This is to certify that the

thesis entitled

AN EVALUATION OF THE INTERACTIVE SIMILARITY ORDERING METHOD OF COLLECTING DATA FOR MULTIDIMENSIONAL SCALING ANALYSIS

presented by

David Edward Ehresman

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Psychology

Raymond W Frankmann

Major professor

Date 10 May 1980

AN EVALUATION OF THE INTERACTIVE SIMILARITY ORDERING METHOD OF COLLECTING DATA FOR MULTIDIMENSIONAL SCALING ANALYSIS

By

David Edward Ehresman

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

1980

ABSTRACT

AN EVALUATION OF THE INTERACTIVE SIMILARITY ORDERING METHOD OF COLLECTING DATA FOR MULTIDIMENSIONAL SCALING ANALYSIS

By

David Edward Ehresman

One drawback to multidimensional scaling techniques is the large number of judgments that are usually needed. One method of reducing the number and difficulty of these judgments is the Interactive Similarity Ordering (ISO) system.

Experiment I used Monte Carlo procedures to investigate the robustness of ALSCAL, a nonmetric multidimensional scaling program, with respect to incomplete row conditional data of the type produced by ISO. This study used configurations of 32 points in two dimensions and varied the amount of error added, the percentage of data analyzed, and the number of partitions of the proximity matrices. The results indicate that with one partition, as few as 40% of the data produce good solutions when the input has moderate error. With two partitions, 60% of the data is needed to produce comparable solutions.

In Experiment II, the ISO method is compared directly with the paired comparison method of collecting data. Ten subjects made judgments about the distances between 16 U.S. cities using both methods. The results were scaled using ALSCAL and the resulting cognitive maps were compared. The mean correlation between the distances of the two cognitive maps produced by a subject was 0.90

indicating that one gets similar results whether one uses the ISO method or the paired comparison method.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to the members of my dissertation committee, Dr. Raymond Frankmann (Chairman), Dr. Neal Schmitt, Dr. Lester Hyman, and Dr. Richard Dubes, whose guidance helped to make this work possible. I would also like to thank Mr. Mark Klein for his assistance in the preparation of this manuscript and Dr. Judith Frankmann for her encouragement and guidance throughout this project. Special thanks to my wife, Mary Anne, for her never-ending faith and understanding support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The large number of published applications in recent years attests to the wide spread use of nonmetric multidimensional scaling techniques in the social sciences. These techniques (e.g. Kruskal, 1964a, b; Takane, Young, and deLeeuw, 1977) construct a configuration of points in a metric space using only the ordinal or rank order information from a similarity of dissimilarity (proximity) matrix.

Typically, a proximity matrix is formed by having a subject judge the similarity or dissimilarity of all the C(n,2) = n * (n-1) / 2 pairs of n stimuli. As an illustration, consider Henley's (1969) Experiment II. She had subjects judge the dissimilarity of 30 animals. Each of the 435 (C(n,30)) pairs of animal names were presented one at a time and subjects were asked to rate them on a scale of 0 (no difference) to 10. These judgments were scaled and the three dimensional solution was chosen as the appropiate representation. The first dimension was interpretated in terms of the size of the animal; the elephant, camel, and giraffe were at one end of the continuum while the rat, mouse, and chipmunk were at the other extreme. The second dimension, with animals like the lion, tiger, and bear at one extreme and the cow, sheep, and deer at the other, was interpreted as a ferocity versus mildness continuum. The third dimension was more difficult to label. It was loosely interpreted as a "resemblance or relatedness to man or something similar" (p. 180).

Unfortunately, the number of pairs that must be rated goes up rapidly with the number of stimuli, n. For example, with n = 16, 120 pairwise judgments are necessary; with n = 32, 496 judgments must be collected to fill the triangular matrix; and with n = 48, there are 1128 pairs of stimuli. This large number of judgments has been a serious impediment to experimental designs that call for relatively large numbers of stimuli.

Several methods have been proposed for forming proximity matrices for large data sets. Young and Cliff (1972) developed a computer program which collects a subset of the C(n,2) pairwise comparisons. The subset of pairs is determined interactively on the basis of the subject's previous responses. Girrard and Cliff (1976) demonstrated by way of a Monte Carlo study that this program works quite well. However, from the point of view of many users, it has one insurmountable deficiency; it is a metric rather than a nonmetric procedure. That is, it assumes that the judgments are Euclidean distances, not merely proximities.

Another way to lower the number of judgments required involves sorting or grouping tasks of various kinds (e.g. Romney, Shepard, and Nerlove, 1972; Rao and Katz, 1971). After the sorting task is complete, a proximity matrix is derived, and the complete matrix is scaled. However, as Spence (in press) points out, it is questionable whether such a matrix really represents a subject's perception of the pairwise interstimulus proximities. Spence indicates that some highly experienced users of these sorting techniques urge that the results be used with the greatest of caution.

Yet another way of reducing the number of judgments a subject must make is to present a subset of all pairwise comparisons that has been chosen a priori. Spence and Domoney (1974), Graef and Spence (1979), and Spence (in press) have suggested several ways of selecting the subset which is to be presented. Among the methods they have discussed and evaluated are cyclic designs, random selection, and selection based on knowledge of the distances in the configuration that is to be obtained. Their Monte Carlo studies indicate that if enough judgments are collected, these partial proximity matrices yield solutions that are very nearly identical to those obtained by scaling the full matrix.

Young, Null, and Sarle (1978) recently developed an interactive computer program for collecting rank order data which can be scaled by the ALSCAL program (Takane, Young, and de Leeuw, 1977). The authors claim that this Interactive Similarity Ordering (ISO) system can collect data for a given stimulus set in a time comparable to that needed to collect enough data using an incomplete pairwise comparison design. In addition, the authors feel that the judgments in the rank ordering task are simpler than those in a pairwise comparison task.

The first part of this study will be a Monte Carlo study to evaluate ALSCAL's ability to analyze data of the type produced by ISO. The second part will compare the solutions obtained from a pairwise comparison task to those obtained from the ISO task.

<u>Monte Carlo Studies</u>

There have been a number of attempts to gain a better understanding of nonmetric multidimensional scaling techniques by means of

Monte Carlo investigations. One line of studies (Klahr, 1969; Stenson and Knoll, 1969; Levine, 1978) investigated the statistical significance of stress. (Stress is a "goodness-of-fit" measure between the input proximity matrix and the recovered distance matrix. See Appendix A and Kruskal (1964b) for a more detailed explanation.) These researchers scaled random data varying a number of parameters and summarized the data to provide a null hypothesis with which to compare stress values obtained in real studies. However, as Levine (1978) notes, Ling (1973) criticized these types of studies on the grounds that most sets of data which are to be scaled have enough structure a priori to reject a null hypothesis of randomness. Ling also notes that not all random permutations are equally probable as is the case in these types of Monte Carlo studies.

