



An Introduction to Nonmetric Multidimensional Scaling

George B. Rabinowitz

American Journal of Political Science, Vol. 19, No. 2. (May, 1975), pp. 343-390.

Stable URL:

<http://links.jstor.org/sici?&sici=0092-5853%28197505%2919%3A2%3C343%3AAITNMS%3E2.0.CO%3B2-F>

American Journal of Political Science is currently published by Midwest Political Science Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/mpsa.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

GEORGE B. RABINOWITZ

The University of North Carolina at Chapel Hill

*An Introduction to Nonmetric Multidimensional Scaling**

Nonmetric multidimensional scaling methods are useful for spatially representing the interrelationships among a set of data objects. In this, they are similar to factor analytic methods. The assumptions and procedures associated with these methods are, however, somewhat different from those associated with factor analysis, and are more appropriate to certain political data. In this paper the logic underlying nonmetric multidimensional scaling methods is described, and some guides for using these procedures are offered.

Nonmetric multidimensional scaling techniques are among the set of procedures available to investigators interested in spatial representation of political objects. These techniques are useful in illuminating the structure hidden in a complex data matrix, and form an important addition to the factor analytic methods which have been widely used in the discipline. They have achieved considerable popularity in recent years, primarily for three reasons. First, they often yield solutions in a sufficiently low dimensionality to permit a visual examination of the structure. This is an invaluable interpretative aid. Second, they permit the investigation of many matrices which cannot be congenitally analyzed using factor analysis. Third, they make only ordinal assumptions about the data, which is often advantageous given the "weak" nature of most social science data. This paper is a general introduction to nonmetric multidimensional scaling.¹

*I wish to thank Stuart Rabinowitz, Herbert Weisberg, the members of the Comparative Politics Discussion Group at the University of North Carolina, and three unknown reviewers for the valuable suggestions they have made.

¹ The strategy used in these procedures is quite flexible and can be applied to a wide variety of problems. These include the direct analysis of preference data (data in which individuals indicate which of a set of objects they prefer) and nonmetric factor analysis. However, the wide familiarity and availability of metric factor analytic procedures and technical problems which arise in the direct analysis of preference data have limited the application of this approach in these areas. In this piece we will limit our focus to more conventional nonmetric multidimensional scaling.

An Intuitive Example

Let us launch the discussion by way of a particular example. Suppose that we were interested in an individual's perception of five political candidates: McGovern, Humphrey, Nixon, Percy, and Wallace. We might ask the individual a variety of questions. For example, we might ask how much he likes each candidate, or what the good and bad points of each candidate are. One task would be particularly rewarding for the information it produces: Ask the individual to order pairs of candidates according to the degree of similarity that he perceives among them. If the individual organizes politics (and politicians) along a liberal-conservative continuum, we would expect him to perceive candidates of like ideology to be relatively similar; if he organizes politics on the basis of party affiliations, we would expect him to perceive candidates of the same party to be relatively similar; if he has an incoherent or idiosyncratic view, we would expect no recognizable pattern to emerge. Let us suppose, then, that we asked him to rank pairs of these candidates on the basis of their similarity, and he ordered them as follows:

Most similar: 1. Humphrey - McGovern
2. McGovern - Percy
3. Nixon - Wallace
4. Nixon - Percy
5. Humphrey - Percy
6. Humphrey - Nixon
7. Humphrey - Wallace
8. McGovern - Nixon
9. Percy - Wallace

Least similar: 10. McGovern - Wallace

This ranking, while of interest, does not adequately convey a sense of the manner in which the individual's perceptions of these political figures are organized. If the ranking were used to locate points representing the candidates in a visualizable space, the underlying structure of his perceptions might be revealed more clearly.

In a technical sense, locating objects in a representational space involves assigning to each object a series of numbers, one number for each of the dimensions in the space. The numbers position the object points on the dimensions. If one dimension is involved, one number is assigned to each object. If the space is two-dimensional, two numbers are associated with each object. The first number positions the point on the first axis; the second number positions the point on the second axis. If three dimensions are involved, three numbers are assigned to each object, and so on.

In order to assign these numbers in a rational way, it is necessary to associate a mathematically meaningful property with the observed data. For example, think of creating an ordinal scale of objects on the basis of their physical weight. The real number properties of *greater*, *equal*, and *less than* can be associated respectively with an object weighing more than, the same as, or less than another object. On this basis, given any finite number of objects and a simple balance, it would be possible to assign a number to each object such that whenever one object is heavier than another and tips the balance in its direction, a higher number is assigned to it.

What mathematically meaningful property might be associated with our pairwise ranking of candidates? It does seem that *the more similar a pair of candidates are perceived as being, the closer the points representing these candidates should be when they are located in the space*. With this as a criterion we can approach the problem of scaling these points.²

Can these points scale in one dimension? The most dissimilar pair is McGovern and Wallace; hence on a single dimension these will have to lie at opposite ends of the scale. Humphrey-McGovern are the most similar pair; hence they must lie next to each other on the scale. This situation is depicted in Figure 1. Now consider the location of the Percy point. Percy is more similar to McGovern than to Humphrey; hence his point should be closer to McGovern's than to Humphrey's. This creates a problem. We have already decided that Wallace and McGovern must lie at opposite ends of the scale; hence Percy cannot be located at P_1 or any point to the left of McGovern. We also decided that Humphrey must be immediately next to McGovern; hence Percy cannot be located at P_2 or any point between McGovern and Humphrey. If we were to try to locate Percy at the Humphrey point or anywhere

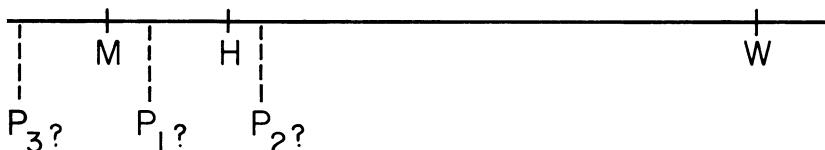


FIGURE 1
Illustration of the Impossibility of a Unidimensional Representation of the Data

² This assumption appears quite reasonable; however, it is by no means the only, or necessarily the "correct," assumption to apply to these observations. Its usefulness depends on the degree to which the individual's psychic perception of proximity corresponds to physical proximity. Nevertheless, unless strong reasons dictate the assumption is *not* valid, it is an excellent choice.

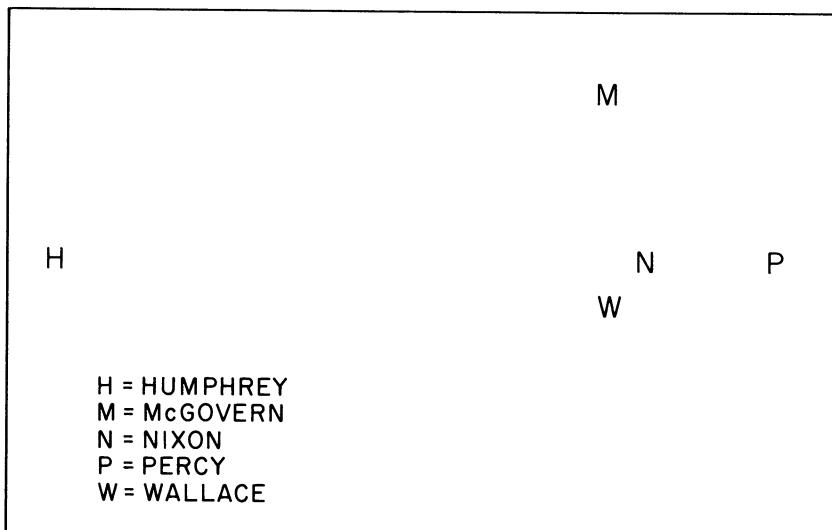


FIGURE 2
 A Random Configuration of the Five Points

to the right of Humphrey (i.e., at P_3), the Percy point would be closer to Humphrey than to McGovern, which reverses the similarity order. With the McGovern, Humphrey, and Wallace points positioned at acceptable locations, there is no adequate place for the Percy point; thus no perfect representation of these similarities is possible on a single dimension.

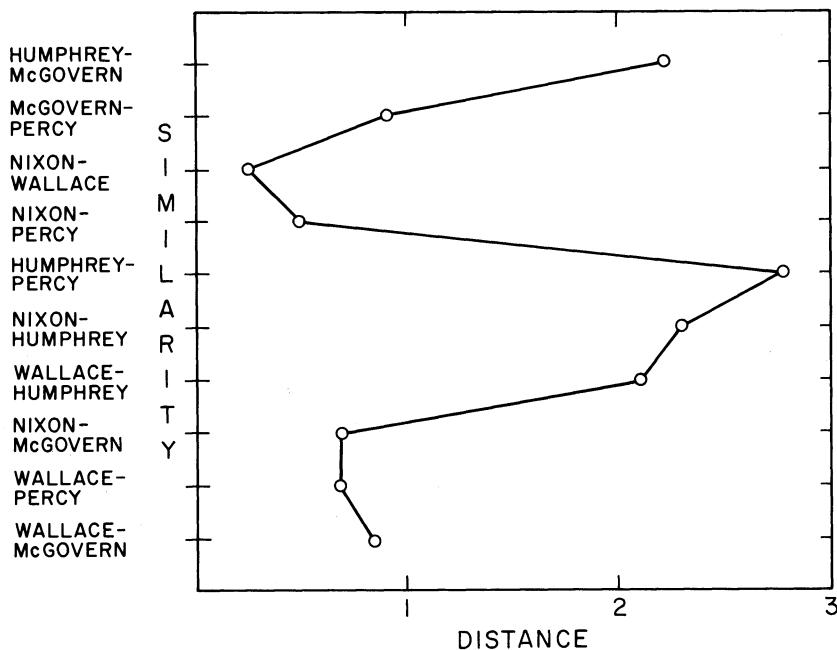
Can we scale the points in two dimensions? Clearly, this is more complicated. Let us start with a rather arbitrary decision. We will initiate the process by randomly locating five points in the space, one point for each candidate.³ In Figure 2 the randomly generated configuration of points is displayed. Obviously, this configuration does not even approach satisfying the scaling criterion. The Humphrey and McGovern points which should be closest are quite far apart and generally the points are scrambled incoherently.

How might we proceed? A naive, but reasonable approach would be to move the points in the space so that the distances between them are more

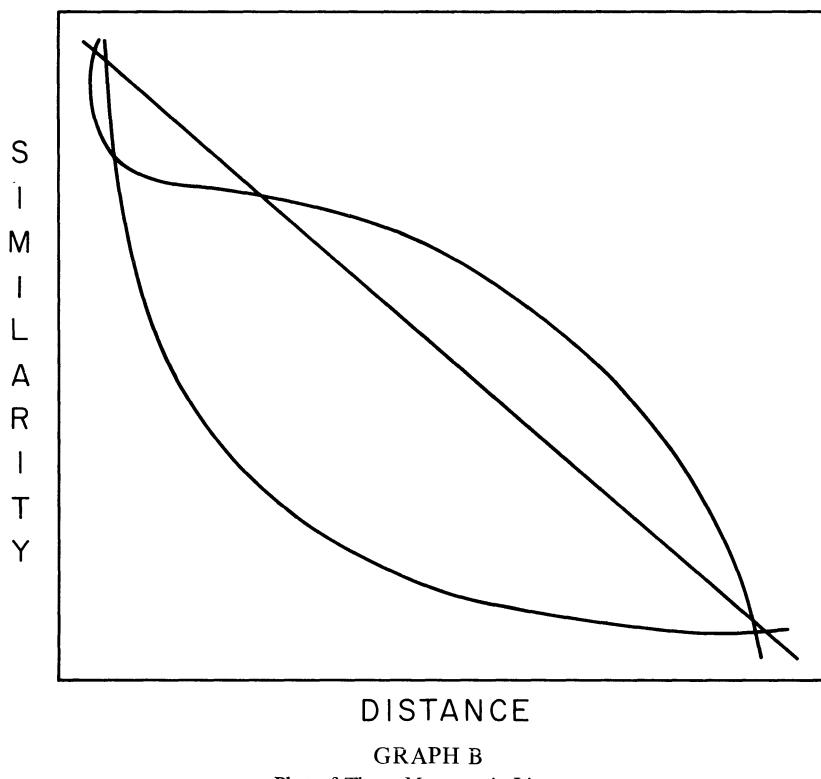
³The process of placing the points can be likened to a dart-thrower tossing five labeled darts, each aimed at the same bull's-eye. The location of each dart would depend only on the chance fluctuation of the throw. Similarly, the location of the five points is entirely arbitrary and depends only on chance factors.

consistent with the similarities. However, it is not entirely clear how the points should be moved. For example, if we moved the Humphrey and McGovern points very close together, it might bring McGovern too close to Nixon and Wallace, it might take McGovern too far from Percy, or somehow violate the relation of Humphrey to the other candidates. Ideally, when moving each point, we would take into consideration all of its interrelations. One reasonable tactic would be to generate a set of target values, one for each pair, which could be used to guide the moves.

