mentioned, the model defined in (19) through (21).

A Monte Carlo Study with Random Data

In order to get a better idea of the characteristics of this analytic model, the authors undertook to study the results of applying the method to completely random data (this can be thought of as a Monte Carlo exploration of the behavior of the method under the "null hypothesis"). The particular method tested in this way was that described in the first sections of this paper for the special case of analysis of individual differences in multidimensional scaling. The random data were generated for a hypothetical set of 25 stimuli and 12 subjects. An off diagonal half matrix of hypothetical similarities was generated for each subject. The hypothetical similarities were independently normal with 0 mean and common variance. These half matrices were folded over to form complete (symmetric) similarities matrices (with diagonals missing). The procedure for estimating an additive constant described in Torgerson [1958, pp. 268–277] was then applied to convert these pseudo-similarities to pseudo-distances. The additive constant estimation scheme used assumes that some three points lie precisely on a straight line in the underlying space, and has the property that it guarantees that the triangle inequality will be satisfied for all triples of points. This means that the "pseudo-distances" are, in fact, distances (in the sense of satisfying the metric axioms), but they are not, in general, distances in a Euclidean space (nor are they likely to be distances in any meaningful metric space of small dimensionality). The pseudo-distances were then converted to pseudo-scalar products, which were normalized and analyzed by the procedure described above. Five separate analyses were done, in one through five dimensions. In this analysis a version of the program was used that ordered the dimensions on the basis of the sum of squares of subject weights (which, as mentioned earlier, corresponds roughly to variance accounted for) and the first $r - 1$ dimensions of the r dimensional solution were then used as starting configuration for the $r - 1$ dimensional solution. To check that this procedure was not biasing the results in any way, one of the solutions (the two dimensional one) was rerun with a random starting configuration. The solution was indistinguishable from the one gotten in the other way, so it was assumed that no such bias existed.

Several things are of interest. One is the way in which variance accounted for changes with dimensionality. These figures are shown in Table 1. Also shown in Table 1 is the ratio of degrees of freedom (*i.e.*, number of free parameters) in the solution to those in the data. The number of degrees of freedom in the solution for an r dimensional solution is just $r(m + n - 2)$, where m and n are the number of subjects and of stimuli respectively. (This number changes linearly with r rather than being negatively accelerated because there are no orthogonality constraints on successive dimensions. The only effective constraints are that the sum of coordinate values be zero and the sum of squares one for each dimension of the stimulus space, which is reflected in the fact that two degrees of freedom are subtracted from $m + n$.) In the present case, with $m = 12$ and $n = 25$, this expression is simply $35r$. The degrees of freedom for the data is simply $mn(n - 1)/2 = 3600$. Thus in the present case the ratio of those degrees of freedom is about $.01r$. This ratio seems relevant because, if the model were strictly linear in its parameters (which it is not), one would expect the proportion of variance accounted for approximately to equal this ratio. In the case of nonlinear models, this will not generally be true, but may still provide a rough guideline. It would thus be of interest to see how that proportion compares to the "degrees of freedom ratio." The last column of Table 1, therefore, shows the ratio of these two numbers. We can see from this column of figures that the variance accounted for tends to be greater than one would expect based on the "degrees of freedom ratio", particularly for the small dimensionalities. Thus, the value of VAF/DFR should, it would seem, be fairly large (certainly larger than 5.0) for one to assume a "significant" fit (the precise figure, would depend, of course, on the exact number of subjects and stimuli, and to establish it would require extensive Monte Carlo and/or theoretical work).

The second point of interest has to do with the weights in the space for the pseudo-subjects. These turn out to be positive in almost every case, even though these data are completely random (the one exception was one subject who was slightly negative on dimension one in the 3, 4 and 5 dimensional solutions). Evidently the fitting of the "additive constant", because it assures satisfaction of the triangle inequality, is sufficient to make almost all the weights positive. This means that positivity of the weights need not be constrained (if the additive constant is fit), but, on the other hand, one would not necessarily expect to detect a poor fit of the model on the basis of occurrence of negative weights. Experience with real data, however, suggests that if the model is *systematically* violated, this may be reflected in negative weights.

Summary and Discussion

While further investigation is clearly indicated, the results presented here are very encouraging as to the validity and utility of a model accounting

TABLE 1

Variance Accounted For in Random Data

Dimensionality	Variance Accounted For (VAF)	D. F. Ratio (DFR)	VAF/DFR
1	.04864	.00972	5.003
2	.09653	.01944	4.966
3	.13302	.02917	4.560
4	.16471	.03889	4.235
5	.19232	.04861	3.956

for individual differences in similarity judgments in terms of differential "saliences" of a common set of underlying perceptual dimensions.