The majority of Monte Carlo studies have been concerned with "metric determinancy." The question these investigations have addressed is: Given the (possibly noisy) ordinal relation between points (stimuli), how well can a scaling algorithm recover a known configuration?

The basic methodology of these studies was (1) to generate a random configuration, (2) generate a proximity matrix by adding noise to the interpoint distances and possibly subjecting the noisy distances to a monotonic transformation, (3) scale the proximity matrix thus derived to generate a configuration, and (4) compute the correlation (or squared correlation) between the "true" and the recovered configurations to determine how well the algorithm recovered the orginal configuration.

Three different ways of adding error to the distances have been reported in the literature. Wagenaar and Padmos (1971) and Graef and Spence (1979) multiplied the distances by a random normal deviate. The normal error distribution had a mean of one and the variance was a parameter that was varied. Any negative deviates that were generated were discarded and a replacement was chosen.

Girrard and Cliff (1976) added error in a way that they argue yields proximities with a distribution similar to the distribution of similarity judgments made by subjects. They added a random normal deviate to the distances, linearly transformed them so most values were between -1.0 and +1.0, took the inverse Fisher Z transform, and then linearly transformed the proximities back to a scale of 1.0 to 9.0.

The most widely used method of adding random error has been the Ramsay method, so named because Ramsay (1969) noted that it is equivalent to sampling the square of a proximity from a non-central chi squared distribution. Error is introduced by adding a random normal deviate to each coordinate before the distance between points is computed. Ramsay (1969) and Young (1970) note that this is a multidimensional analogue of Thurstone's (1927) discriminal process.

Of the Monte Carlo investigations that have used the Ramsay model, there have been three different ways of specifying the variance of the normal distribution that is sampled to obtain the error deviates. Young (1970) specified the error variance, σ_e^2, relative to variance of the distances of the configuration, σ_d^2, such that $\sigma_e^2 = E^2 \sigma_d^2$, where E was an error level parameter. Sherman (1972) and Young and Null (1978) specified the error variance, σ_e^2, relative to the variance of the coordinates of the configuration, σ_c^2, i.e. $\sigma_e^2 = E^2 \sigma_c^2$. In the Young and Null study, σ_c was standardized to .333 for each dimension of all configurations. The final way of specifying σ_e^2 is as an arbitrary error level, $\sigma_e^2 = E^2$. This is the procedure used by Spence (1972), Spence and Domoney (1974), Graef and Spence (1979), and Spence (in press).

The ISO System

The Interactive Similarity Ordering (ISO) system (Young, Null, and Sarle, 1978) can collect several types of data. The type that is of interest in this study is called asymmetric or row conditional. The data are called row conditional because each judgment in the ith row is relative to the ith stimulus; this gives rise to a square asymmetric matrix.

In order to produce a row conditional matrix, the subject's task is as follows: Given a "standard" (one of the stimuli) and a list of the remaining stimuli, choose the stimulus from the list which is most similar to the standard. This task is repeated until all n-1 stimuli have been rank ordered relative to the standard, thus filling one row of the data matrix.

If the number of stimuli, n, is relatively large, it will take a subject a considerable amount of time to choose his or her response from the complete list of remaining stimuli. Therefore, ISO allows the experimenter to choose the maximum list length, i.e. the maximum number of alternatives presented to a subject at one time. ISO then uses a sorting algorithm called a merge sort (Knuth, 1973) to interactively minimize the number of judgments required by using the transitive relationship,

$$(r_{ij} < r_{ik} \text{ and } r_{ik} < r_{il}) \Rightarrow r_{ij} < r_{il}, \tag{1}$$

where r_{ij}, r_{ik}, and r_{il} are the rank order of the jth, kth, and lth stimulus with respect to the standard, stimulus i. Note that this technique uses only the ordinal information of the response, thus making it a nonmetric technique.

By setting the maximum list length to less than the number of stimuli, one increases the number of judgments that must be made with respect to a given standard. Because not all the stimuli are presented at once, additional judgments are necessary to determine the relative order of stimuli that do not initially appear on the same sublist. However, the judgments are simpler because there are fewer alternatives to choose from, and can therefore be made more quickly. Young, Null, and Sarle (1978) indicate that by partitioning the stimuli into two sublists, one increases the number of standards a subject can order in an hour. This increase is larger for medium list length than for small list length.

The experimenter can also shorten the time it takes to complete an experiment by using only a random subset of the stimuli as standards. This is analogous to the method of presenting a random subset of pairwise comparisons as described by Spence and Domoney (1974).

The user of the ISO system thus has a range of options in deciding how much data to collect and how to collect it. It is the purpose of this study to help the experimenter make an intelligent choice when using ISO as a data collection tool.

EXPERIMENT I: MONTE CARLO STUDY

The general procedure used to evaluate ALSCAL's ability to analyze data of the type produced by ISO (rank order, row conditional data) is as follows: (1) generate a number of random configurations, (2) from each configuration, produce a proximity matrix by adding a random error component to the coordinates before calculating the Euclidean distance between pairs of points, (3) from each proximity matrix, produce a row conditional, rank order matrix by rank ordering each row (or partition of a row) in the proximity matrix, (4) scale the row conditional rank order matrix using the ALSCAL program, and (5) compare the configuration produced by ALSCAL with the "true" configuration.

Procedure

The "true" configurations were generated by using the method described by Spence (1972). Coordinates were obtained by randomly sampling from the uniform distribution on the interval (-1.0, +1.0) with the added constraint that all points be within a hypersphere of radius 1. Following a trend in the literature, five configurations, each consisting of 32 points in two dimensions, were generated in this manner. These served as the true or population configurations in this study, thus giving five replications.

This study consisted of a complete factoral design of 2 x 2 x 4 with five replications, where the factors were (1) the amount of error added to the coordinates, (2) the number of partitions or sublists, and (3) the number (percentage) of standards which were ordered. The levels of each of these factors is described in detail below.

Error was added to the coordinates using the Ramsay model. Perturbed distances, d', were computed as,

$$d'_{ij} = \left[\sum_{a=1}^{m} (x'_{ia} - x'_{ja})^2\right]^{\frac{1}{2}}, \tag{2}$$

where $x'_{ia} = x_{ia} + e_{ria}$, x_{ia} is the true configuration coordinate for point i on dimension a, and $e_{ria} : N(0, \sigma_r^2)$. Equivalently,

$$d'_{ij} = \left[\sum_{a=1}^{m} (x_{ia} - x_{ja} + e'_{rija})^2\right]^{\frac{1}{2}}, \tag{3}$$

where $e'_{rija} : N(0, 2\sigma_r^2)$. Fresh error deviates were used each time a distance was calculated as implied by the subscripts on e. The error level, r, took on two levels; $r = 1$ had $\sigma = 0.0$ and $r = 2$ had $\sigma = 0.15$. Spence and Domoney (1974) refer to these error levels as yielding errorless and moderately perturbed distances.