Let us draw a graph. Each point on the graph will represent a pair of candidates. Location on the Y axis will be determined by the rank order similarity of the pair, and position on the X axis by the interpoint distance in the scaling space. It is important to distinguish between the *scaling space* and the *graph*. The *scaling space* appears in Figure 2; it is the space in which the candidate points are represented. It contains five points, one for each candidate, and is two-dimensional, but could be of one, three, four, or any finite



GRAPH A
Plot of Distance Against Similarity—Random Configuration



dimensionality. It is the end product of the scaling procedure. The *graph* has ten points, one for each pair of candidates, is *always* two-dimensional (y = similarity, x = distance), and is useful in creating the target distances. The graph associated with the scaling space in Figure 2 appears in Graph A.

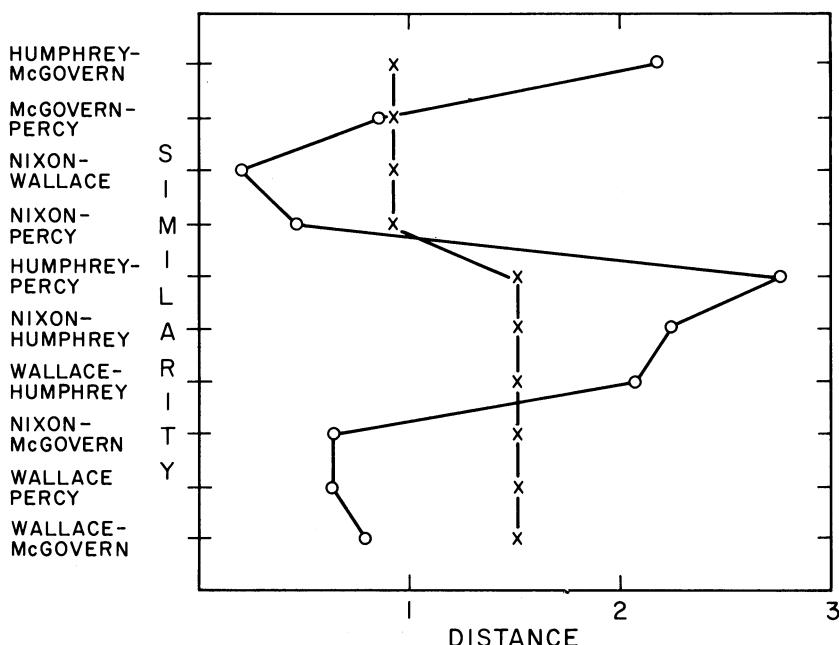
What would this graph look like had our scaling effort been successful? Our scaling objective is to locate the points so that the distance between pairs of points increases as pairs are perceived to be less similar. Had we successfully located the candidate points, when we connected the ten points on the graph the resulting line would have moved down and to the right, indicative of increasing distance as we moved from the most to the least similar pair.

Lines which move consistently in one direction are called monotonic lines. In Graph B three lines appear. Notice, all three lines are monotonic—they continually move to the right, indicative of increasing distance as one moves

down from the most to the least similar pair. However, the three lines are quite different: one line is almost straight, another a smooth hyperbolic curve, and the third slightly S-shaped. That each of the lines satisfies the scaling criterion illustrates the nonmetric nature of the scaling goal. If the procedure were metric, there would be an exact relationship between the similarity measure and the interpoint distances. For example, if we insisted that our interpoint distances be a linear function of the similarities, then only straight lines would be acceptable. Algebraically, insisting on a linear function requires that for every unit change in the similarity measure there be a fixed change in interpoint distance. Any such assumption presupposes that the units of the similarity measure are meaningful. Since in this case we started with only the rank order of pairwise similarity, making this assumption would be inappropriate. In general, by allowing for *any monotonic relation* we allow greater flexibility in our attempts to locate points in the scaling space and make only ordinal assumptions about the measure of interpair similarity.

Now let us look back at the graph associated with Figure 2. Quite clearly, the line in Graph A is not moving uniformly down and to the right. In Graph C this line reappears along with another line. The second line is constructed to be as similar to the first line as possible (in a least squares sense) with the restriction that it never move down and to the left. Using the more technical vocabulary, the distance values on the second line are restricted to be *monotonically* decreasing with (as a function of) perceived similarity. Therefore, as one moves from more to less similar pairs, the distance associated with each successive pair on the graph never decreases. This second line will always move either straight down or down and to the right. (In the literature, the graph in which both the actual and monotonic distances appear is called a *Shepard Diagram*.)

Our object in the scaling process is to have distances between pairs decrease monotonically as pairs are perceived to be more similar. Our object in constructing targets is to generate a set of values which can guide us when we move the points in the scaling space. Notice that when the first line is zigzagging, not at all satisfying the original scaling criterion, the second line runs straight down and seems to be a "smoothed out" version of the first line which is consistent with the scaling criterion. Clearly, it would be a considerable improvement of the configuration which appears in Figure 2, if the actual interpoint distances were more like the values associated with the second of these two lines. Hence, these second values will make excellent targets, and we shall use them for that purpose. (In the literature, target values are called *disparities*. They are also sometimes identified as d -hats, \hat{d} , or d -stars, d^* .)



GRAPH C

Plot of Distance Against Similarity for Actual Distances (O's) and Target Values (X's) - Random Configuration

The numerical values of the distances and targets appear in Table 1. The strategy for calculating the targets is to set them equal to the actual distances except when violations of the monotonicity requirement occur. When violations do occur, targets are calculated by averaging as few actual distances as possible to resolve the violation. Note that when a series of target values are the same, they are simply the average of the actual distances over the series of pairs. Hence, a pair whose actual distance is satisfactory will tend to have a target value equal to the actual distance; a pair whose actual distance is smaller than that of pairs which are perceived to be more similar will tend to have a target value larger than its actual distance; and a pair whose actual distance is larger than that of pairs which are perceived to be less similar will tend to have a target value smaller than its actual distance. A detailed example showing how these target values are calculated appears in Appendix A.⁴

⁴This least squares method is not the only way we could calculate targets; it is,

The strategy for moving the points is exceedingly straightforward. If two points are farther apart than they should be—that is, if the actual distance for the pair is greater than the target value—the points will be moved closer together; if the points are closer together than they should be, they will be moved farther apart. The relative amount of movement will be determined by the difference between the actual and target distances. The greater the difference, the more the two points will be moved. Points will always be moved directly toward or directly away from each other—that is, they will be moved along the line linking the two points.

Generally, when the points are moved the new actual distances will *not* equal the target values. It would be possible to pick any *single* distance, such as the Humphrey-McGovern distance, and make that actual distance equal to its target; it is ordinarily not possible to make *every* distance correspond. The targets provide a *guide* for moving the points. Our goal in moving the points is only to improve the configuration. Once we have succeeded in improving the solution by making the actual distances more like these targets, we can then calculate new targets and repeat the entire procedure in the hope of further improving the configuration. Eventually, we should obtain a solution in

TABLE 1
Actual Distances and Target Values Corresponding
with the Initial Configuration Which Appears
in Figure 2

Pair	Actual Distance	Target Value
Humphrey - McGovern	2.218	0.961
McGovern - Percy	0.909	0.961
Nixon - Wallace	0.239	0.961
Nixon - Percy	0.479	0.961
Humphrey - Percy	2.766	1.554
Nixon - Humphrey	2.288	1.554
Wallace - Humphrey	2.130	1.554
Nixon - McGovern	0.662	1.554
Wallace - Percy	0.666	1.554
Wallace - McGovern	0.815	1.554

however, a particularly reasonable approach. Any method we use would have to share two features with it: first, the target values must be based on the actual distances; second, the target values must be a monotonic function of the original rank order of pairs. Another common method is the rank image method. An example using this method appears in Appendix B.

which the actual distances are a monotonic function of similarity, if such a solution is possible.

Let us now go down the list of Table 1, moving pairs of points. The first pair, Humphrey-McGovern, should be moved toward each other and moved quite a lot since the difference between the actual and target distances is considerable. The next pair, McGovern-Percy, should be moved slightly apart, since the target distance is a little larger than the actual distance. This is the second move for the McGovern point, and in general, each point will be

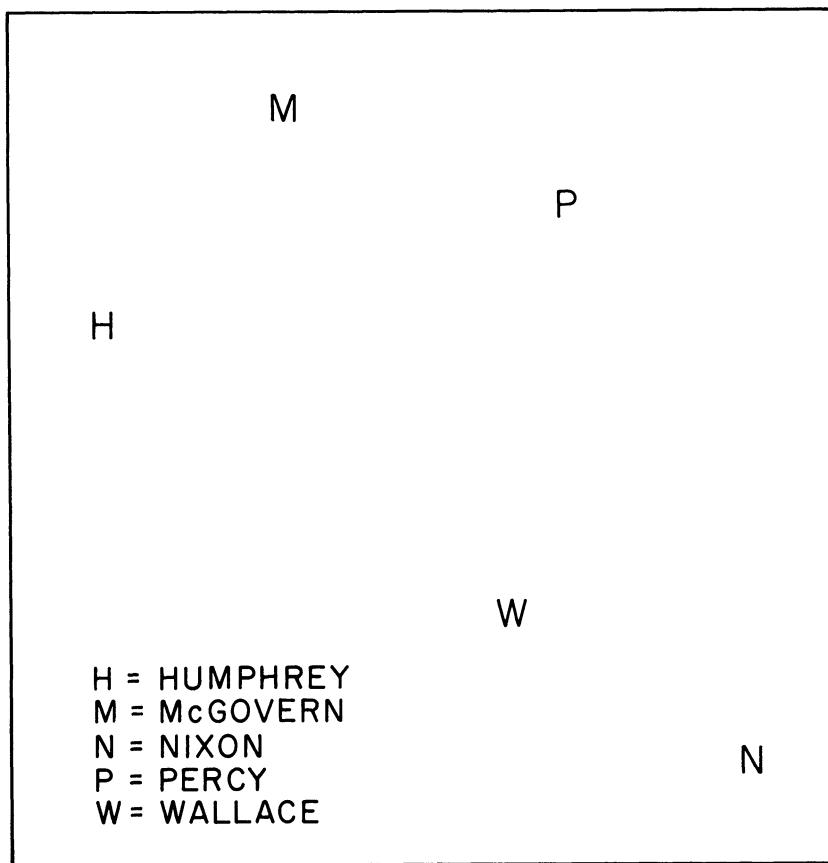


FIGURE 3
The Configuration of Five Points after the First Set of Moves

TABLE 2

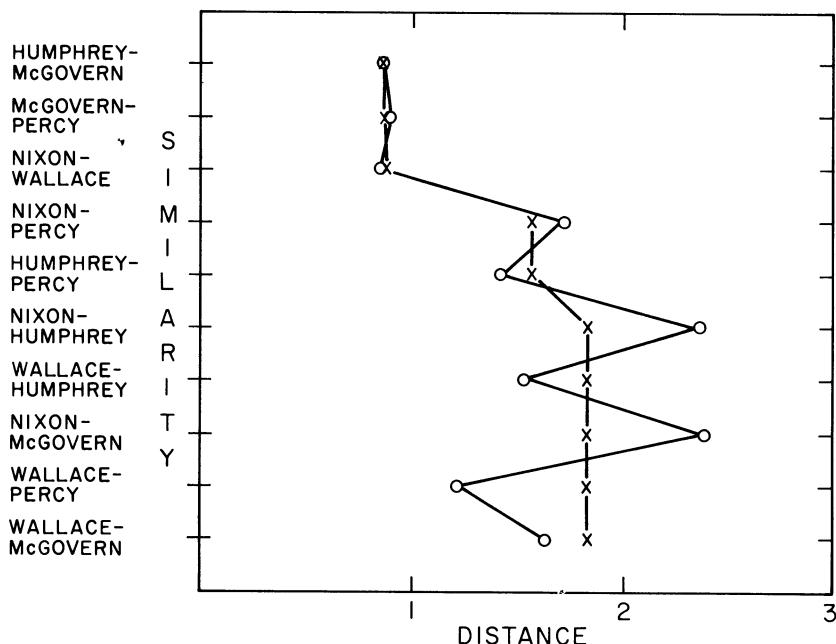
Actual Distances and Target Values Corresponding with Configuration after the First Iteration Which Appears in Figure 3

Pair	Actual Distance	Target Value
Humphrey - McGovern	0.856	0.856
McGovern - Percy	0.864	0.860
Nixon - Wallace	0.856	0.860
Nixon - Percy	1.739	1.588
Humphrey - Percy	1.437	1.588
Nixon - Humphrey	2.366	1.822
Wallace - Humphrey	1.511	1.822
Nixon - McGovern	2.381	1.822
Wallace - Percy	1.218	1.822
Wallace - McGovern	1.634	1.822

moved several times in the course of any particular sequence. After making the ten moves indicated on the list, we would have a new configuration of points. Such a configuration appears in Figure 3. Clearly, this is a considerable improvement over the configuration which appeared in Figure 2. However, it is still far from perfect. For example, Nixon and Percy are farther apart than Percy and Wallace, while they should be closer together. Once more, we can move the points and try to further improve the configuration.

The distances and target values associated with the pairs of points as they are located in Figure 3 appear in Table 2 and are plotted in Graph D. The new target values are again calculated to be as similar to the new distances as possible while still satisfying the monotonicity requirement. Notice that the new target values are much closer to the new distances than were the previous targets to the previous distances. If the solution were "perfect," the target values would be identical to the distances.