The *method* proposed for analyzing such data promises to be very useful, not only for individual differences in scaling, but for analysis of other, more general, kinds of data as well. There are a number of practical benefits of this method in the present situation. While the analysis may, in many cases, require no more computer time than a single multidimensional scaling analysis of averaged data, it provides, in a meaningful way, simultaneous scalings of the data for all the individual subjects. It is certainly more efficient than doing separate multidimensional scaling analyses for the individual subjects. Such individual analyses may not be feasible, furthermore, because the data for any given individual may be too weak or noisy to support such an analysis. The present method will tend to overcome such deficiencies by taking advantage of communalities among subjects. The result of this is likely to be more useful than individual scalings, since the summary provided by the "group" stimulus space together with the "subject" space defined by the weights will usually be considerably more succinct and easily comprehended than a series of unconnected individual spaces. The individual spaces can be constructed if desired, by appropriately weighting the "group" space. For many purposes, however, this may not be necessary for an adequate comprehension of the data. Another advantage is that the "group" space in many cases provides a useful basis of comparison of the subjects—relating to such other criteria as cognitive styles, political ideology, and the like.

A principal disadvantage of the method is that it is limited to the case in which individual subject spaces are related by *linear* transformations of a common space. Even the linear transformations allowed are not general,

but are restricted to those given by *diagonal* transformation matrices. The method may require too many dimensions in cases where the perceptual spaces represent *nonlinear* distortions of a common space (or, where more general linear transformations are required). Our evidence to date indicates that the model underlying the present method of analysis accounts very well for data in a number of different domains. Furthermore, the method can hardly fail to give a more useful description of the data than would individual analyses of subjects' data. We believe, too, that its results are likely to be more useful than those obtainable from a "Points of view" analysis, even in cases involving nonlinear distortions.

As described above the individual differences scaling method must be classified as a "metric" rather than a "nonmetric" one. Thus it requires the relatively strong assumption of a linear relation between data (on similarities or proximities) and distances, or that the function relating data to distances be known or assumed. A version has been produced that is at least "quasi-nonmetric," however [Carroll and Chang, 1970], by using a device very similar to one employed by Torgerson, in his version of non-metric scaling (see Young and Torgerson, 1968). As applied in the present case, this involves: [1] attaining a solution based on linear assumptions, [2] computing inter-stimulus distances for each subject by the weighted Euclidean formula of (1), [3] using the monotone regression algorithm described in Kruskal (1964) to find the monotone function of the data (either for all subjects combined or separately for each subject) best agreeing with these distances, and [4] *replacing* the data dissimilarities with values from this best-fitting monotone function of the data values (while retaining the original data, or at least the rank orders of data values, for purposes of future monotone regressions). Steps [1] through [4] are then repeated, iteratively, until no further change occurs in the solution. Such a procedure does not lead to a solution satisfying a least squares criterion of fit (or any other explicit badness-of-fit function, but ought to provide a very good approximation to such a solution.

Explorations with both real and artificial data have indicated, however, that this "quasi-nonmetric" version of individual differences scaling yields solutions in practice which are virtually indistinguishable from those attained by the "metric" version. Evidently the additional constraints imposed by the use of data from different subjects make this procedure even more "robust" against nonlinearities in the "distance function" than is ordinary metric scaling. While the "quasi-nonmetric" version may have some advantage in dimensionality estimation, it does not appear to provide sufficient advantage in determination of configuration (given that the dimensionality is correctly chosen) to make it worth the additional cost in computer time involved (at most one or two replications of steps (1) through (4) above would be recommended in any case). A "fully nonmetric" procedure could, conceivably, behave quite differently, but this seems dubious to the present authors.

To make the procedure fully nonmetric one would have to introduce additional iterations directed at numerical optimization of an explicit badness-of-fit function. Such a numerical procedure would also allow one to handle missing data in a coherent fashion. In the present method the only approach available for handling missing data is to use some scheme for estimating the missing entries. This is, in general, not a completely satisfactory solution.

Perhaps one of the strongest points of this method is its potential generalization to the "higher-way" case, making it possible, for example, to analyze confusions data for different people at different points in the learning process, thus generating, in addition to stimulus dimensions and subject weights, another set of weights characterizing the learning trials, or to analyze different *kinds* of proximities data for different subjects. The method of "canonical decomposition of N-way tables" also has more general potential applications to quite different kinds of data, such as data arising in the context of factorially designed experiments, or as a new variety of multi-way factor analysis.

REFERENCES

Bloxom, B. Individual differences in multidimensional scaling. Research Bulletin 68-45. Princeton, N. J.: Educational Testing Service, 1968.

