DATA COLLECTION EFFECTS ON NONMETRIC MULTIDIMENSIONAL SCALING SOLUTIONS

MARIE ADELE HUMPHREYS
University of Oklahoma

This study investigates four widely used methods of collecting direct similarity judgments for nonmetric multidimensional scaling: Rating Scale, Triads, Rank Order of Pairs, and Conditional Rank Order. The results indicate that: (1) when class structure is very striking in the attributes to be scaled, respondents under the Rating Scale method place significantly more weight on class structure than on continuously varying attributes in determining similarity; (2) nonmetric multidimensional scaling solutions account for a greater proportion of the variance in Rank Order of Pairs data than in data collected by the other methods studied, although the differences are so small that they would have little operational impact; and (3) the four methods do not differ significantly in test-retest reliability.

RESEARCHERS have generally proceeded on the assumption that the effects of data collection methods on multidimensional scaling (MDS) solutions are operationally equivalent. However, the results of some studies have cast doubt on this assumption. For example, in a study by Wish (Wish and Carroll, 1974), different dimensions were elicited when different data collection methods were used. And Green and Carmone (1970) found that a dimension present under one data collection method did not appear when other data collection methods were used.

Previous studies have directly addressed the question of data collection effects on multidimensional scaling configurations (e.g. Green, Maheshwari, and Rao, 1969; Green and Rao, 1969; Kinnear and Taylor, 1974; Rao, 1974; Sherman, 1972; Taylor and Kinnear, 1971; Whipple, 1976). However, in these studies the perceived dimensionality of the scaled objects was unknown to the research-

ers, potentially large for a single respondent, and often highly variable among respondents. When the attributes of the objects which will be perceived and used by the respondents in making similarity judgments are not known, a priori, by the researcher, then additional procedures (e.g., considerations of stress values, proportion of variance accounted for by additional dimensions, and interpretability) must be used by the researcher to determine the dimensionality of the MDS solution and sometimes to determine the orientation of the axes. Since configurations obtained from scaling data in different numbers of dimensions can be vastly different (Torgerson, 1965), these studies contain a residual uncertainty about whether data collection methods differentially affected the recovery of the perceived dimensions. This uncertainty may confound the results of the primary analysis of differences due to data collection techniques. In addition, in some of the studies responses were averaged across respondents. Thus, differences due to experimental conditions were confounded by the averaging process (See Horan, 1969 and Fisher and Micko, 1972 for a discussion of averaging effects). Therefore, the findings in previous studies are of limited usefulness to the researcher who must choose among data collection methods.

This study was undertaken to provide some carefully controlled experimental evidence with respect to four widely used methods for collecting direct similarity judgments for nonmetric MDS: Triads, Rating Scale, Rank Order of Pairs, and Conditional Rank Order.[1] Although the methods are directly suitable only for small to moderate size data sets, the cognitive tasks represented by these four methods are very similar to those employed in several data collection methods for large stimulus sets (Rao and Katz, 1971), and thus, findings may be applicable to other data collection methods as well.

The questions asked were: (1) Is the same pattern or configuration of objects obtained in a nonmetric MDS solution regardless of which data collection method is used? (2) Is variance accounted for in scaling the same among data collection methods? (3) Are data collection methods reliable in the sense of repeatability?

Research Design

Choice of Stimuli

The attributes of the stimuli were designed to meet two requirements: (1) the perceived dimensionality of the solution configuration

---

[1] Appendix 1 contains a short description of these data collection methods and the criteria used in their selection.

should be known, a priori, and (2) the attributes of the stimuli should exhibit both class structure and continuous variation. The first requirement was most crucial since it assured the researcher that all subjects would use the same attributes of the stimuli in making their similarity judgments, and therefore, that the pattern of responses could be unambiguously interpreted.

Research has shown that respondents often make reliable similarity judgments that reflect the quantitative difference between class and continuous variation and that nonmetric MDS procedures can appropriately be used to recover both types of variation (Aiken, 1972; Degerman, 1970; Green and Rao, 1969). Since both types of variation are often found in stimulus sets chosen for nonmetric MDS, the attributes of the objects scaled here were selected to exhibit both class structure and continuous variation, thus allowing a test of how different data collection methods might differentially affect the recovery of each type of variation.

The stimuli used in this study were triangles and rectangles varying in three sizes and four levels of brightness (shades of grey). All triangles were isosceles with base equal to height, and the corresponding rectangles were of the same height and area as the triangles.