We can now proceed to move the points, making one modification in our strategy this time. While we again will move the points in relation to the difference between the actual and target distances, we will in general move the points less, since we are interested in changing the configuration less radically. (In the nonmetric multidimensional scaling literature, the amount a configuration is changed from iteration to iteration is called the *step-size*.) The configuration of points recovered after this second move is displayed in Figure 4. The configuration is better than the previous one, but still not perfect. Clearly, we could continue to repeat the process indefinitely in the



GRAPH D

Plot of Distance Against Similarity for Actual Distances (O's) and Target Values (X's)—Configuration after First Set of Moves

hope of improving the configuration each time. We would stop when the actual distances and target values are in exact correspondence, indicating that the solution is perfect, or should that never occur, when we succumb to general fatigue. The configuration recovered after three additional iterations appears in Figure 5, and in Table 3 are the interpoint distances and target values associated with Figures 4 and 5. The final configuration displayed in Figure 5 is perfect; the distances consistently increase from the most similar Humphrey-McGovern pair to the least similar Wallace-McGovern pair.

The two-dimensional solution in Figure 5 gives us an image of the way in which the individual perceived these five candidates. It seems that both party identification and left-right ideology might have influenced his perceptions. We will postpone further discussion of the configuration until a later section when we will consider both the general interpretation problem and the interpretation of this configuration in more detail.

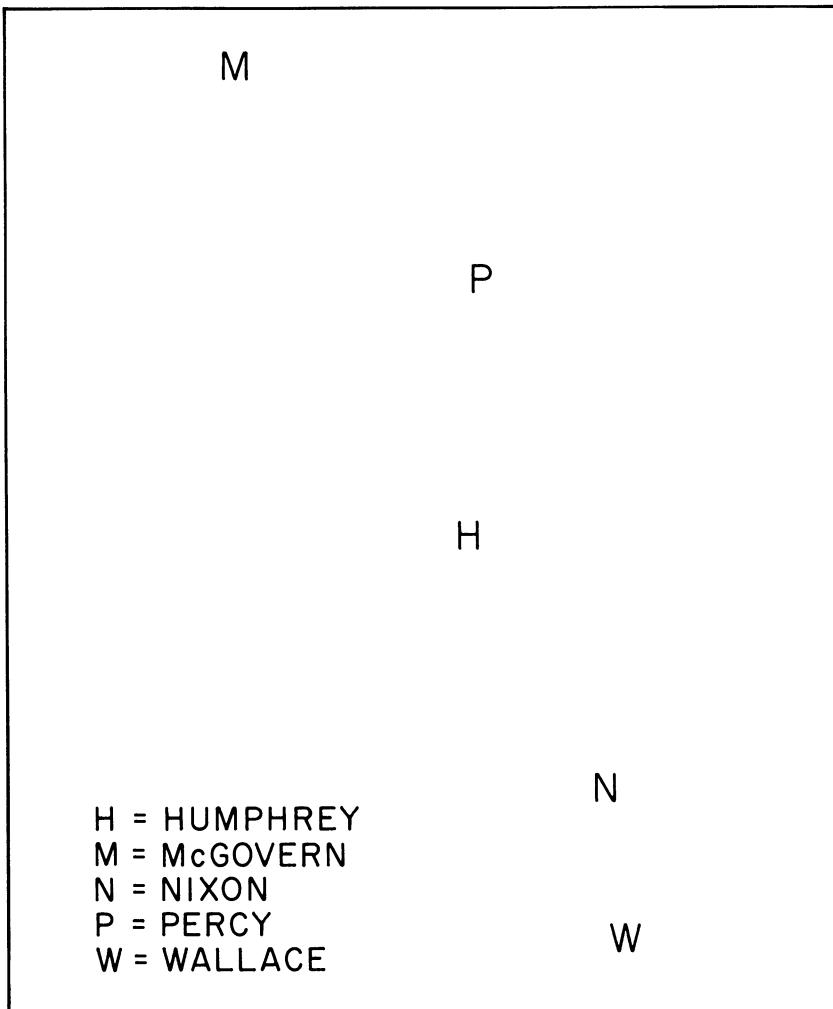


FIGURE 4
The Configuration of Five Points after the Second Set of Moves

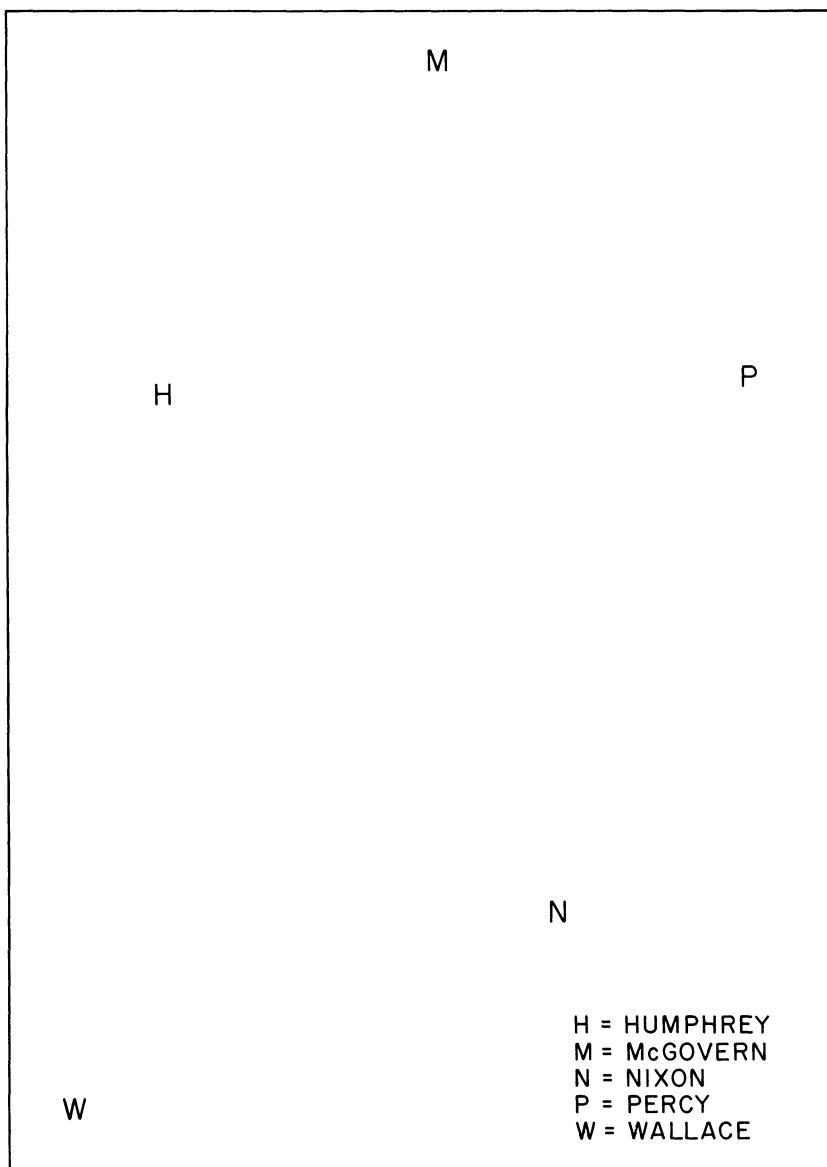


FIGURE 5

The Final Configuration of Five Points Obtained after the Fifth Set of Moves

TABLE 3

Actual Distances and Target Values Corresponding to the Configurations Displayed in Figures 4 and 5

Pair	Figure 4		Figure 5	
	Actual	Target	Actual	Target
Humphrey-McGovern	1.501	.954	.937	.937
McGovern-Percy	.914	.954	.973	.973
Nixon-Wallace	.447	.954	1.141	1.141
Nixon-Percy	1.503	1.021	1.257	1.257
Humphrey-Percy	.742	1.021	1.283	1.283
Nixon-Humphrey	.818	1.021	1.408	1.408
Wallace-Humphrey	1.243	1.243	1.568	1.568
Nixon-McGovern	2.318	2.133	1.876	1.876
Wallace-Percy	1.948	2.133	2.195	2.195
Wallace-McGovern	2.741	2.741	2.421	2.421

The Analytic Procedure

While the procedure we used to locate the points seemed successful in this five-candidate example, it had two very critical drawbacks. First, there was no measure available to indicate the adequacy of the solution. Second, unless we achieved a perfect solution, there was no way of knowing when we could stop the procedure. Both of these would constitute serious problems in more complex and realistic situations. The motivation for performing nonmetric multidimensional scaling procedures is to help elucidate the structure in a complex data matrix. Even when such matrices are quite structured, idiosyncratic features are almost invariably present and mitigate against perfect solutions. Realistically, our goal is not to find a perfect configuration, but rather the best one possible. It is therefore essential to confront the problems of (a) defining a measure which reflects the adequacy of any particular solution and (b) delineating a method for determining when the iterative procedure should be terminated.

The strategy used to move the points involved two basic steps. First, given the set of similarities and the interpoint distances in the scaling space, the target values were calculated. Second, the points were moved in order to make the actual distances more like the target values. The object of the scaling process is to have the actual distances a monotonic function of the

original similarities. The target values were created to be as similar to the distances as possible, with the restriction that they be a monotonic function of the original similarities. A natural measure of how close the solution is to satisfying the monotonicity goal would be the average (or mean) squared difference between the actual distance and the target values. The smaller the average difference, the better the solution would be. Formula A represents the mean squared difference.

Formula A

$$\sum_{i=1}^{\text{\# of pairs}} \frac{[(\text{Actual distance pair } i) - (\text{Target value pair } i)]^2}{\text{\# of pairs}}$$

It should be apparent that the formula is sensible. If the actual distances were a monotonic function of the similarities, then the actual distances and target values for each pair would be identical, and all the differences would be zero. As the actual distances depart from the monotonicity goal, the actual distances and target values will diverge and the differences will increase. Since the differences are squared, their signs will not influence the measure.

There is, however, a major problem with this measure; it is quite sensitive to the scale factor used to calculate distances. No basic unit of measurement is present when one performs nonmetric multidimensional scaling. All that is critical is the relative distance between points. Were we, for example, to double every interpoint distance, the solution would be neither better nor worse than before. However, as a result of doubling, our measure would change dramatically. The actual distances are doubled; hence the targets would double. Since both target values and actual distances are doubled, their differences would double. Based on squared differences, the measure would quadruple.

Formula B

$$\frac{\sum_{i=1}^{\text{\# of pairs}} \frac{[(\text{Actual distance pair } i) - (\text{Target value pair } i)]^2}{\text{\# of pairs}}}{\sum_{i=1}^{\text{\# of pairs}} \frac{[(\text{Actual distance pair } i) - (\text{Mean actual distance over all pairs})]^2}{\text{\# of pairs}}}$$

This problem can be alleviated by adding an appropriate denominator to the measure. Formula B modifies Formula A by dividing it by the variance of the interpoint distances. Should the scale factor be changed, both the numerator and denominator would be affected equally, and hence the effects would cancel out, leaving the measure unchanged.

The revised measure can be interpreted as the proportion of the total variance in interpoint distances *inconsistent* with the monotonicity constraint. It is similar to the ratio of unexplained to total variance central to regression and analysis of variance. This measure will vary between 0 and 1. For perfect solutions it will be zero; as solutions depart from the monotonicity constraint, its value will increase to a maximum of one.

In keeping with statistical tradition, we might make one more change in the measure. Generally, one does not work directly with variances or squared correlations, but rather with their square roots, standard deviations and simple correlations. Hence, we shall define a final measure which is the square root of the previous one. This measure appears in Formula C and is rewritten in Formula C'. In Formula C' the “# of pairs” term which appears in the numerator and denominator simply cancels out and is removed.

Formula C

$$\left[\frac{\sum_{i=1}^{\# \text{ of pairs}} \frac{[(\text{Actual distance pair } i) - (\text{Target value pair } i)]^2}{\# \text{ of pairs}}}{\sum_{i=1}^{\# \text{ of pairs}} \frac{[(\text{Actual distance pair } i) - (\text{Mean actual distance over all pairs})]^2}{\# \text{ of pairs}}} \right]^{\frac{1}{2}}$$

Formula C'

$$\left[\frac{\sum_{i=1}^{\# \text{ of pairs}} \frac{[(\text{Actual distance pair } i) - (\text{Target value pair } i)]^2}{\# \text{ of pairs}}}{\sum_{i=1}^{\# \text{ of pairs}} \frac{[(\text{Actual distance pair } i) - (\text{Mean actual distance over all pairs})]^2}{\# \text{ of pairs}}} \right]^{\frac{1}{2}}$$

The final measure will also vary from 0 for perfect solutions to a maximum of 1.

In the nonmetric multidimensional scaling literature this measure is called *Stress 2*. The term “Stress” follows from the fact that the measure increases

as the solution gets worse, the "2" from the fact that it was the second of two Stress formulas developed.⁵

Given a measure of the badness of fit of a particular solution, the problem of determining when to terminate the iterative procedure can be easily handled. Generally, we would expect that each successive solution would be closer to satisfying the monotonicity requirement than the solution which preceded it. Hence, if the Stress does not improve over several iterations, the procedure can be reasonably terminated.

We now have enough information to delineate a complete procedure.

1. Initially, locate the points, one for each of the objects to be scaled, in a space of fixed dimensionality.
2. On the basis of the interpoint distances and the similarities, construct a set of target values.
3. Calculate the Stress of the solution.
4. If the solution is perfect or if the basic procedure has been repeated several times (undergone several iterations) and the Stress has not improved, stop the entire procedure. Otherwise, go on to step 5.
5. Move the points in order to make the actual distances more similar to the target values, with the magnitude of the move adjusted to reflect the desired amount of change.
6. Repeat the process starting at step 2.