Bricker, P. D. & Pruzansky, S. A comparison of sorting and pair-wise similarity judgment techniques for scaling auditory stimuli. Unpublished manuscript, Bell Telephone Laboratories, 1970.

Bricker, P. D., Pruzansky, S., & McDermott, B. J. Recovering spatial information from subjects' clusterings of auditory stimuli. Paper presented at the meeting of The Psychonomic Society, St. Louis, October–November 1968.

Carroll, J. D. Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, and S. Nerlove (Eds.) *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences.* 1971, in press.

Carroll, J. D. & Wish, M. Multidimensional scaling of individual differences in perception and judgment. Unpublished manuscript, Bell Telephone Laboratories, 1970.

Carroll, J. D. & Chang, J. J. A "quasi-nonmetric" version of INDSCAL, a procedure for individual differences multidimensional scaling. Paper presented at the meeting of the Psychometric Society, Stanford, March, 1970.

Cliff, N. The 'idealized individual' interpretation of individual differences in multidimensional scaling. *Psychometrika*, 1968, 33, 225–232.

Eckart, C. & Young, G. The approximation of one matrix by another of lower rank. *Psychometrika*, 1936, 1, 211–218.

Green, P. E. & Carmone, F. Stimulus context and task effects on individuals' similarities judgments. In C. W. King & D. Tigert (Eds). *Attitude research reaches new heights.* American Marketing Association, 1971, in press.

Green, P. E. & Morris, T. W. Individual difference models in multidimensional scaling: an empirical comparison. Unpublished manuscript, University of Pennsylvania, 1969.

Green, P. E. & Rao, V. R. Dissimilarities data-disaggregate analysis. Unpublished manuscript, University of Pennsylvania (Chapter 3 of monograph, in preparation, entitled: *User's guide to multidimensional scaling: A comparison of approaches and algorithms*), 1970.

Horan, C. B. Multidimensional scaling: combining observations when individuals have different perceptual structures. *Psychometrika*, 1969, **34**, 139–165.

Kruskal, J. B. How to use MDSCAL, a program to do multidimensional scaling and multidimensional unfolding. Unpublished report, Bell Telephone Laboratories, 1968.

Lyttkens, E. On the fix-point property of Wold's iterative estimation method for principal components. In Krishnaiah, P. R. (Ed.) *Multivariate analysis*. New York: Academic Press, 1966. Pp. 335–350.

McGee, V. E. Multidimensional scaling of N sets of similarity measures: A nonmetric individual differences approach. *Multivariate Behavioral Research*, 1968, **3**, 233–248.

Rao, V. R. The salience of price in the perception and evaluation of product quality: A multidimensional measurement model and experimental test. Unpublished doctoral dissertation, University of Pennsylvania, 1970.

Ross, J. A remark on Tucker and Messick's "points of view" analysis. *Psychometrika*, 1966, **31**, 27–31.

Shepard, R. N. Metric structures in ordinal data. *Journal of Mathematical Psychology*, 1966, **3**, 287–315.

Torgerson, W. S. *Theory and methods of scaling*. New York: Wiley, 1958.

Tucker, L. The extension of factor analysis to three-dimensional matrices. In Frederiksen, N. and Gulliksen, H. (Eds.) *Contributions to mathematical psychology*. New York: Holt, Rinehart and Winston, 1964. Pp. 109–127.

Tucker, L. R. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 1966, **31**, 279–311.

Tucker, L. R. & Messick, S. An individual difference model for multidimensional scaling. *Psychometrika*, 1963, **28**, 333–367.

Wish, M. Comparisons among multidimensional structures of nations based on different measures of subjective similarity. In A. Rapoport (Ed)., General Systems, Vol. 15, 1970, in press (a).

Wish, M. Individual differences in perceived similarity and friendship of nations. Submitted to Journal of Personality and Social Psychology. (b).

Wish, M. Individual differences in perceptions and preferences among nations. In C. W. King & D. Tigert (Eds). *Attitude research reaches new heights*, American Marketing Association, 1971, in press.

Wish, M., Deutsch, M. & Biener, L. Differences in perceived similarity of nations. In R. N. Shepard, A. K. Romney & S. Nerlove (Eds). *Multidimensional scaling: Theory and applications in the behavioral sciences*, 1971, in press.

Wold, H. Estimation of principal components and related models by iterative least squares. In Krishnaiah, P. R. (Ed.) *Multivariate analysis*, New York: Academic Press, 1966, Pp. 391–420.

Young, F. W. and Torgerson, W. S. TORSCA, a FORTRAN IV program for Shepard-Kruskal multidimensional scaling analysis. *Behavioral Science*, 1967, **12**, 498.