The variance of the error distribution used in this study was fixed, i.e. it was not relative to the variance of the coordinates or the distances. This is the method that has been used by Spence and his coworkers (e.g. Spence and Domoney, 1974). Since the mean variance of the interpoint distances was 0.4293, an error level of 0.15 would be approximately equivalent to an error level of 0.35 if the error variance was proportional to the variance of the distances as in Young (1970). The mean variance of the coordinates was 0.5069 so the 0.15 error level would be approximately equal to an error level of 0.30 if the error variance was proportional to the variance of the coordinates as in Sherman (1972).

For each perturbed distance matrix, a row conditional proximity matrix was formed by rank ordering the distances in each row using the values 1 to n. This yields the full matrix which ISO would produce if the distance matrix represented the subject's perception of the interstimulus proximities, if all the stimuli were used as standards, and if one sublist was used (containing all of the stimuli except the standard).

The number of partitions factor took on two levels, one and two. The partition of one is the matrix described in the previous paragraph. For a partition of two sublists, one needs two incomplete matrices. These two proximity matrices (to be scaled as replications of one subject) were formed by randomly assigning each element of a row to one of the two partitions, thus halving each row of the perturbed distance matrix into two submatrices. The elements in each row of the first submatrix were converted to ranks and placed into one matrix and similarly the second submatrix was converted to ranks to obtain the second matrix.

Finally for each full and partitioned matrix, 40%, 60%, 80%, and 100% of the rows were randomly choosen to remain in the matrix to be submitted to ALSCAL. This represents the ISO option of choosing the number of standards to be ordered.

This 2 x 2 x 4 design with five replications thus gives rise to 80 data matrices. These were submitted to ALSCAL (version 2.03) as implemented on the University of Michigan's Amdahl computer running the MTS operating system and was accessed via the Merit network. The ALSCAL parameters were set as shown in Table 1. Note particularly

Table 1. ALSCAL parameter values* used in the Monte Carlo study.

Number of stimuli	32
Number of subjects	1 or 2 (depending of the number of partitions)
Measurement level	ordinal
Data type	asymmetric dissimilarity
Measurement process	discrete
Measurement conditionality	row conditional
Model type	simple Euclidean
Dimension of solution	2
Initial scaling	between subjects
Negative weights permitted	no
Nonmetric transformation	Kruskal's least squares
Convergence criterion	0.001

*For parameter definitions, see Takane, Young, and de Leeuw (1977).

that the nonmetric (ordinal), asymmetric matrix, and row conditional options were used.

Results

ALSCAL's ability to recover the known configuration was measured by calculating the product-moment correlation between the distances of the true configuration and the distances of the recovered configuration. This correlation, r_{TR}, or its square, is commonly used as the dependent measure in multidimensional scaling Monte Carlo studies. These correlations (averaged across replications) are plotted in Figure 1 as a function of error level, number of partitions, and percentage of standards (rows) analyzed. The raw correlations were converted to approximate normals using the Fisher Z transformation, averaged, and then converted back to correlations before plotting. Note that although r_{TR} decreases as the percentage of standards analyzed gets smaller and as the error level and number of partitions increase, all of the correlations are quite large. The lowest correlation is 0.86. An analysis of variance was performed using the correlations between the true and recovered configurations, converted to approximate normals, as the dependent measure. The cell means, plotted as Fisher Zs, are shown in Figure 2; the results of the analysis are shown in Table 2. The only effect that was not significant at the 0.05 level is the interaction between the number of partitions and the number of standards analyzed. Note that the data plotted in Figure 1 and in Figure 2 are the same data. Figure 2 uses the scale units that were used in the analysis of variance while Figure 1 uses the more familar correlation scale.

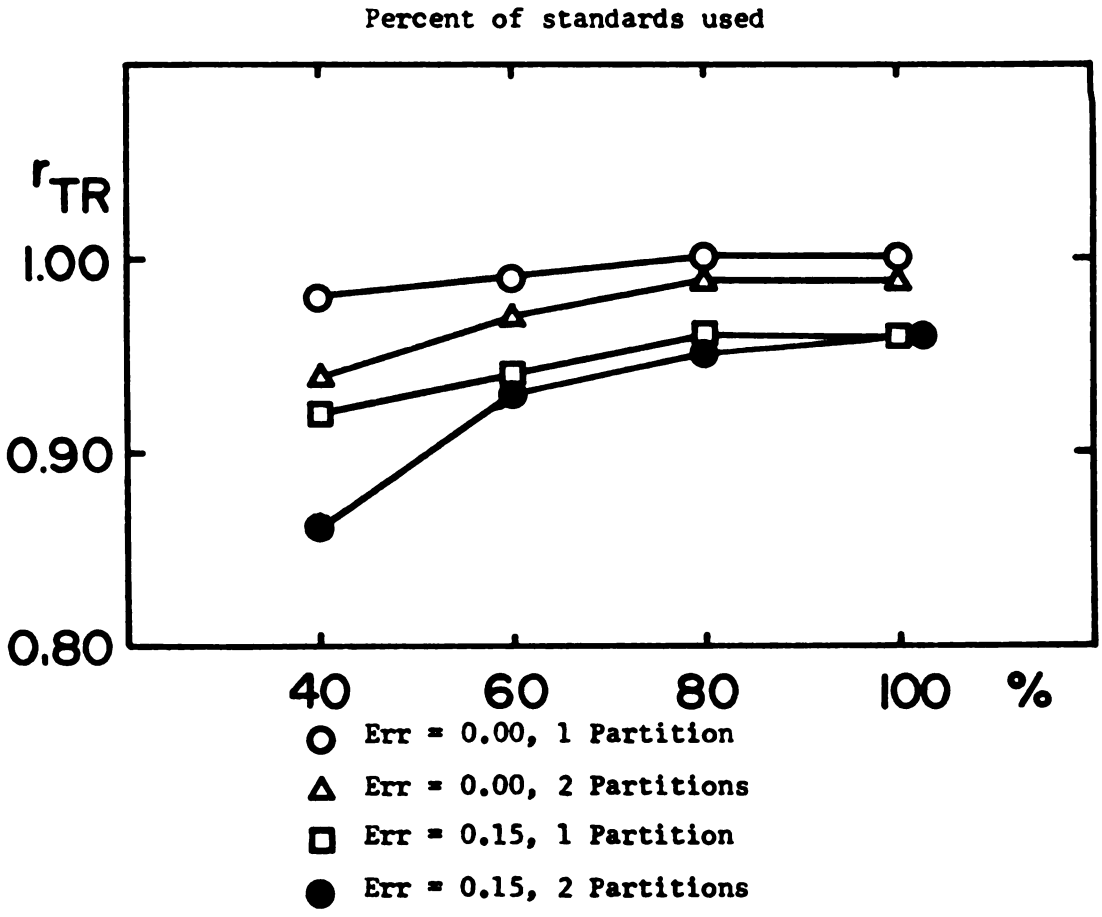

Figure 1. Mean correlations between true and recovered distances.