In the stimuli described above, class variation is present in the attribute of shape (triangle and rectangle), and continuous variation is present in the attributes of brightness and size. Furthermore, previous research has shown that these attributes are unidimensional and comparably perceived by all subjects. Attneave (1950) and Degerman (1970) have both used stimuli similar to those chosen in this research and have shown that, when objects such as squares and triangles varying in area and brightness are scaled, the a priori dimensions of size, brightness, and shape may be assumed.

If all combinations of size, brightness, and shape were used, there would have been 24 (2 × 3 × 4) stimuli. Out of the 24 possible, 12 were chosen for this study. The number of stimuli used was determined by two requirements. First, the tasks resulting from the data collection methods should not be ones which would either overload the information processing capability of the subjects or result in data collection times for each method being so long that subjects could not complete each method in one continuous time period without undue fatigue, since either of these conditions might have introduced unacceptable error levels in the responses.

With twelve stimuli, the estimated task completion times for any method did not exceed one to one and one-half hours. Further, it was felt that, with twelve stimuli, neither the number of judgments nor the complexity of the judgments were excessive.

The second requirement was that, in order to justify the scaling solution to be used in the comparisons, a combination of dimensions and number of stimuli which would result in sufficient metric determinacy of the solution should be selected. Based on the work of Young (1970), it was determined that since the error in the research configuration should be fairly low, sufficient metric determinancy of the solution was assured with the selected combination of three dimensions and twelve stimuli.

From the 24 possible combinations of size, brightness and shape, 12 were chosen such that the resulting correlations between the solution dimensions would be minimized. The configuration used in this study is shown in Figure 1. (The levels of brightness correspond to Zip-a-tone 95 line screens of 10%, 20%, 50%, and 70%.)

Data Collection

The subjects were 56 students enrolled in undergraduate and graduate business classes and in an introductory psychology class. The students were volunteers who were paid for participating in the experiment.

Fourteen respondents gave similarity responses under each data collection method. The order of presentation for both the Triads (T) and Conditional Rank Order (CR) methods was randomized before the experiment. In the Rating Scale (RS) and Rank Order of Pairs (RP) methods, all pairs of stimuli were initially presented to the respondents in the order described by R. T. Ross (1934).

A familiarization period was added to the experimental procedures before actual similarity judgments were collected in order to
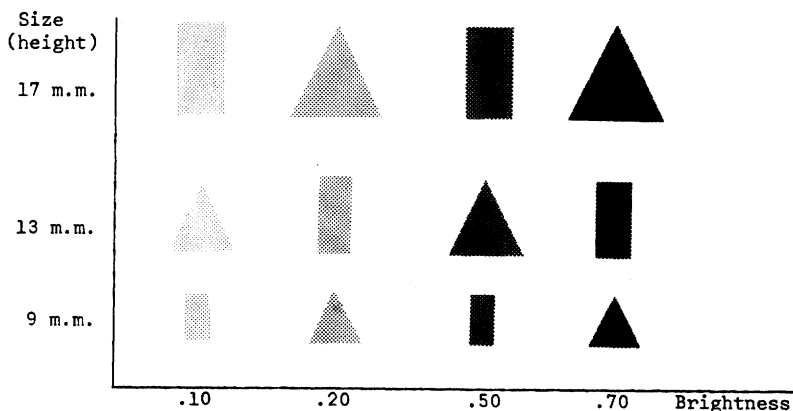

Figure 1. Configuration of Stimuli

eliminate the effect of an initial shift in response patterns which was noticed during a pilot study.

In the Rating Scale method, a continuous scale was utilized to alleviate the problem of degeneracy, which can occur when the proximity measures contain a large proportion of tied values. In the Rank Order of Pairs method, respondents were instructed to follow a specific procedure, which involved a series of sorts of the pairs into piles of differing degrees of similarity, before performing the final complete ranking of pairs.

After one complete set of similarity judgments had been completed, a set of similarity judgments sufficient to provide a rank ordering of eleven pairs of stimuli was repeated to provide data for reliability tests. Then subjects rated each stimulus figure on a light-dark scale and a large-small scale to provide data for a check of the MDS solution axes.

In order to obtain a complete rank order of all stimulus pairs for input to the scaling algorithm, the Triads and Conditional Rank Order responses were preprocessed by means of the TRICON algorithm (Carmone, Green, and Robinson, 1968) to remove intransitivities and obtain implied disjoint comparisons. After preprocessing, a complete proximity matrix was available for each of the subjects.

Data Analysis

Spatial Analysis

All 56 complete proximity matrices were simultaneously scaled in three dimensions using NINDSCAL (Carroll, 1970; Carroll and Chang, 1971). An examination of stress values and an examination of the projections of the scaled objects on the axes of the derived configuration were made to verify that the three-dimensional solution resulting from the NINDSCAL analysis was the correct solution for interpretation.