⁵ The original Stress formula, *Stress 1*, differs only in the denominator. In Stress 1 the variance of distances term is replaced by the mean square distance. The formula for Stress 1 is:

$$\left[\frac{\sum_{i=1}^{\text{# of pairs}} [(\text{Actual distance pair } i) - (\text{Target value pair } i)]^2}{\sum_{i=1}^{\text{# of pairs}} (\text{Actual distance pair } i)^2} \right]^{\frac{1}{2}}$$

Stress 1, like Stress 2, is sensitive to the difference between the actual distances and target values and insensitive to the scale factor. The Stress 1 denominator will always be greater than or equal to that for Stress 2. Hence, Stress values will be lower using Stress 1 than Stress 2. While Stress 2 has a more intuitive interpretation and provides certain technical advantages over Stress 1, both measures perform well as measures of badness of fit. In the text, when we refer to the Stress of a solution we will mean the Stress calculated according to any satisfactory measure of badness of fit, such as Stress 2 or Stress 1.

The procedure seems quite reasonable. However, its development has not been very rigorous. It would be useful to approach the task of scaling the points in a more analytic fashion. Underlying much of the procedure is the notion that the lower the Stress, the better the solution. Analytically, we could define the problem as one of finding the configuration of points for which the Stress value is minimized.

In minimizing such complex functions mathematicians use what are called *negative gradient* or *steepest descent* techniques. These methods work as follows:

1. One determines how much the measure will improve for a very small change in each of the independent variables. (The independent variables in the Stress formula are the locations of the points on the axes. Each coordinate location is something that we can manipulate. Hence, if we have 5 points to locate in 2 dimensions, we have $5 \times 2 = 10$ independent variables). Technically, the amount the function improves with a very small change of an independent variable is called the *partial derivative* of the function with respect to that variable.
2. One then changes the values of each independent variable in proportion to the relative amount of improvement; variables are changed more if the improvement (partial derivative) calculated was large. One then repeats the procedure iteratively until a change in any of the independent variables will cause no (or negligible) improvement. This occurs when all the partial derivatives are zero or close to zero.⁶

Analytically, we can solve the problem of minimizing Stress by calculating the partial derivatives and using them to relocate the points. When one goes through the analytic solution, it is remarkably like the intuitive procedure we developed earlier. That is, it turns out that points are always moved directly toward or away from each other with the relative amount of movement determined by the difference between the actual and target distances. However, placing the problem in an analytic context is still useful. First, we know

⁶ If we list these partial derivatives in order, starting with that associated with the first point on the first dimension and continuing to the last point on the last dimension, the resulting set of ordered partial derivatives is called the *gradient*. The gradient is a vector whose direction summarizes the direction in which the points are moved. The size of the gradient is the square root of the sum of the squared partial derivatives, and hence offers some information about the rate of change for a very small movement of the points. It is easier to technically discuss some aspects of the procedure in terms of the gradient; hence this vector is regularly referenced in the literature.

that our intuitive solution did not miss an alternative approach which would have been easier and more accurate. Second, when we moved the points in the intuitive procedure we worked pair by pair. In the intuitive procedure each point was moved several times in the course of a single iteration; these moves may have been in the same or in quite different directions. The partial derivatives for each point on each axis represent the sum of all the pairwise moves. Hence, using the partial derivatives is slightly more efficient.

Third, our only stopping criterion occurred when the Stress failed to improve after several iterations. Now when the partial derivatives are all zero, we know all our moves cancel out and we have reached some kind of optimum solution. Fourth, we never had a satisfactory way to decide how much to change a configuration on any given iteration. It happens that if we change our solution just enough so that our last move gives us no information about our next move, we are changing our configuration just the right amount. While this is a little technical, it nevertheless is useful to know that once we have broached the problem analytically we have a better notion of how radically we should change the configuration from iteration to iteration.

Comparison with Factor Analysis

The procedure is well defined and can be programmed to run on a computer. Once programmed, it can easily be applied to problems of substantive interest. Since factor analysis has been the most widely used multidimensional scaling procedure, an obvious question is "how do these methods compare?"⁷

Obviously, they differ in terms of the level of measurement they presume; factor analysis assuming interval level data, nonmetric multidimensional scaling, ordinal level data. However, this distinction is somewhat artificial. It is feasible to construct a nonmetric factor analysis procedure or a metric multidimensional scaling technique. There is a more fundamental difference between the two methods.

Most critically, the basic models underlying the methods are different. Factor analysis is based on a *scalar product* model; nonmetric multidimensional scaling is based on a *distance* model. While this is a rather technical sounding distinction, it is of some significance. In Figure 6 two points, X and

⁷A full discussion of factor analysis is beyond the intended scope of this paper. In this section the most critical difference between factor analysis and multidimensional scaling is emphasized. Because the discussion is brief, it is necessarily somewhat more technical than those sections which precede and follow it.

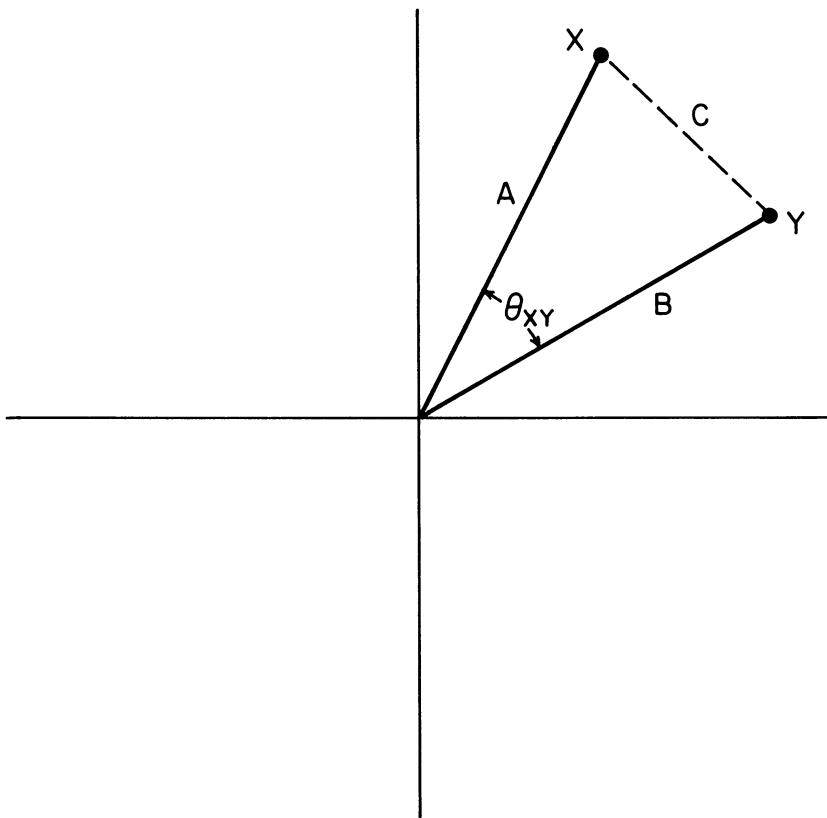


FIGURE 6
Basic Scalar Products with Points Equidistant from Origin

Y, appear. The length of the line labeled C is the distance between the two points. The length of the line labeled A is the distance of point X from the origin. In keeping with standard notation, we will denote this distance $|X|$. In general, the distance of a point from the origin will be denoted by the symbol representing the point surrounded by two vertical lines. The length of the line labeled B is the distance of the Y point from the origin or $|Y|$. The angle made by the intersection of these two lines we will denote as θ_{XY} . The scalar product between X and Y is equal to the length of X from the origin times the length of Y from the origin times the cosine of the angle between them; that is, $|X| |Y| \cos \theta_{XY}$.

How is this scalar product related to the interpoint distance? When the points are all equidistant from the origin, the distance and the scalar product will depend only on the angular separation. As the angle becomes larger, the distance will increase and the scalar product will decrease. Under this condition there is an inverse monotonic relation between distances and scalar products. The larger the distance, the smaller the scalar product and vice versa.

In instances where the distances of the points from the origin are not equal, this relationship breaks down. In Figure 7 we see three points, X, Y, and Z. Here $|X| = |Y|$, but $|Z|$ is quite a bit larger than $|X|$ and $|Y|$. The largest

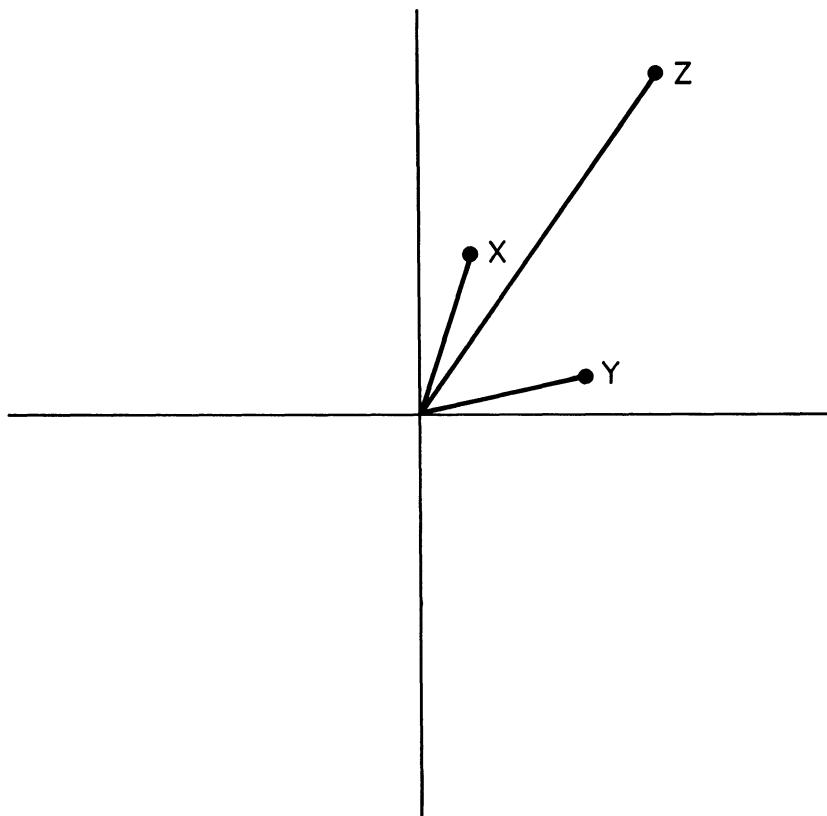


FIGURE 7
Illustration Where Scalar Products Are Not Monotonically Related to
Interpoint Distance

scalar product will be between X and Z, since $|Z|$ is large, and the angle between X and Z is small. The next largest scalar product will be between Y and Z; $|Y| |Z| = |X| |Z|$, but the angle between Y and Z is greater than that between X and Z. The smallest of the scalar products will be between X and Y, since $|X|$ and $|Y|$ are both small and the angle between them is the largest of all. Now look at the distances; it is obvious from inspection that the XY distance is the smallest, the XZ distance next, and the YZ distance the largest. Here, there is no monotonic relation between distance and scalar product; the pair with the smallest scalar product has the smallest distance, the pair with the largest scalar product has the middle distance, and the pair with the middle scalar product has the largest distance. In general then, when all points are equidistant from the origin, the scalar product is monotonically related to the distance; when points are not equidistant from the origin, the scalar product is not monotonically related to the distance.

The basic difference between factor analysis and nonmetric multidimensional scaling should now be clearer. Factor analytic procedures treat input data as scalar products; a "perfect" configuration following a factor analysis is one in which the interpoint *scalar products* match the values in the input data matrix. In contrast, a "perfect" configuration following a nonmetric multidimensional scaling analysis is one in which the *interpoint distances* match the input data values.

Both covariance matrices and Pearson product-moment correlation matrices are scalar product matrices. However, Pearson product-moment correlations are scalar products between standardized variables, hence between variables of identical length. Either factor analysis, which is directly a scalar product model, or nonmetric multidimensional scaling, which requires data to be a monotonic function of distance, can legitimately be used to analyze product-moment correlation matrices. The standard nonparametric measures of association are conceptually similar to standardized scalar products; hence they lend themselves to either nonmetric multidimensional scaling analysis or (nonmetric) factor analysis.

How similar will the results be using these methods? Since the models underlying the two methods differ, we would in general expect them to produce somewhat different results. The principal component variant of factor analysis keeps all variables of constant length and should produce results more similar to those obtained using nonmetric multidimensional scaling. Since nonmetric multidimensional scaling procedures are nonmetric, which implies less restrictive criteria of fit, they usually achieve solutions in the same or lower dimensionalities than metric factor analysis or principal component analysis.

In many instances we collect similarities data which are not scalar product measures, but are measures the analyst wishes to treat as a monotonic function of distance. For example, a measure of internation hostility or internation trade, sociometric choice, individual perceptions of interobject similarities (such as in the five-candidate example), and the line-of-sight measure of pairwise similarity are all proximity rather than scalar product measures.⁸ In these instances the nonmetric multidimensional scaling model is correct to apply, and the factor analytic model is not.