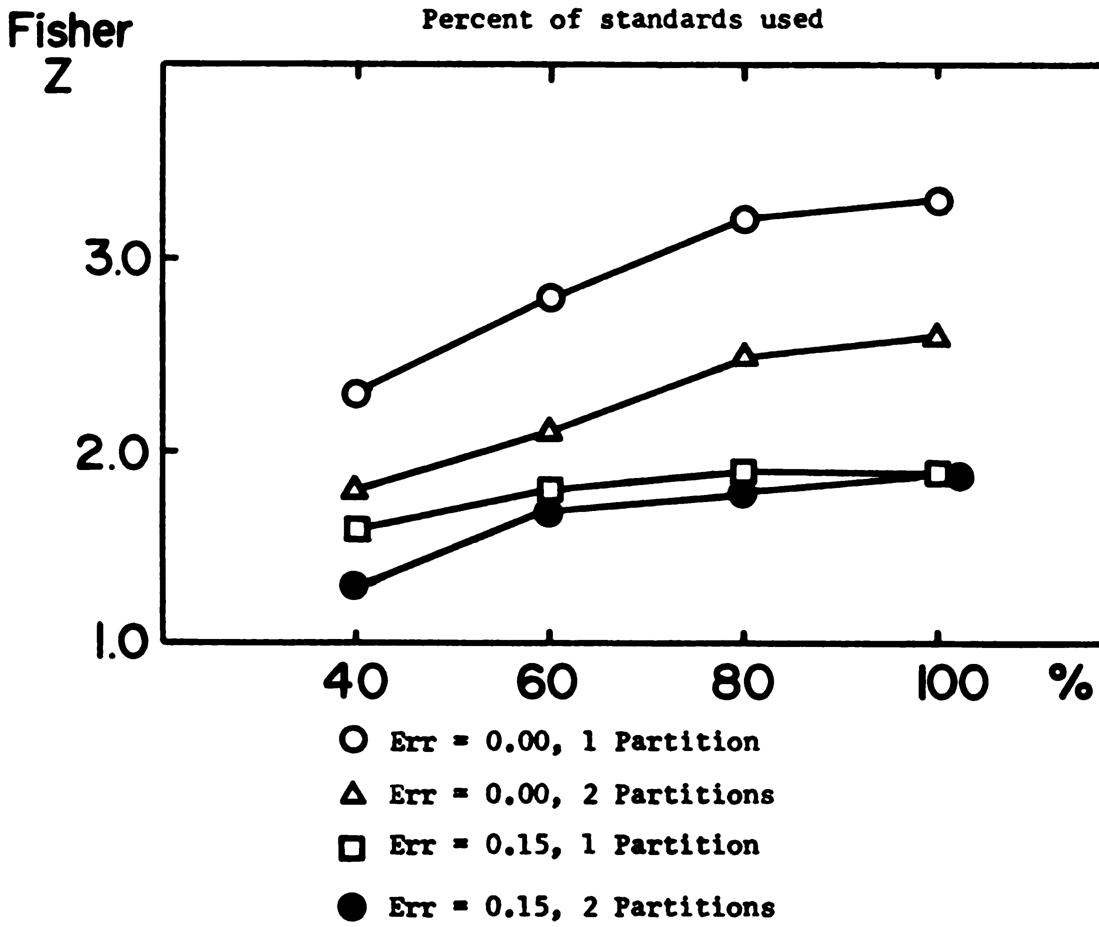

Figure 2. Mean Fisher Zs between true and recovered distances.

Table 2. Monte Carlo analysis of variance.

Source	Error Term	Degrees of Freedom	Mean Square	F Ratio	Approximate F Probability
A (error level)	AD	1	.135354E+02	457.024	< 0.0005
AD	None	4	.296163E-01		
B (no. of partitions)	BD	1	.345949E+01	311.763	< 0.0005
BD	None	4	.110965E-01		
C (% of standards)	CD	3	.186499E+01	81.644	< 0.0005
CD	None	12	.228430E-01		
AB	ABD	1	.156782E+01	168.527	< 0.0005
ABD	None	4	.930309E-02		
AC	ACD	3	.202602E+00	13.365	< 0.0005
ACD	None	12	.151589E-01		
BC	BCD	3	.450126E-03	0.126	0.943
BCD	None	12	.358172E-02		
ABC	ABCD	3	.466844E-01	4.603	0.023
ABCD	None	12	.101426E-01		
D (replications)	None	4	.115876E+00		
Total		79			

Figure 3 plots the SSTRESS, the stress-like "goodness-of-fit" measure that ALSCAL minimizes (See Appendix A). SSTRESS for errorless proximities is lower than for the proximities with error added. For the error free proximities, SSTRESS is higher for partitions of two than for partitions of one, and higher for low percentages of standards analyzed than for high percentages of rows analyzed. For the proximities with error added, the situation is reversed. SSTRESS is lower for a partition of two than for a partition of one and it gets smaller as the percentage of rows analyzed decreases.

Discussion

Although the three way interaction confounds any statistical interpretation of the main effects, much can be learned from the data plotted in Figure 1. These results indicate that when using the row conditional option of ALSCAL, one need not rank all the stimuli. This is in agreement with the work done by Spence and his coworkers (Spence and Domoney, 1974; Graef and Spence, 1979; Spence, in press) with pair comparison judgments.

For error free input and a stimulus set of 32, one could safely use as few as 40% of the stimuli as standards. This is true regardless of whether one chooses to use one or two partitions of the input matrix. Given the time savings reported by Young, Null, and Sarle (1978) for a partition of two, this would be the preferred method when using the ISO system.