The stress value for the three-dimensional solution is .0113, indicating an excellent fit of the solution to the input values. In addition, the three-dimensional solution explains 96.11% of the total variance in the original data. Both results support the choice of three dimensions for scaling.

To verify the interpretation of the derived axes, a linear regression was used to locate directions in the stimulus space which optimized the linear correlation between a fitted vector and scale values of stimulus attributes (or properties.) The algorithm PROFIT (Chang

and Carroll, n.d.) was used to perform this analysis. The average rating of size and brightness given by the subjects for each object was used as an independently determined measure of that property. The correlation between the first fitted vector and the property of brightness is .9767, and the direction cosine of the vector with axis two is .9949. The correlation between the second fitted vector and the property of size is .9892, and the direction cosine of this vector with axis three is .993.

The above analysis plus a visual check of plots of the group stimulus space confirmed that the orientation of the solution provided by the NINDSCAL derived axes could be interpreted without rotation in terms of the actual attributes of shape, brightness, and size.

The range of correlations between each subject's unique solution distances and original data for each collection method are shown in Table 1. The high correlations show that the NINDSCAL analysis accounts satisfactorily for every subject's data.

Data Transformation

Since the above spatial analysis confirmed the *a priori* interpretation of the configuration, the weights in the subject space were transformed to separate the effects of variance explained and dimensional salience (Carroll and Wish, 1974, p. 433; Humphreys, 1978) as follows:

$$r_i = (w_{i1}{}^2 + w_{i2}{}^2 + w_{i3}{}^2)^{1/2} \qquad (1)$$

$$\theta_i = \arccos(w_{i1}/r_i) \qquad (2)$$

$$\phi_i = \arctan(w_{i2}/w_{i3}) \qquad (3)$$

where,

(w_{i1}, w_{i2}, w_{i3}) is the vector of weights obtained from the NINDS-CAL subject space for respondent i (w_{it} is the weight given by respondent i to dimension t of the group stimulus space);

TABLE 1
Correlations between Unique Solution Distances and Original Data

Method	Range	Mean
Conditional Rank Order	.945–.997	.979
Rank Order of Pairs	.955–.998	.989
Rating Scale	.947–.994	.977
Triads	.939–.996	.976

r_i^2 is the proportion of respondent i's variance[2] explained by the NINDSCAL analysis;

θ_i is an inverse measure of the variance explained by shape (w_{i1}^2) relative to the total variance explained by the analysis.

and, θ_i is a direct measure of the variance explained by the brightness (w_{i2}^2) relative to the variance explained by size (w_{i3}^2)

For the purpose of this research, the vector $P_i = (r_i,\ \theta_i,\ \theta_i)$ was interpreted as a description of the response pattern of subject i under the data collection method used by that subject.

Results

Response Pattern Analysis

Using the algorithm *Multivariance* (Finn, 1976), a multivariate analysis of variance with vector P as the dependent variable was used to test for differences in response patterns among the data collection methods. Prior to the analysis, univariate frequency plots of r, ϕ and θ were used to check the normality assumption. Subsequently, the Fisher r to z transformation was used to normalize the r measures. In addition, a test of the null hypothesis of equality of variances was performed using the test criterion of Box as presented in Cooley and Lohnes (1971). The null hypothesis was accepted at the .05 level of significance ($F = 1.57$; $df = 18,\ 955$).

The omnibus null hypotheses that no differences exist in response patterns to the four data collection methods, i.e.,

$$H_O: P_{RP} = P_{CR} = P_T = P_{RS}$$

was tested by the Wilks lambda criterion. The test statistic is obtained from an F transformation given by Rao, which is an approximation for the distribution of lambda. The null hypothesis was rejected. In order to isolate those comparisons between variables which were responsible for differences, univariate F tests were performed on each variable. The univariate tests showed that variables r and θ merited interpretation as the cause of the differences. Table 2 summarizes these results.

Since the omnibus F test was significant, and the univariate F tests were significant for the variables r and θ, the interpretation of t

---

[2] See Appendix 2 for a description of variance explained in INDSCAL solutions and an estimate of the effect of nonorthogonality.