Which method should be chosen to analyze a particular set of data? As a general rule, factor analytic procedures should be used when one is creating standard indices, such as an efficacy or civil rights index, or if one is interested in testing the dimensionality or viability of items for forming potential indices. This type of scaling is metric and is usually consistent with the linear model underlying classic factor analysis. Nonmetric multidimensional scaling methods are most useful in analyzing proximity structures such as party or candidate spaces, internation relations, and legislative voting patterns. Nevertheless, it is the measure of similarity which determines which method to apply, and an analyst should give careful thought to the substantive implication of the particular similarity measure he applies. If the measure is a proximity measure, a nonmetric multidimensional scaling method should be used. If the measure is an unstandardized scalar product measure, a factor analytic method should be used. If the measure is a standardized scalar product, it is often useful to apply both methods.⁹

Guidelines for Use

A researcher interested in using a nonmetric multidimensional scaling program should be familiar with some basic guidelines. We will now discuss those issues which an investigator confronts when he applies this methodology to research problems.

Data

Data input into a nonmetric multidimensional scaling program is usually a matrix of values reflecting the similarity or dissimilarity among pairs of

⁸ The line-of-sight measure of pairwise similarity is discussed in Rabinowitz (1973, Chapter 2). It is designed to measure the relative similarity between pairs of objects (such as political parties) from individual ratings of those objects, under the assumption that the individuals have common perception but different evaluations of the objects.

⁹ A full treatment of proximity measures is beyond the scope of this paper. See MacRae (1970), Weisberg (1968), Morrison (1972), and Rabinowitz (1973) for discussion of some of the standard measures of association and their utility for spatial analyses.

TABLE 4
Matrix of Perceived Intercandidate Similarity

	McGovern	Humphrey	Nixon	Percy	Wallace
McGovern	—				
Humphrey	1	—			
Nixon	8	6	—		
Percy	2	5	4	—	
Wallace	10	7	3	9	—

objects. In our five-candidate example we had arranged our ten candidate pairs in a list proceeding from most to least similar. We could have supplied the same information in a conventional matrix form. Such a matrix appears in Table 4. An entry in this matrix is the rank order similarity of a candidate pair. Hence, a 1 is assigned to the Humphrey-McGovern pair, a 2 to the McGovern-Percy pair, and so on through the list. When this matrix is input to a nonmetric multidimensional scaling program the numerical values will be used only to sort the pairs from most to least similar, and a list identical to the one we used in the five-candidate problem will be formed by the program. Correlation matrices contain essentially similar information. Each entry in a correlation matrix is a number reflecting the similarity between two objects, and these numbers can be used to rank the object pairs from most to least similar.

If one is working with a correlation matrix, it is necessary to be sure that items are coded consistently. For example, if we are analyzing a set of roll calls with Yea coded 1 and Nay, 3, and if on some of these votes Yea is liberal and on others Yea is conservative, all the items should be recoded so that a liberal vote receives consistently one score and a conservative vote another. Artificial negative (or positive) correlations due to question wording or coding must be avoided, since this will lead to artificially large (or small) distances between object points.

Interpretation of Results

The usual objective in performing a nonmetric multidimensional scaling analysis is to uncover the structure present in a complex data matrix. To do this an analyst must deal with three basic questions:

1. What is the correct dimensionality for the spatial representation of the data matrix?

2. How satisfactory a solution has been achieved? That is, what does any particular Stress value tell us about the solution?
3. How can the spatial configuration recovered be substantively interpreted?

Determining dimensionality. Every nonmetric multidimensional scaling solution takes place in a space of *fixed* dimensionality. The analyst first chooses the number of dimensions in which he wishes to work and then obtains a solution in that dimensionality. For example, in the five-candidate problem, we first tried a one-dimensional solution, decided a perfect one-dimensional solution was impossible, and then went on to try a separate two-dimensional solution.

The primary guide in selecting the number of dimensions in which to perform the analysis is the substantive knowledge the user brings to the problem. Invariably the analyst has an *a priori* notion concerning the number

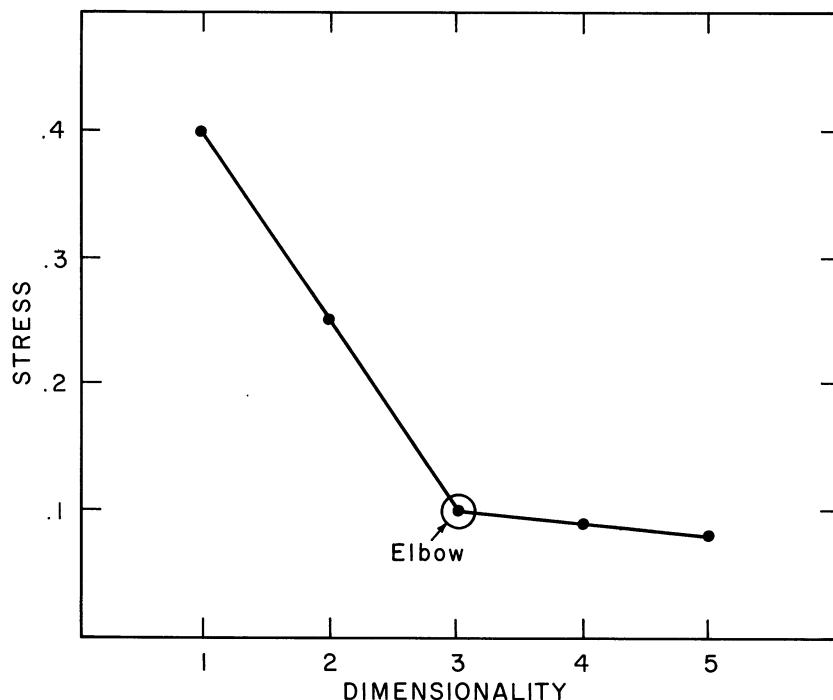


FIGURE 8
Illustrative Plot of Stress against Dimensionality with a Clear Elbow at Three Dimensions

of dimensions underlying his data matrix. If very strong substantive reasons indicate that only a specific dimensionality is appropriate, the analyst can obtain solutions in only that dimensionality. More usual is the case where the analyst feels the solution should be within a certain range of dimensions, but is not sure which dimensionality is correct. The normal procedure in this case is to obtain a solution in each of the relevant dimensionalities and to use these results to select the correct one.

Figure 8 contains a plot of Stress against dimensionality for some hypothetical set of solutions. The Stress values for the one- to five-dimensional solutions are respectively .4, .25, .1, .09, .08. Notice that Stress decreases markedly as one goes from one to two and two to three dimensions, but then seems to level off, decreasing only slightly as we add dimensions after that. We know Stress will always decrease as we add dimensions, since as we add dimensions we are increasing the number of coordinates which we will estimate and hence increasing the number of manipulable independent variables.¹⁰ In this example, it seems that the first three dimensions are each capturing a significant part of the structure present in the original data matrix, while the fourth and fifth dimensions seem to be capturing idiosyncratic features, or perhaps only random error. Notice the elbow in the curve at the third dimensional point. This elbow occurred because of the drop in the rate of change in Stress after the third dimension was added. The appearance of an elbow in the Stress against dimensionality curve provides a good indication of the correct dimensionality. In this example the dimensionality is three.

Unfortunately, in many empirical instances there is not a clearly discernible elbow. In these instances an analyst must draw on his substantive expertise to select the dimensionality he will report. When there is neither a clear substantive nor a clear empirical basis for selecting a specific dimensionality, then it is often useful to report the solution in several different dimensionalities.

Interpretation of Stress. The following guidelines are useful in relating Stress to the quality of solution.

<i>Quality of Solution</i>	<i>Stress 2</i>	<i>Stress 1</i>
Perfect	0.00–0.05	0.000–0.025
Excellent	0.05–0.10	0.025–0.050
Good	0.10–0.20	0.050–0.100
Fair	0.20–0.40	0.100–0.200
Poor	0.40–1.00	0.200–1.000

¹⁰Occasionally, Stress will be higher for a higher than for a lower dimensional

While valuable, these guidelines must be applied cautiously. Stress, ideally, would be insensitive to the number of points used in the analysis. However, as the number of points increases, Stress tends to increase. In addition, one-dimensional solutions tend to be disproportionately prone to high Stress. If one is analyzing data with a large number of objects ($N > 30$) or if one is working in one dimension, it is reasonable to extend these ranges slightly. On the other hand, if the number of points analyzed is small ($N < 10$) or if when we form the ratio of the number of points to the number of dimensions, this ratio is small (ratio < 4), the resulting Stress might be lower than the quality of solution warrants. Nevertheless, in most situations the guidelines are satisfactory.

Interpreting a Configuration: Traditional Method. When interpreting a spatial representation of data, our first instinct should be to delineate the basic cluster pattern present in the configuration. One of the great advantages of nonmetric multidimensional scaling is its tendency to produce solutions in a limited number of dimensions, thus giving us a visualizable picture of the major interrelationships among the objects in our data matrix. By getting this overview, we are much less likely to fall into those misinterpretations which arise when we are forced to work dimension by dimension, without a sense of how the entire space fits together.

In Figure 9 the final configuration of the five candidate points is redrawn with the representational axes included and some of the major groupings demarcated. We can identify two broad patterns. The regular Democrats, Humphrey and McGovern, are drawn together to the upper left side of the space; the two regular Republicans, Percy and Nixon, to the middle right. Similarly, the more liberal group of McGovern, Percy, and Humphrey are located toward the top of the space, while the more conservative group of Nixon and Wallace lies toward the bottom. The clustering pattern is weak, since the distances within clusters are quite large; nevertheless, it does seem that party and ideology effects are represented in the space.

If we want to move from this general description to an analysis of the structure in dimensional terms, we move to mathematically tenuous grounds. The problem is familiar to anyone who has studied factor analysis. In a

solution. *This result only occurs artificially*, either because the procedure was terminated prematurely, or because a false or local minimum was obtained. (In a later section the problem of local minima will be discussed.)

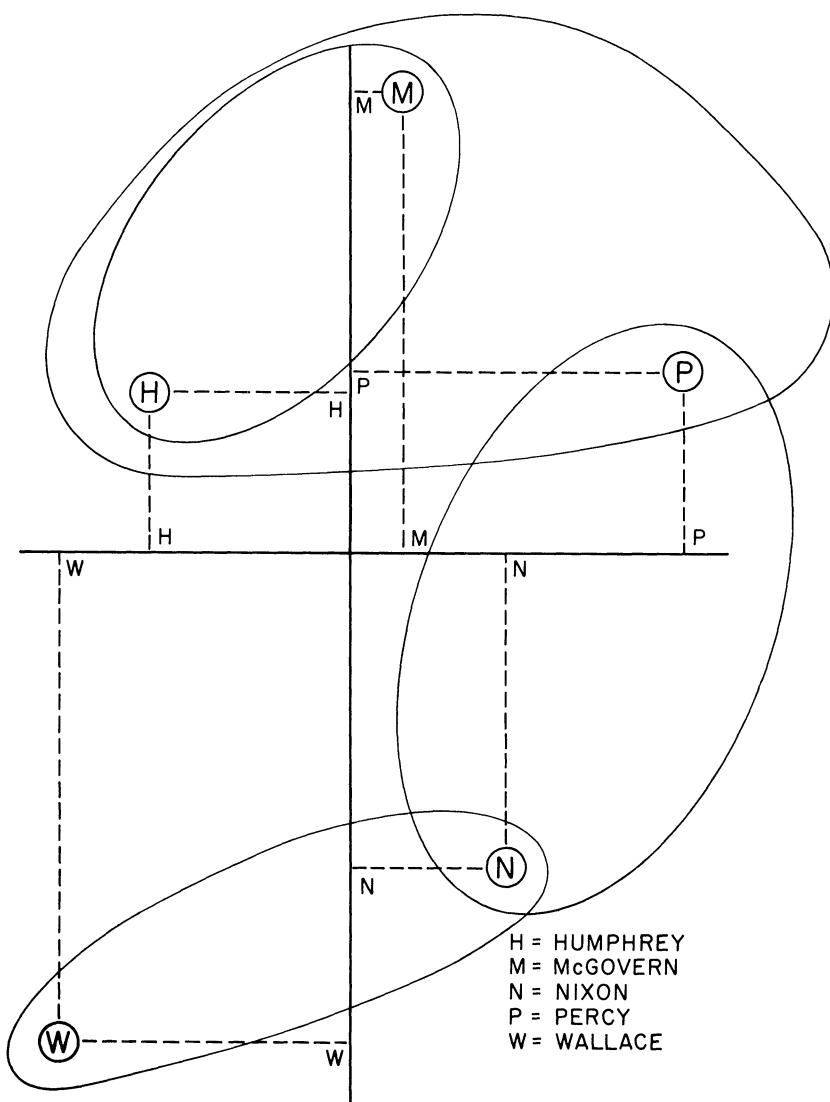


FIGURE 9
The Final Configuration of Five Points Redrawn with the Representational
Axes Included and the Major Groupings Demarked

Euclidean space, the particular axes we use as a basis on which to locate our points are entirely arbitrary. If, for example, we had rotated our axes and represented our points as in Figure 10, the solution would have identical Stress and be mathematically equivalent to our initial one. Notice, however, that the X and Y coordinates for the two solutions are quite different. When analysts interpret spatial configurations dimensionally, they do so in terms of the location of the points on specific axes. This makes the selection of the axes very critical, and puts the analyst in the somewhat awkward position of interpreting a configuration in terms of specific dimensions which are mathematically irrelevant to the solution obtained.