For two dimensional data containing moderate error and a stimulus set of 32, one could again use as few as 40% of the stimuli as standards when using a partition of one. When using a partition of two, the recovery correlation drops to 0.86 when 40% of the data

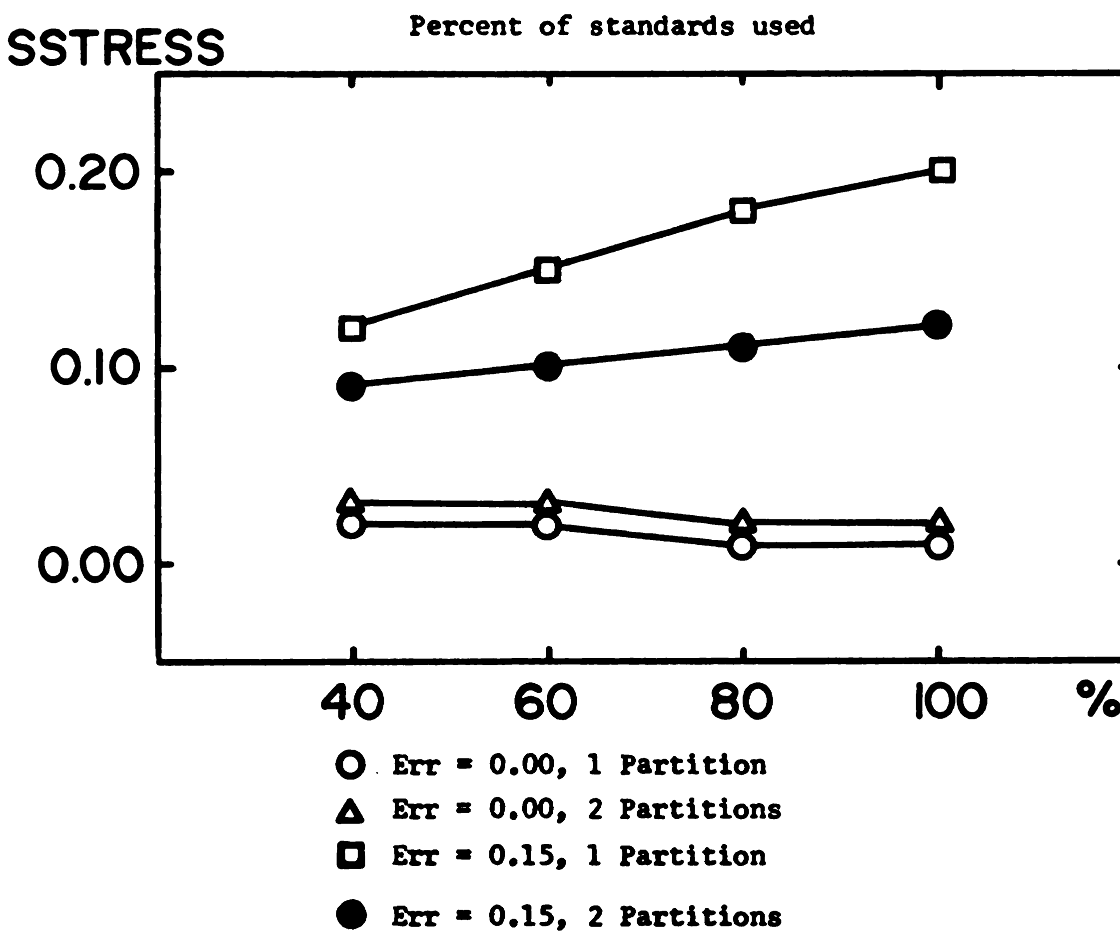

Figure 3. Mean SSTRESS for ALSCAL solutions.

is analyzed. While this is by no means a poor fit, it differs fairly sharply from the correlation of 0.93 for 60% of the standards analyzed. Thus, one might well prefer to use 60% of the stimuli as standards. This is still a sizeable reduction in the task demanded of the subject.

Spence and Domoney (1974) suggest collecting a minimum of 50% and 55% of the pairwise judgments for data with zero and moderate error when analyzing a three dimensional configuration of 32 points. This corresponds very well to the data in Figure 1 although their recommendation is based only on analysis of 1/3, 2/3, and complete data. They also present more complete data for 40 and 48 points. Not surprisingly, the larger the stimulus set, the lower the percentage of data that must be analyzed. This should also hold for row conditional data although it has not been tested. Graef and Spence (1979) obtained similar results for 31 points in two dimensions in a study that compared cyclic deletions and deletion based on a priori knowledge of the size of the distances between stimuli.

Figure 3, which displays SSTRESS as a function of the parameters of this study, should serve as another warning against using stress measures to evaluate the quality of a scaling solution. For the errorless data, SSTRESS is <u>inversely</u> monotonically related to the recovery correlation with a partition of two having the <u>higher</u> SSTRESS. For the data with moderate error, SSTRESS is <u>directly</u> monotonically related to the recovery correlation with the partition of two having the <u>lower</u> SSTRESS.

EXPERIMENT II: ISO VERSUS PAIRED COMPARISONS

The purpose of this study is to compare the solutions one gets from an actual paired comparison task with the solutions one gets from a rank order ISO task using the same stimuli. It is desirable to separate this question from the question of the robustness of ALSCAL with respect to missing data which was discussed in the first study. Therefore, all paired comparisons and rank orders were obtained; this meant that a relatively small number of stimuli were used in this study.

Procedure

Sixteen U.S. cities were choosen to serve as stimuli. They are listed in Table 3. Ten subjects were recruited from undergraduate and graduate level psychology students at Michigan State University. Subjects were paid $5.00 to participate in the study.

Each subject performed two tasks: (1) judging the distances between all pairs of cities, and (2) rank ordering the distances of all 15 cities to the remaining one for all 16 cities. The paired comparison stimuli were presented in random order on a computer CRT screen. The subject had to rate the distance between each pair of cities on a scale of one to nine by typing in the appropriate number. One represented a judgment of "very close together" and nine represented a judgment of "very far apart." The rank order stimuli were presented in random order by the ISO system using all stimuli as standards and one partition. Young, Null, and Sarle (1978) indicate

Table 3. The 16 cities used in Experiment II.

Boston
New York
Washington, D.C.
Miami
Atlanta
Cincinnati
Detroit
Chicago
St. Louis
New Orleans
Dallas
Salt Lake City
Denver
Los Angeles
San Francisco
Seattle

that the maximum list length does not affect the time it takes to order a standard when only one partition is used, but that standards are ordered more quickly for longer list lengths when the stimuli are partitioned into two sublists. On the other hand, one of the advantages of the rank order task is that the judgments are simpler than the paired comparison judgments, and the shorter the maximum list length, the simpler the judgment should be to make. Since it is expected that most researchers using ISO will choose to partition their stimuli into sublists, it was decided not to use the shortest maximum list length of two. However, to keep the judgments quite simple, the maximum list length was set to four. Each subject produced two data matrices which were submitted to ALSCAL using the parameters listed in Table 4.