Factor analysts cope with this problem by defining specific *simple structure* criteria for selecting axes. The idea underlying simple structure is to have

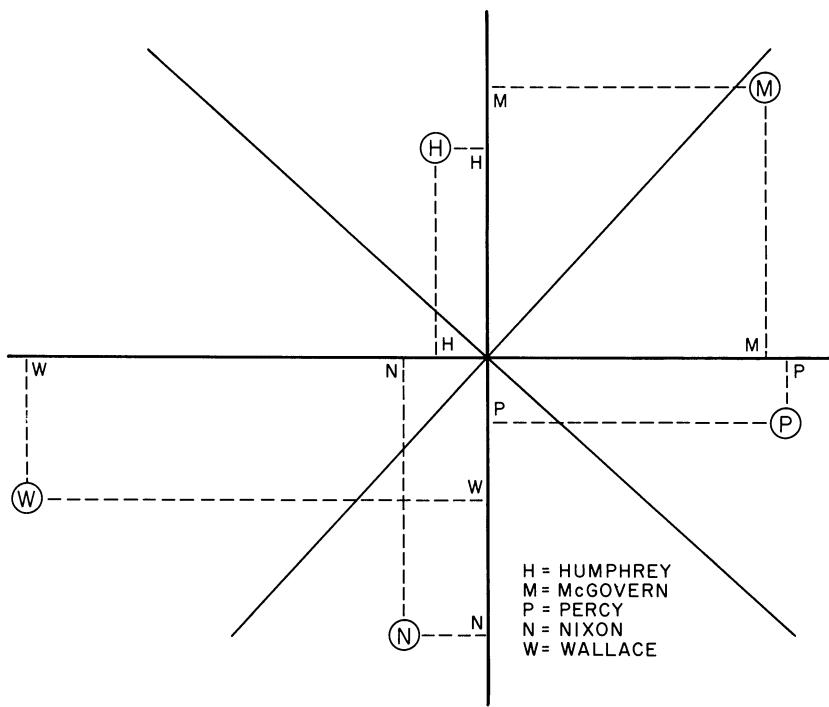


FIGURE 10
The Final Configuration Redrawn after a Rotation
(Previous representational axes are also shown.)

axes run through clusters of points. Axes are then interpreted in terms of the clusters they pass through. When clusters are well defined, this tends to encourage interpretations similar to those one would make when visually overviewing the structure. If no strong clustering pattern is present, then these rotational methods are less useful. In addition, if we are interested in interpreting dimensions as underlying factors or causes and clusters are formed because of the interrelation of several different factors, these methods produce misleading results.

The standard rotational methods used in factor analysis are divided into two types: orthogonal, in which the axes are kept at right angles and hence are independent; and oblique, in which the axes are allowed to correlate. The most common orthogonal rotation is the varimax rotation. The most common oblique rotations are the direct and indirect oblimin rotations. In general, the orthogonal rotations are most useful as a vehicle for overviewing structure, since the nonassociation between axes makes comparisons between objects easier. The oblique rotations are better suited to delineating dimensions, since "real world" attributes of legitimate substantive interest are likely to be interrelated.

It is appropriate to apply these standard rotational procedures when performing a nonmetric multidimensional scaling analysis as well. However, there is a caveat: in a nonmetric multidimensional scaling solution the origin is arbitrary. That is, the origin can be changed without altering the Stress or any essential attribute of the structure. Since rotations are around a fixed point, this could be a problem. Fortunately, the centroid of the configuration, the point in the middle of the space, is a natural origin for these solutions and is a very reasonable point about which to perform rotations. The standard nonmetric multidimensional scaling programs all locate the origin at the centroid.

It is important, of course, *not to* interpret the origin and distance from the origin as one would in a factor analysis. Factor analysis is a scalar product model; hence the origin is fixed and distance from the origin is an important piece of information. When performing a nonmetric multidimensional scaling analysis, even after a "standard rotation" interpretation of axes must rest on the *relative*, not the absolute, location of points on the dimensions. In addition, because of the nonmetric nature of the algorithm, the structure of the space will be insensitive to any factors which are common to the entire set of items. For example, in a factor analysis, the first principal component will often have high loadings for all the items when the items in the matrix tend to be positively correlated. In a nonmetric factor analysis or a nonmetric multidimensional scaling analysis this first component will not appear, be-

cause it does not discriminate between items.¹¹ *Nonmetric multidimensional scaling solutions emphasize those dimensions which lead to discrimination between items.*

Interpreting a Configuration: External Criteria. In both nonmetric multidimensional scaling and factor analysis, there are other and perhaps better ways to approach the problems of dimensional interpretation. These involve using external criteria to help select dimensions. If an investigator is seriously interested in explicating the structure among his data objects, he is likely to have an a priori sense of what the underlying dimensions are. If the research design can be controlled, it should be possible to include in the analysis both the means to test dimensional hypotheses and the means to locate axes in the space.

When we asked the individual to rank the candidate pairs according to how similar he perceived them to be, we noted that if the individual organized politics (and politicians) along a liberal-conservative continuum, he would perceive candidates of like ideology to be relatively similar; if he organized politics on the basis of party affiliation he would perceive candidates of the same party to be alike; if he had an incoherent view of politics he would display no discernible pattern in his pairwise perceptions. The two underlying dimensions which we identified as most likely to shape his political orientations were party and ideology. Let us suppose that when we asked him for his perceptions of pairwise similarities, we also asked him to locate the five candidates on two independent continua, one party, the other liberal-conservative. This produced two dimensions independent of (external to) the space created on the basis of the similarity judgments, which are displayed in Figure 11.

How might we use this external information to clarify our understanding of the similarity space? If the individual had used either party or ideology exclusively as a guide in his perceptions of candidate similarity, then the configuration recovered on the basis of the similarity judgments should have been unidimensional and should have corresponded to one or the other of the two external dimensions. If the individual had used both party and ideology, we would expect the similarity space to be two-dimensional and reflect both a party and an ideology component. The space is two-dimensional and seems to be influenced by party and ideology, but it is unclear to what extent each component influences the spatial structure and how an axis representing each component can be located in the space.

¹¹ A nonmetric factor analysis differs from a metric factor analysis in that the scalar products need only preserve the same rank order as the input data values rather than duplicate the actual values.

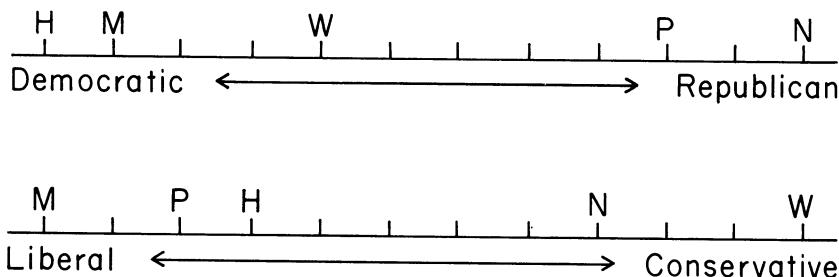


FIGURE 11
Individual's Perception of the Five Candidates on a Democratic-Republican
and a Liberal-Conservative Dimension

Inserting a new axis in the space is geometrically equivalent to drawing a line through the space running through the origin. If we are interested in seeing whether or not the spatial structure can reflect a specific external dimension, like the party or ideology dimension, it would be reasonable to determine the single axis in the space on which the projections are most like those of the externally determined dimension. If the projections on the internally drawn axis are quite similar to those on the external axis, we would have evidence which supports the hypothesis that the external dimension is one of the underlying causes of the structure; if the projections on the best fitting internal dimension and the external dimension are quite different, the hypothesis could be rejected. This raises two closely related questions: First, "How can we determine the internal axis which most closely corresponds to the external axis?" Second, "How can we measure the degree of fit between the projections on the internal and external axes?"¹²

The basic numerical information we have at the end of a nonmetric multidimensional scaling procedure is numbers which locate the object points on each of the (arbitrary) representational axes. Associated with each axis are as many numbers as there are points in the space. Conceptually, we can think of the axes as standard analytic variables. Suppose we performed an unstandardized multiple regression using as the independent variables the representational axes and using as the dependent variable the externally determined axis. We would obtain the linear combination of the representational axes which most closely corresponds with the external axis. That is, for each

¹²To clarify any ambiguities in the text, we mean by an "external axis" an axis defined completely independent of the recovered configuration, and by an "internal axis" an axis actually drawn through the space.

representational axis we will have a weight which reflects how important that axis is in predicting position on the external dimension. If we orient a new axis in the space such that it corresponds with each representational axis in proportion to the magnitude of the regression coefficient (weight), we will have the single internal axis which is most like (in a least squares sense) the external axis. The degree of fit can be measured by the Pearson product-moment correlation between the internal and external axes. This bivariate correlation is identical to the multiple correlation between the representational axes and the external dimension.

The regression equation to determine the orientation of the party axis would be

$$b_1 \text{axis}_1 + b_2 \text{axis}_2 + \text{constant} + \text{error} = \text{external party dimension},$$

where the b values are unstandardized regression coefficients, the constant and error terms are those associated with an ordinary least squares regression, and axis_1 and axis_2 are variables representing the location of the points on the representational axes. On each variable (axis_1 , axis_2 , and the external party dimension) there are as many observations as points in the space, in this case five. While the particular representational axes used will influence the b values and the constant term, the eventual orientation of the axis representing party will be entirely independent of them. Similarly, the multiple correlation coefficient will not be influenced by the particular representational axes used.

To determine the orientation of the axis, we calculate its slope by taking the ratio of the b coefficients. For example, if $b_2/b_1 = 3$, the axis should be drawn so that for each one-unit change on the first representational axis there is a three-unit change on the second representational axis. When we actually perform the regression with these data our results are:

$$10.8 \text{axis}_1 - 5.7 \text{axis}_2 + .002 + \text{error} = \text{external party, and}$$

$$1.2 \text{axis}_1 + 6.5 \text{axis}_2 + .200 + \text{error} = \text{external ideology.}$$

The b_2/b_1 ratio used to orient the party axis internal to the space is $-5.7/10.8 = -0.528$; the b_2/b_1 ratio used to orient the internal ideological axis is $6.5/1.2 = 5.417$. Both of these axes are drawn in Figure 12.¹³ The

¹³ A simple procedure which can be used to construct an axis once the b coefficients have been determined is to mark the point with coordinate (b_1, b_2) and then draw the line determined by that point and the origin. This line is the appropriate internal axis.

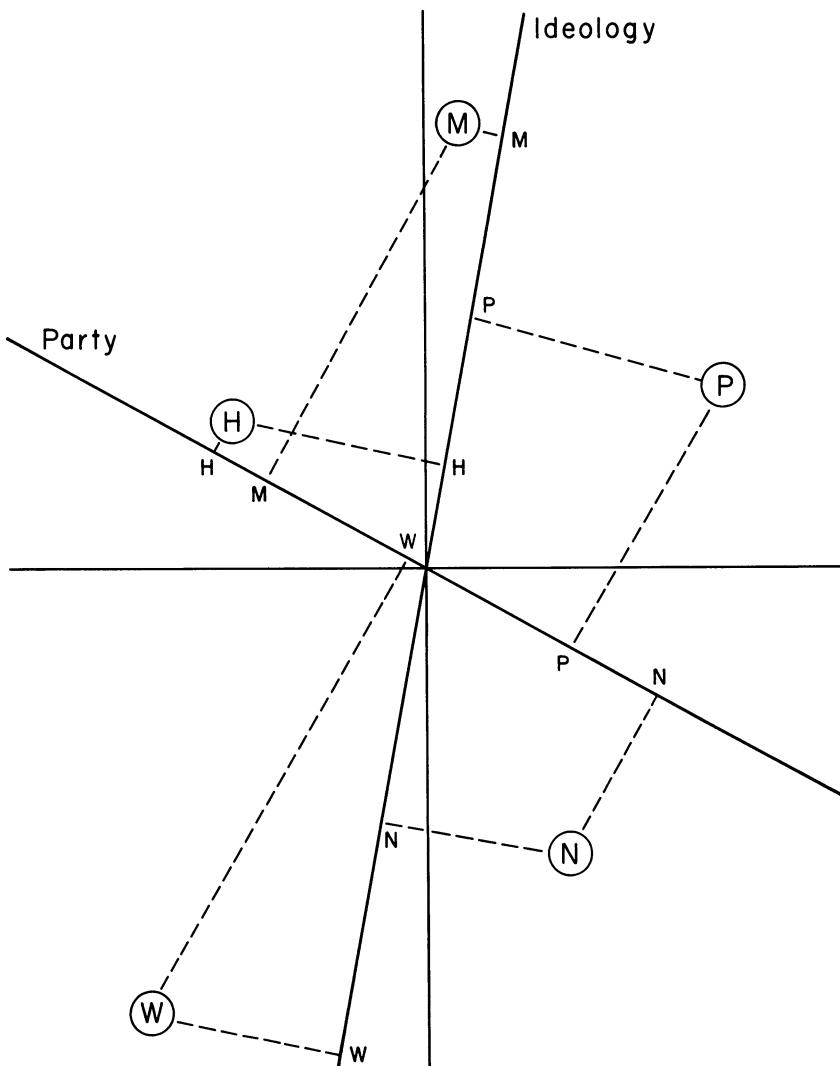


FIGURE 12

The Final Configuration of Five Points with Representational and Party Identification and Ideology Axes Inserted

correlation between the projection of points on the internal party axis and external party dimension is .996, and that between the internal ideological axis and the external ideological dimension is .995. The close correspondence is apparent if one compares the projection of points on the internal axes in Figure 12 with the external dimensions in Figure 11.