Results

The two configurations obtained for each subject were compared to each other using the correlation between the interpoint distances as the measure of correspondence. The mean correlation between the rank order cognitive map and the pairwise comparison cognitive map was 0.90. This was obtained by converting the correlation coefficients to approximate normals using the Fisher Z transformation, averaging, then converting back to correlations. If one drops the lowest correlation (0.57), the average increases to 0.92. There is some justification for dropping the low correlation. It was an obvious outlier, being the only one below 0.82. In addition, the subject was averaging 2 to 3 seconds per judgment towards the end of the rank order task. This was much quicker than her earlier response times and much quicker than the average response time of other

Table 4. ALSCAL parameter values* used to analyze cognitive distances.

Parameter	Rank order task	Paired comparison task
Number of stimuli	16	16
Number of subjects	1	1
Measurement level	ordinal	ordinal
Data type	asymmetric dissimilarity	symmetric dissimilarity
Measurement process	discrete	discrete
Measurement conditionality	row conditional	unconditional
Model type	simple Euclidean	simple Euclidean
Dimension of solution	2	2
Initial scaling	between subjects	between subjects
Negative weights permitted	no	no
Nonmetric transformation	Kruskal's least squares	Kruskal's least squares
Convergence criterion	0.001	0.001

*For parameter definitions, see Takane, Young, and de Leeuw (1977).

subjects. This suggests that the subject was just trying to get finished and was not being particularly careful about her responses. Six of the ten subjects had correlations of 0.90 or higher; the three others had correlations greater than 0.80.

Correlations were also computed between the distances as measured from a U.S. map and the recovered distances from the two cognitive maps. The average correlation for the rank order task was 0.90; the average correlation for the paired comparison task was 0.86. Within subjects the rank order correlation was never less than the paired comparison correlation. Informally, a few subjects indicated that the rank order map was a better representation of their perception than was the paired comparison map.

Discussion

Based on the high correlation between the distances from the rank order cognitive maps and the distances from the paired comparison cognitive maps, one can conclude that one gets very similar solutions regardless of which task a subject performs. The correlations between the cognitive maps and the actual U.S. map hint that the rank order task may produce slightly better solutions than the paired comparison task. (See Appendix B for an example of a cognitive map.)

Young, Null, and Sarle (1978) indicate that one advantage of the ISO system is that the judgments are simpler to make than paired comparison judgments. Subjects informally reported that although the rank order task was more tedious than the paired comparison task because there were many more judgments, the rank order judgments were indeed simpler. The tedium of the rank order task should be greatly reduced by partitioning the stimuli and using only a subset of them as standards as discussed in the Monte Carlo study. Young, Null, and Sarle (1978) suggest that an ISO task with partitioned stimuli should take no longer than a paired comparison task to collect equal amounts of data.

GENERAL DISCUSSION

The results of the Monte Carlo experiment presented earlier indicate that in the case studied, one can reduce the number of stimuli used as standards without sacrificing the quality of the solution. One can also greatly reduce the number of judgments by partitioning the stimuli into two sublists and rank ordering each subset to the standard. The Monte Carlo study also indicates that this partitioning can be done without sacrificing the quality of the solution. Thus using these two methods one can greatly reduce the number of judgments needed.

The second experiment indicates the results one obtains from the rank order task are equivalent to the results one gets from the paired comparison task. Since the judgments required in the rank order task tend to be simpler than those of the paired comparison task, researchers who do multidimensional scaling studies with large stimulus set may well want to consider using the rank order task.

There are a number of obvious ways that this study could be extended. One could explore other dimensions, error levels, number of points, number of partitions, and so on. Perhaps even more useful would be a direct comparison of the cyclic and random deletion paired comparison tasks with the ISO task using the same randomly generated configurations as the basis for the input matrices. Another useful avenue of study would be a more thorough exploration of the trade offs of the various ISO options. In particular, it would be helpful to have data as to the length of time it takes subjects to complete an experiment when the number of stimuli, number of standards, number of partitions, and the maximum list length are varied.

APPENDICES

APPENDIX A

APPENDIX A

THE ALSCAL ALGORITHM

The general ALSCAL model (Takane, Young, and de Leeuw, 1977) is

$$d^2_{ijk} = \sum_{a=1}^{m} v_{ia} w_{ka} (x_{ia} - x_{ja})^2, \tag{4}$$

where d^2_{ijk} is the squared distance between points i and j for replication k, v_{ia} is a weight for point i on dimension a, w_{ka} is a weight for replication k on dimension a, and x_{ia} and x_{ja} are the coordinates for points i and j on dimension a. With $v_{ia} = w_{ka} = 1$, this general model simplifies to the simple Euclidean model that was used in this study.

The (unnormalized) objective function that ALSCAL minimizes is

$$\phi = \sum_{k=1}^{N} \phi_k = \sum_{k=1}^{N} \sum_{ijk}^{n} [d^2_{ijk} - f_{ik}(o^2_{ijk})]^2, \tag{5}$$

where o^2_{ijk} is the squared value of the observed dissimilarity between stimulus i and j on the kth replication, and f_{ik} is the transformation between observations and Euclidean distances. Note that a unique transformation function may be used for each row in each replication of the input (dissimilarity) matrix. This permits the row conditional scaling that was done in this study.

In nonmetric multidimensional scaling, the transformation function, f_{ik}, is a monotonic one. In ALSCAL the regression equation is

$$d^{*2}_{ijk} = f_{ik}(\phi^2_{ijk}), \tag{6}$$

subject to the linear inequalities which define monotonicity. This is a problem in isotonic regression for which solutions are available (Kruskal, 1964b).

In equation (5), if the function of the dissimilarity is replaced by the best fitting estimate of the squared distance, d^{*2}_{ijk}, and if d^{*2}_{ijk} is also used to normalize the equation, the result is the ALSCAL goodness-of-fit measure, SSTRESS, for the unconditional Euclidean model,

$$\phi_u^2 = \frac{\sum_i \sum_j \sum_k (d^2_{ijk} - d^{*2}_{ijk})^2}{\sum_i \sum_j \sum_k d^{*4}_{ijk}} . \quad (7)$$

This is quite similar to Kruskal's (1964b) STRESS goodness-of-fit measure, S, defined as

$$S^2 = \frac{\sum_i \sum_j (d_{ij} - d^*_{ij})^2}{\sum_i \sum_j d^2_{ij}} . \quad (8)$$

For the ALSCAL row conditional model, a normalized SSTRESS is computed for each row and then an average is computed over all these SSTRESSs as in equation (9),

$$\phi_r^2 = \frac{1}{Nn} \sum_k^N \sum_i^n \frac{\sum_j (d^2_{ijk} - d^{*2}_{ijk})^2}{\sum_j d^{*4}_{ijk}} . \quad (9)$$

The basic steps of the ALSCAL algorithm as used in this study are quite simple, although the implementation is rather complex. First an initial configuration of points is computed from the dissimilarities. The point and replication weights are set equal to one. Next the interpoint squared distances are computed from the configuration of points and the best fitting transformation between the squared dissimilarities and these squared distances are found for each row. New estimates of the squared distances, d^{*2}_{ijk}, are calculated using the transformations just found and the squared dissimilarities. SSTRESS is then computed and if it is small enough, the process is finished. Otherwise, a new configuration is found using the new best estimates of the squared distances. This process is repeated by finding a new transformation as described above.