The results support the hypothesis that party and ideology were critical in determining the observed spatial structure. Of course, the small N of five candidates warrants considerable caution. In general, the use of external dimensions to reveal the underlying basis for a spatial structure is a potent addition to the rotation methods. By using external dimensions, the analyst determines axes with real substantive interpretability and obtains a measure of fit between the hypothesized determinants of the structure and the spatial structure which has been observed.

Problems with the Method

As with most analytic procedures, these methods present problems to which an investigator should be sensitive before applying them in his research. Problems arise from two sources, one the nonmetric nature of the procedures, the other the iterative strategy used in arriving at a best fit solution.

Problems Associated with the Nonmetric Assumptions

These techniques make only ordinal assumptions and return metric results. The jump in level of measurement occurs because of the metric constraints present in this type of data. One need work only a few examples by hand to realize that for most similarity matrices only a single representation is possible. However, when the number of dimensions becomes large relative to the number of variables, the constraint essential to these procedures starts to evaporate. One can then obtain solutions with low Stress that do not represent the data well. Hence, *the procedure should not be applied (or should be applied very cautiously) to data when the ratio of objects to dimensions is small (ratio < 4)*. Similarly, *analysis with less than ten, and certainly with less than eight points should be avoided*. (We have violated both these caveats with our five-candidate problem in order to present an example with greater intelligibility.)

There is another caveat which arises from the nonmetric nature of the procedure. Let us again violate the constraint restrictions and suppose that we have six points whose correct location is displayed in Figure 13. Notice that the five points, excluding f , are located in two distinct clusters: the A,B,C cluster and the D,E cluster. The similarity order obtained using just the ten

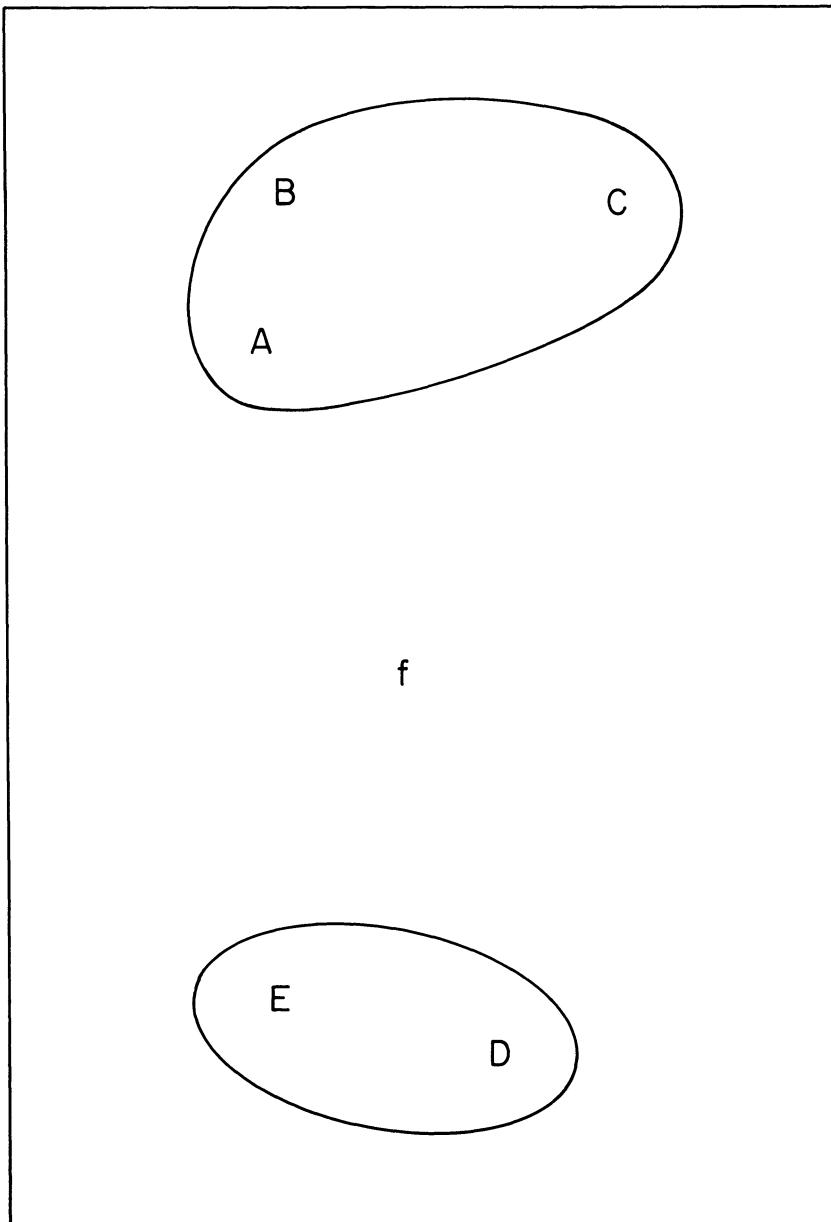


FIGURE 13

Illustration of a Potentially Degenerate Configuration
(Excluding f , all within-cluster distances are less than between-cluster distances.)

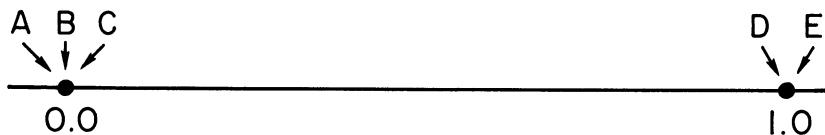


FIGURE 14

A Degenerate Solution Incorrectly Locating All Within-Cluster Points at the Same Position

pairs formed from these five points is AB, DE, BC, AC, AE, AD, BE, CD, CE, BD. This ordering can be broken into two distinct halves—the within-cluster pairs AB, DE, BC, and AC, followed by the between-cluster pairs AE, AD, BE, CD, CE, BD. When this condition arises a nonmetric procedure cannot be appropriately applied, for a “perfect” solution can always be obtained by locating each point in a cluster at exactly the same point in the solution space. For example, using just the five points, the obviously incorrect unidimensional solution in Figure 14 will be “perfect.” The AB, DE, BC, and AC pairs will all have an actual distance of 0.0 and the AE, AD, BE, CD, CE, and BD pairs will all have an actual distance of 1.0. These distances satisfy the monotonicity requirement so targets will equal distances and the Stress will be zero.

This condition usually arises when an analyst is using a very limited number of points. For example, the presence of the *f* point in this case would be enough to prevent the solution from degenerating.¹⁴ Nevertheless, it is important, particularly if the data are expected to be severely clustered, to be on the lookout for this type of degeneracy. It is easy to detect; Stress will be very low, even in one dimension, and points will be located in *very tight* distinct clusters. The occurrence of this degeneracy informs the analyst that his data are quite clustered, but gives him no insight into their finer structure. The problem can be circumvented only by performing a metric analysis.

¹⁴The rank order of pair similarity with *f* included is: AB, DE, BC, Ef, Af, AC, Df, Bf, Cf, AE, AD, BE, CD, CE, BD. If A, B, and C were located at the same position, the AC distance would be zero. This implies that the Af distance must be zero, since Af precedes AC in the similarity order. This in turn implies the *f* point would also have to be located at the same position as A, B, and C. However, the Ef pair precedes the Af pair in the similarity order; hence, the Ff distance would also have to be zero. Clearly, Stress would become very high were E positioned with A, B, and C; hence a different solution would be sought and the degeneracy would not occur.

Problems Associated with the Numerical Method (the Iterative Strategy)

In Figure 15 a curve is drawn. Notice that it has two low points, one at A and one at B. Point A is called a global minimum, since it is associated with the lowest point on the curve. Point B is called a local minimum, since it is associated with the lowest point on the curve in its immediate vicinity. If we

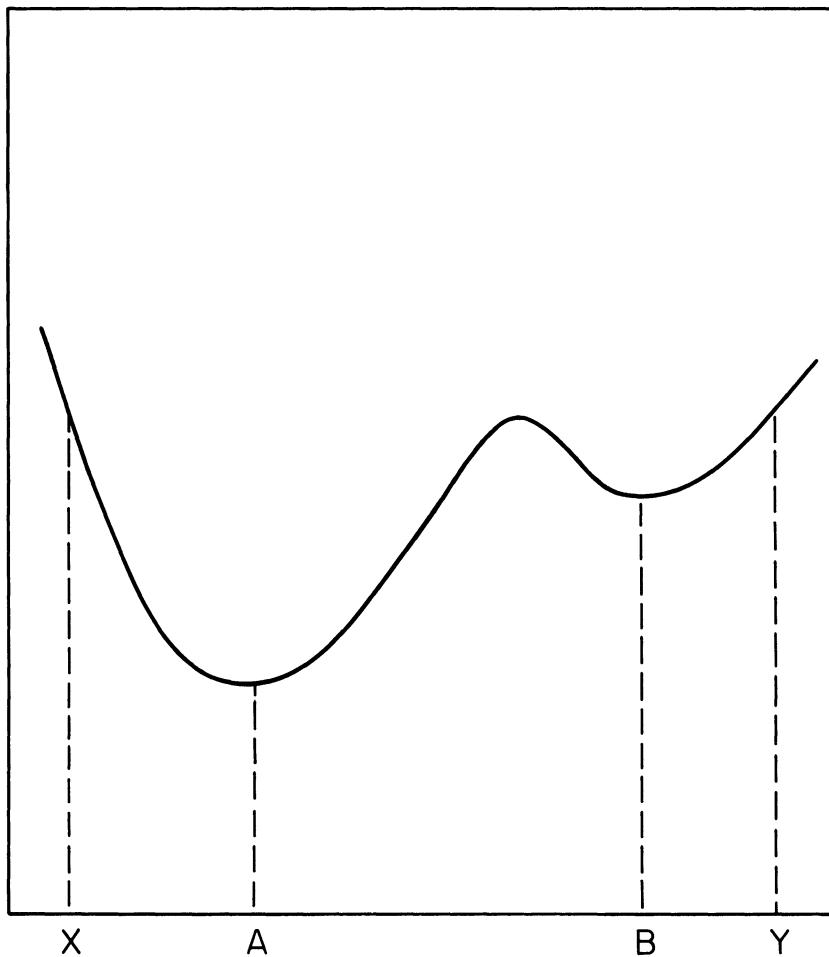


FIGURE 15
Illustration of a Local Minimum at B and a Global Minimum at A

start at point X and move to a minimum, we will wind up at A, the global minimum. If we start at point Y and move to a minimum, we will wind up at B, the local minimum. Notice that if we had a procedure oriented to obtaining a minimum from a fixed starting point, we would not know that B was a local and, not a global minimum. In analogous fashion, the procedure we use to minimize Stress is not sensitive to whether the minimum found is local or global. In either case, Stress would decline fairly consistently from iteration to iteration until the minimum was achieved.

If one does have a local minimum, it is usually easy to detect, in that Stress is high and *the solution does not make sense empirically*. Often, this is not a real problem since only one minimum exists, which must therefore be a global one. However, to help insure the analyst that the solution he reports is indeed the global minimum, several strategies are available.

The particular starting configuration we use is an important determinant of the likelihood of a local minima solution. We started our five-candidate problem by randomly (arbitrarily) locating five points in the space. If we had started with a more reasonable configuration, for example, one based on a metric analysis of the same data matrix, we would have been much closer to the eventual best solution. In general, a rational (as opposed to a random) start is an excellent way to reduce the risk of hitting a local minimum. Another good strategy is simply to perform the analysis several times using *different random* starting configurations, and then choosing the one solution with the lowest Stress. Either of these approaches is effective in nullifying potential local minima problems. The "safest" of all strategies is to run one solution with a rational start and several with random starts to check on the result. Most of the nonmetric multidimensional scaling programs allow for either a series of random starts, some type of rational start, or both.

Programs Available

There are several programs available to perform nonmetric multidimensional scaling. The three which have been most widely used are MDSCAL (new version KYST), TORSCA (new version POLYCON), and SSAI (this program is regularly updated). *The programs produce quite similar results and may be used interchangeably*. The main distinctions between the programs are technical, and involve the way in which target values and the measure of fit (Stress) are calculated and how one obtains the initial configuration to start the analysis. The basic algorithm in each of these programs is essentially identical to the one described in this paper.

Summary

We have completed a general introduction to nonmetric multidimensional scaling. Let us now summarize the major points.

Assumption

When points representing objects are located spatially, if, according to the measure used, one pair of objects is more similar than another pair, the points in the more similar pair should be located closer in the space.

Procedure

Points representing the objects are located in a space of fixed dimensionality and are then moved iteratively until a minimum Stress configuration is obtained. Stress is a measure of badness to fit; it ranges from 0.0 for a perfect solution to 1.0 for the worst possible configuration.

Guidelines for Use

1. Input data is a matrix which reflects the similarity or dissimilarity between pairs of objects.
2. The researcher specifies the dimensionalities in which he wishes to obtain solutions. The eventual decisions on dimensionality are based on substantive expertise and the plot of Stress against dimensionality.
3. The adequacy of a solution is determined using the following guidelines, realizing that the guidelines are not perfect and are sensitive to the number of points used in the analysis.

<i>Quality of Solution</i>	<i>Stress 2</i>	<i>Stress 1</i>
Perfect	0.00–0.05	0.000–0.025
Excellent	0.05–0.10	0.025–0.050
Good	0.10–0.20	0.050–0.100
Fair	0.20–0.40	0.100–0.200
Poor	0.40–1.00	0.200–1.000

4. The solution should be interpreted initially by describing the overall clustering pattern. If a dimensional interpretation is desired, either the standard rotational methods or the use of external criteria is appropriate.

Caveats

1. Be very circumspect about solutions where the ratio of the number of points to the number of dimensions is less than four.
2. Do not work with less than eight and preferably not less than ten points.
3. If Stress is very low and points are located in tight distinct clusters, the solution is probably degenerate. Check the original matrix to see if all the within-cluster pairs are more similar than any between-cluster pairs.
4. Beware of potential local minima problems.

Annotated Bibliography*General*

There is an extensive nonmetric multidimensional scaling literature. The articles which develop the theory underlying the three major nonmetric multidimensional scaling programs are: Shepard (1962a, 1962b), Kruskal (1964a, 1964b), Guttman (1968), and Young (1968); of these, the Kruskal (1964a) piece is the most readable for a nontechnical audience. The algorithm used in nonmetric multidimensional scaling is quite flexible and can be applied to a wide variety of problems. A sense of the variety of potential uses is provided in Young (1972) and Lingoes (1972). The most common alternate use is in the direct analysis of preference data; procedures appropriate to this application are discussed in Green and Carmone (1970), Gleason (1969), and Rabinowitz (1973). An overview of current scaling procedures not restricted to nonmetric multidimensional methods appears in Shepard (1972). Many of the programs Shepard discusses are considered in more detail in Green and Rao (1972).

MacRae and Schwarz (1968) and Weisberg (1968) compare nonmetric multidimensional scaling and factor analysis in the analysis of legislative roll calls. Weisberg (1974) compares principal component analysis and nonmetric multidimensional scaling more generally. Weisberg and Rusk (1970), Rusk and Weisberg (1972), Mauser (1972), and Rabinowitz (1973) provide applica-

tions of nonmetric multidimensional scaling techniques. They all analyze the dimensions underlying candidate evaluations. Taken together, the Weisberg and Rusk, Mauser, and Rabinowitz pieces illustrate the scope of data to which the method can be legitimately applied and the importance of the similarity measure in determining the eventual spatial configuration.

Sources

The nonmetric multidimensional scaling programs referred to in the paper are available from the sources listed below.

MDSCAL and KYST

Computer Program Librarian
Bell Laboratories
Murray Hill, New Jersey 07974

TORSCA and POLYCON

Professor Forrest Young
Psychology Department
University of North Carolina
Chapel Hill, North Carolina 27514

SSAI

Professor J. C. Lingoes
1000A N. University Building
The University of Michigan
Ann Arbor, Michigan 48104

For a comparison of the effectiveness of TORSCA, SSAI, and MDSCAL, see Spence (1972) and Lingoes and Roskum (1973).

Metric Constraint and Interpreting Stress

The degree of spatial constraint present in an ordinal data matrix is examined in Shepard (1966).

Detailed investigations of the relation between Stress and the adequacy of solution appear in Young (1970) and Sherman (1972). Spence and Ogilvie

(1973) present tables which delineate realistic upper bounds for Stress values given a fixed number of points and a fixed dimensionality.

Factor Analysis and Metric Multidimensional Scaling

There are many texts dealing with factor analysis. An excellent general text is Harman (1967). A less technical and quite readable text is Rummel (1970).

A method for performing metric multidimensional scaling appears in Torgerson (1958, pp. 247-297).

Interpreting Structure

For a good discussion of the use of unidimensional scales in interpreting spatial structure, see Cliff and Young (1966). A very useful, but fairly difficult article which deals with some alternate approaches to interpreting a nonmetric multidimensional scaling solution is Degerman (1972).

Appendix A

Calculation of Target Values Using the Least Squares Criteria

Let us suppose that the actual interpoint distances were those displayed in the second column of Table 5. Let us calculate the target distances associated with these actual distances. The calculations are displayed in the next five columns of Table 5.

We start by simply listing the actual distances until we have a violation of the monotonicity requirement. The first four pairs all have their distances in the correct order. Hence, initially these target distances are the same as the actual distances. The fifth distance, however, is smaller than the fourth distance; this violates the monotonicity requirement. To try to rectify this violation we will simply average the fourth and fifth distances. The fourth distance is 7.0; the fifth distance is 2.0; their average is $(7.0 + 2.0)/2 = 9.0/2 = 4.5$. However, this still does not solve the problem; our new target of 4.5 for the fourth pair is less than the target of 6.0 for the third pair. To keep the targets monotonic we must also include the third pair in the average. The new average is $(6.0 + 4.5 + 4.5)/3 = 15.0/3 = 5.0$. The target values for the first five pairs are now monotonic (2.0, 4.0, 5.0, 5.0, 5.0), so we can proceed to the sixth pair. Again we will set the target distance equal to the actual distance unless there is a violation in the monotonicity requirement. The

TABLE 5

Illustration of Target Value Calculation Using Least Squares Method

Pair	Actual Distance	Calculation of Targets					Final Target	Actual Distance
1	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
2	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
3	6.0	6.0	6.0	5.0	5.0	5.0	5.0	6.0
4	7.0	7.0	4.5	5.0	5.0	5.0	5.0	7.0
5	2.0	2.0	4.5	5.0	5.0	5.0	5.0	2.0
6	7.0			7.0	7.0	7.0	7.0	7.0
7	8.0			8.0	8.0	8.0	8.0	8.0
8	9.0			9.0	9.0	9.0	9.0	9.0
9	10.0			10.0	9.5	9.5	10.0	
10	9.0			9.0	9.5	9.5	9.0	

sixth, seventh, eighth, and ninth distances have actual distances consistent with the monotonicity requirement, so their targets are set equal to their actual distance. The tenth pair has an actual distance of 9.0, which is less than the target distance of 10.0 associated with the ninth pair. These two must be averaged to form the targets for the ninth and tenth pairs. Their average is $(10.0 + 9.0)/2 = 19.0/2 = 9.5$. This creates no new violations of the monotonicity requirement; hence our ten targets are now calculated.

The full set of targets appears in the next-to-last column of Table 5. The actual distances reappear in the last column. Notice that the target distances do seem to serve their purpose. The pair most clearly out of order is the fifth pair, where the actual distance is too small. Here the target distance is considerably greater than the actual distance. Were the two pairs immediately preceding it in the order closer together, the fit would also improve, and for them the target distances are less than the actual distances. For the first, second, sixth, seventh, and eighth pairs the target and actual distances are the same, and these pairs do seem to have reasonable interpoint distances. The ninth and tenth pairs are slightly out of order, and here the ninth target is slightly smaller and the tenth slightly larger than their respective actual distances.

Appendix B

To calculate the targets using the rank image method, one simply sorts the actual distances and uses the smallest distance as the target for the most

TABLE 6

Illustration of Target Value
Calculations Using Rank
Image Method

Pair	Actual Distance	Target Distance
1	2.0	2.0
2	4.0	2.0
3	6.0	4.0
4	7.0	6.0
5	2.0	7.0
6	7.0	7.0
7	8.0	8.0
8	9.0	9.0
9	10.0	9.0
10	9.0	10.0

similar pair, the next smaller distance as the target for the next most similar pair, and so on. The actual and target distances appear in Table 6 below. Notice, the targets are simply the actual distances, now sorted to run from the most to least similar.

Manuscript submitted February 7, 1974.

Final manuscript received December 19, 1974.

REFERENCES

Cliff, Norman, and Young, Forrest W. 1968. "On the Relation Between Unidimensional Judgments and Multidimensional Scaling," *Organizational Behavior and Human Performance*, August 1968, pp. 269-285.

Degerman, Richard L. 1972. "The Geometric Representation of Some Simple Structures," in R. Shepard, A. K. Romney, and S. B. Nerlove, eds., *Multidimensional Scaling, Volume I, Theory*. New York: Seminar Press, pp. 193-211.

Gleason, Terry. 1969. *Multidimensional Scaling of Sociometric Data*. Ann Arbor: Institute for Social Research.

Green, Paul E., and Carmone, Frank J. 1970. *Multidimensional Scaling and Related Techniques in Marketing Analysis*. Boston: Allyn and Bacon.

Green, Paul E., and Rao, Vithala R. 1972. *Applied Multidimensional Scaling*. New York: Holt, Rinehart, and Winston.

Guttman, Louis. 1968. "A General Technique for Finding the Smallest Coordinate Space for a Configuration of Points," *Psychometrika*, December 1968, pp. 469-506.

Harman, Harry H. 1967. *Modern Factor Analysis*. Chicago: University of Chicago Press.

Kruskal, Joseph. 1964a. "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis," *Psychometrika*, March 1964, pp. 1-27.

_____. 1964b. "Multidimensional Scaling: A Numerical Method," *Psychometrika*, June 1964, pp. 115-129.

Lingoes, James C. 1972. "A General Survey of the Guttman-Lingoes Nonmetric Program Series," in R. Shepard, A. K. Romney, and S. B. Nerlove, eds., *Multidimensional Scaling, Volume I, Theory*. New York: Seminar Press, pp. 52-68.

Lingoes, James C., and Roskam, Edward E. 1973. "A Mathematical and Empirical Analysis of Two Multidimensional Scaling Algorithms," *Psychometrika*, December 1973, Monograph Supplement, pp. 1-93.

MacRae, Duncan, Jr. 1970. *Issues and Parties in Legislative Voting*. New York: Harper and Row.

MacRae, Duncan, Jr., and Schwarz, Susan B. 1968. "Identifying Congressional Issues by Multidimensional Models," *Midwest Journal of Political Science*, May 1968, pp. 181-201.

Mauser, Gary A. 1972. "A Structural Approach to Predicting Patterns of Electoral Substitution," in R. Shepard, A. K. Romney, and S. B. Nerlove, eds., *Multidimensional Scaling, Volume II, Applications*. New York: Seminar Press, pp. 245-287.

Morrison, Richard. 1972. "A Statistical Model of Legislative Roll Call Analysis," *Journal of Mathematical Sociology*, July 1972, pp. 235-248.

Rabinowitz, George B. 1973. *Spatial Models of Electoral Choice: An Empirical Analysis*. Chapel Hill: Institute for Research in Social Science.

Rummel, Rudolph J. 1970. *Applied Factor Analysis*. Evanston: Northwestern University Press.

Rusk, Jerrold G., and Weisberg, Herbert F. 1972. "Perceptions of Presidential Candidates: A Midterm Report," *Midwest Journal of Political Science*, August 1972, pp. 388-410.

Shepard, Roger N. 1962a. "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function, I," *Psychometrika*, June 1962, pp. 125-140.

_____. 1962b. "The Analysis of Proximities: Multidimensional Scaling with an

Unknown Distance Function, II," *Psychometrika*, September 1962, pp. 219-246.

_____. 1966. "Metric Structures in Ordinal Data," *Journal of Mathematical Psychology*, July 1966, pp. 284-315.

_____. 1972. "Introduction to Volume I," in R. N. Shepard, A. K. Romney, and S. B. Nerlove, eds., *Multidimensional Scaling, Volume I, Theory*. New York: Seminar Press, pp. 1-22.

Sherman, Charles R. 1972. "Nonmetric Multidimensional Scaling: A Monte Carlo Study of the Basic Parameters," *Psychometrika*, September 1972, pp. 323-355.

Spence, Ian. 1972. "A Monte Carlo Evaluation of Three Nonmetric Multidimensional Scaling Algorithms," *Psychometrika*, December 1972, pp. 461-486.

Spence, Ian, and Ogilvie, John C. 1973. "A Table of Expected Stress Values for Random Rankings in Nonmetric Multidimensional Scaling," *Multivariate Behavioral Research*, October 1973, pp. 511-517.

Torgerson, Warren S. 1958. *Theory and Methods of Scaling*. New York: John Wiley and Sons.

Weisberg, Herbert F. 1968. *Dimensional Analysis of Legislative Roll Calls*. Unpublished doctoral dissertation, University of Michigan.

_____. 1974. "Dimensionland: An Excursion into Spaces," *American Journal of Political Science*, November 1974, pp. 743-776.

Weisberg, Herbert F., and Rusk, Jerrold G. 1970. "Dimensions of Candidate Evaluations," *American Political Science Review*, December 1970, pp. 1167-1185.

Young, Forrest W. 1968. "TORSCA-A FORTRAN IV Program for Nonmetric Multidimensional Scaling," *Behavioral Science*, July 1968, pp. 343-344.

_____. 1970. "Nonmetric Multidimensional Scaling: Recovery of Metric Information," *Psychometrika*, December 1970, pp. 455-473.

_____. 1972. "A Model for Polynomial Conjoint Analysis Algorithms," in R. Shepard, A. K. Romney, and S. B. Nerlove, eds., *Multidimensional Scaling, Volume I, Theory*. New York: Seminar Press, pp. 69-104.