APPENDIX B

APPENDIX B

ONE SUBJECT'S COGNITIVE MAPS

Figures 4 and 5 are an example of the cognitive maps generated in the second study. The map in Figure 4 was generated with data collected with the rank order (ISO) task; the map in Figure 5 was generated with data collected with the paired comparison task.

Note that three cities, Cincinnati, Boston, and Dallas, are in noticably different locations on the two maps. In general, the rank order map (Figure 4) appears to be a better representation of an actual map.

Table 5 shows the correlations between the interpoint distances in the two cognitive maps and an actual U.S. map. The correlation between the two cognitive maps is 0.91 which is quite close to the group average of 0.90 (or 0.92 if one subject is dropped from the analysis). Also note that the correlation between the rank order cognitive map and the U.S. map is higher than the correlation between the paired comparison cognitive map and the U.S. map. This is representative of the other subjects.

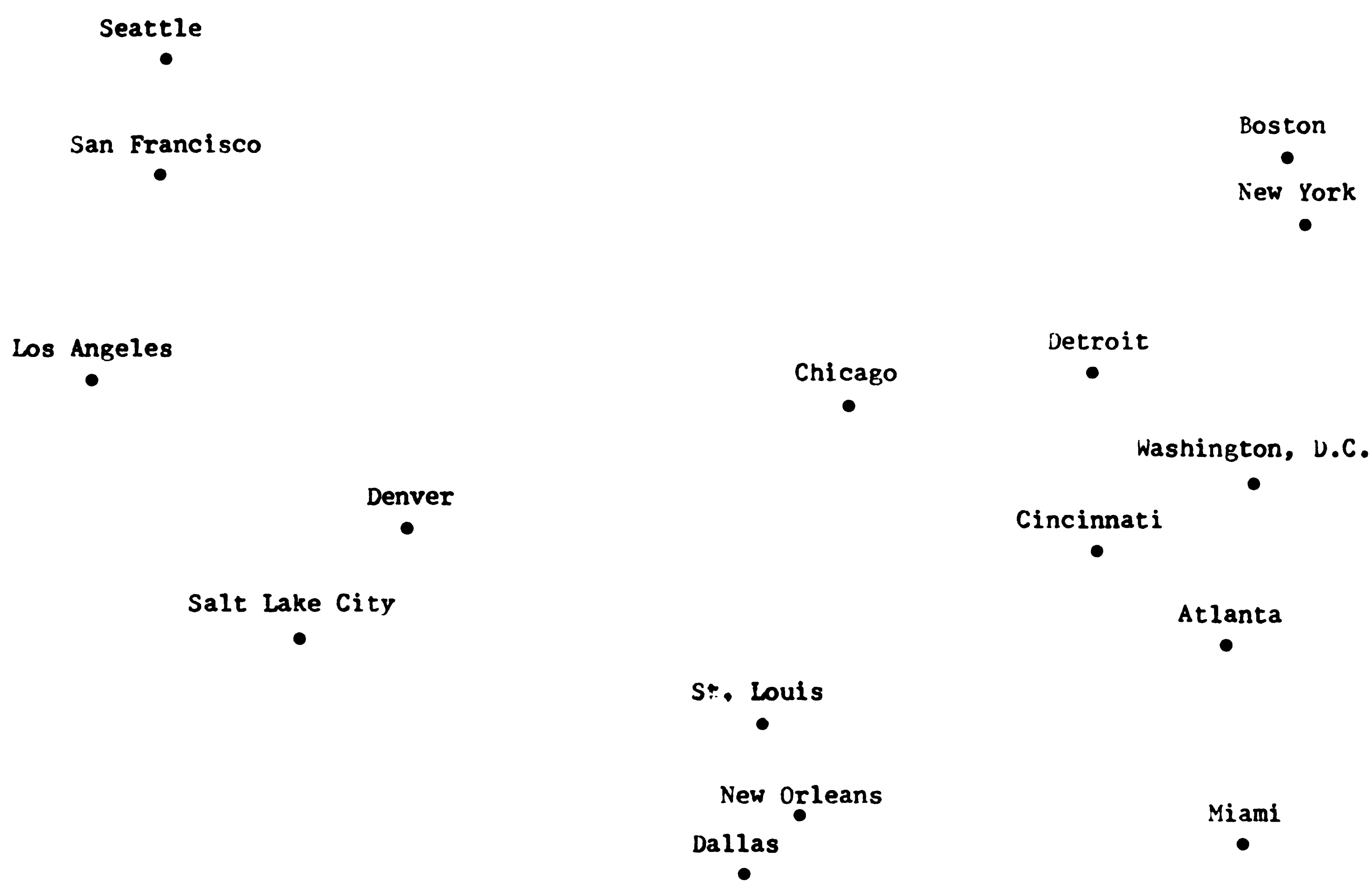

Figure 4. An example of a rank order cognitive map.

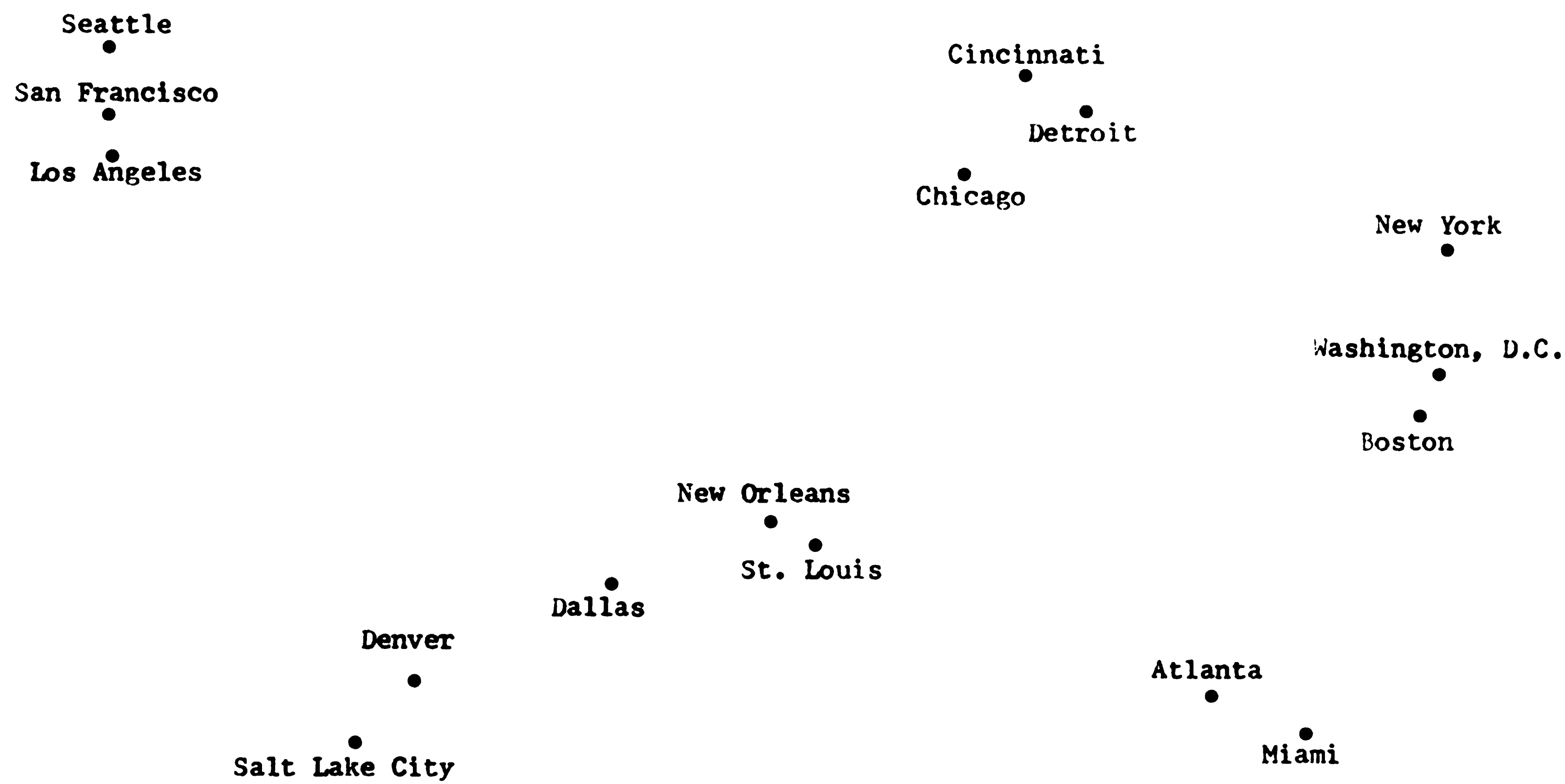

Figure 5. An example of a paired comparison cognitive map.

Table 5. The correlations between cognitive maps.

	Rank Order Distances	Pairwise Distances	Actual Distances
Rank Order Distances	1.00		
Pairwise Distances	0.91	1.00	
Actual Distances	0.94	0.88	1.00

LIST OF REFERENCES

LIST OF REFERENCES

Girrard, R. and Cliff, N. A Monte Carlo evaluation of interactive multidimensional scaling. Psychometrika, 1976, 41, 43-64.

Graef, J. and Spence, I. Using distance information in the design of large multidimensional scaling experiments. Psychological Bulletin, 1979, 86, 60-66.

Henley, N. M. A psychological study of the semantics of animal terms. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 176-184.

Klahr, D. A Monte Carlo investigation of the statistical significance of Kruskal's nonmetric scaling procedure. Psychometrika, 1969, 34, 319-330.

Knuth, D. E. Sorting and searching. The Art of Computer Programming (Vol. 3). Reading, Mass.: Addison-wesley, 1973.

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 1964, 29, 1-27. (a)

Krushal, J. B. Nonmetric multidimensional scaling: A numerical method. Psychometrika, 1964, 29, 115-129. (b)

Levine, D. M. A Monte Carlo study of Kruskal's variance based measure on stress. Psychometrika, 1978, 43, 307-315.

Ling, R. F. A probability theory of cluster analysis. Journal of the American Statistical Association, 1973, 68, 154-164.

Ramsay, J. O. Some statistical considerations in multidimensional scaling. Psychometrika, 1969, 34, 167-182.

Rao, V. R. and Katz, R. Alternative multidimensional scaling methods for large stimulus sets. Journal of Marketing Research, 1971, 8, 488-494.

Romney, A. K., Shepard, R. N., and Nerlove, S. B. (Eds). Multidimensional Scaling: Theory and Applications in the Behavioral Sciences. Vol. II. New York: Seminar Press, 1972.

Sherman, C. R. Nonmetirc multidimensional scaling: A Monte Carlo study of the basic parameters. Psychometrika, 1972, 37, 323-355.

Spence, I. A Monte Carlo evaluation of three nonmetric multidimensional scaling algorithms. Psychometrika, 1972, 37, 461-486.

Spence, I. Incomplete experimental designs for multidimensional scaling. In R. B. Colledge and J. N. Rayner (Eds.), Multidimensional Analysis of Large Data Sets. Columbus, Ohio: Ohio State University Press, (in press).

Spence, I. and Domoney, D. W. Single subject incomplete designs for nonmetric multidimensional scaling. Psychometrika, 1974, 39, 469-490.

Stenson, H. H. and Knoll, R. L. Goodness of fit for random rankings in Kruskal's nonmetric scaling procedure. Psychological Bulletin, 1969, 71, 122-126.

Takane, Y., Young, F. W., and de Leeuw, J. Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. Psychometrika, 1977, 42, 7-67.

Thurstone, L. L. A law of comparative judgment. Psychological Review, 1927, 34, 273-286.

Wagenaar, W. A. and Padmos, P. Quantitative interpretation of stress in Kruskal's multidimensional scaling technique. British Journal of Mathematical and Statistical Psychology, 1971, 24, 101-110.

Young, F. W. Nonmetric multidimensional scaling: Recovery of metric information. Psychometrika, 1970, 35, 455-473.

Young, F. W. and Cliff, N. Interactive scaling with individual subjects. Psychometrika, 1972, 37, 385-415.

Young, F. W. and Null, C. H. Multidimensional scaling of nominal data: The recovery of metric information with ALSCAL. Psychometrika, 1978, 43, 367-379.

Young, F. W., Null, C. H., and Sarle, W. Interactive similarity ordering. Behavior Research Methods & Instrumentation, 1978, 10, 273-280.