TABLE 2
Multivariate and Univariate Tests of Significance

	Variable	F-ratio	P-value
Multivariate test	P	4.8255^a	0.0001
Univariate tests	r	3.6284^b	0.0188
	ϕ	1.7412^b	0.1701
	θ	3.1424^b	0.0329

[a] Degrees of Freedom, 9 and 121.8.
[b] Degrees of Freedom, 3 and 52.

tests for pairwise comparison of those variables between groups was justified (See Bock, 1975, pp. 422–24). In order to provide independent t tests, orthogonal contrasts in which each treatment effect was contrasted with the mean of succeeding treatment effects were used (See Finn, 1974, p. 232). The selection of specific contrasts and direction of tests were guided by informal hypotheses concerning possible response patterns which the author formulated prior to the study. Due to the lack of previous research, the hypotheses were speculative in nature and therefore were not advanced as formal hypotheses in this paper. However, the author's reasoning is presented as a possible interpretation of the findings in the discussion of results.

Table 3 and 4 present the results of the pairwise comparisons. The test results shown in Table 3 indicate that there is no statistically significant difference in the total variance explained for the CR, T, and RS methods, and the total variance explained for the RP method is greater than for any other method studied.

The test results shown in Table 4 indicate that there is not a statistically significant difference in the proportion of total explained variance accounted for by shape for the CR, RP, and T methods, and the proportion of explained variance accounted for by shape is greatest for the RS method.

Reliability Analysis

For each subject, eleven ranks obtained from retest similarity judgments were correlated with the corresponding ranks obtained

TABLE 3
Results of Comparisons on r

Contrast	t-statistic	P-value
$r_T - r_{RS}$	−0.007	$.9920^a$
$r_{CR} - (1/2)\,(r_T + r_{RS})$	0.673	$.6028^a$
$r_{RP} - (1/3)\,(r_{CR} + r_T + r_{RS})$	3.230	$.0006^b$

[a] Two-tailed test.
[b] One-tailed test.

TABLE 4
Results of Comparisons on θ

Contrast	*t*-statistic	*P*-value
$\theta_{RP} - \theta_T$	−0.775	.4383[a]
$\theta_{CR} - (1/2)(\theta_{RP} + \theta_T)$	−0.302	.5840[a]
$\theta_{RS} - (1/3)(\theta_{CR} + \theta_{RP} + \theta_T)$	−2.956	.0016[b]

[a] Two-tailed test.
[b] One-tailed test.

from the complete set of judgments. A summary of the reliability data is shown in Table 5. The Spearman rank correlation coefficients were transformed by the Fisher *r* to *z* transformation for the purpose of obtaining a more symmetric distribution and analyzed by a one-way analysis of variance. The results support the null hypothesis of no differences in test-retest reliability among the four data collection methods (*p*-value = .1974).

Discussion of Findings

Total Variance Explained

There are three reasons why the NINDSCAL analysis may not explain all of the variance in a respondent's data. First, the respondent's similarity judgments may contain random error. Second, additional idiosyncratic dimensions may be necessary to fully account for the respondent's judgments. That is, the respondent may be responding reliably to a set of dimensions which is not completely contained in the set of dimensions of the group space chosen for interpretation. Third, the NINDSCAL model may not account for the differences in the spacing or ordering of stimuli by this respondent on the dimensions of the group space.

Because of the careful choice of stimuli to be scaled in this study, all respondents should have perceived and used the same atttributes of the objects, i.e., shape, size and brightness, in making their similarity judgments. Further, the ordering of the stimuli with respect to size and brightness should have been the same for all respondents, and the spacing of the projections of the stimuli with

TABLE 5
Test-Retest Reliability Measures

Data Collection Method	Spearman Rank Correlation	
	Range	Mean
Rating Scale	.72–.95	.87
Rank Order of Pairs	.43–.98	.81
Conditional Rank Order	.58–1.00	.85
Triads	.81–.99	.90

respect to size and brightness should have been similar. Based on this analysis, the most probable cause for unexplained variance in the analysis would seem to be random error.

The finding that the total variance explained for the Rank Order of Pairs data is significantly greater than for the other methods studied may thus be interpreted as implying that random errors will occur less often with the Rank Order of Pairs method. One reason this may occur is that, because the respondent sees each pair of stimuli and its relation to other pairs being ranked many times while ranking the pairs of objects in the Rank Order of Pairs method, the respondent has time to formulate a decision rule and use it with more precision and consistency than in any of the other methods. In all other methods, it may be more difficult for the respondent to check with previous judgments in an attempt to impose some consistency in criteria used to make the similarity judgments.

Variance Accounted for by Shape

When similarity judgments are obtained for a set of stimuli in which only class variation is perceived by the subject, the configuration resulting from MDS should exhibit a structure where the m class prototypes are located at the corners of the geometrical m-simplex in m-1 dimensions (Degerman, 1972). Psychological distances underlying similarity judgments which are restricted to a consideration of only class structure may tend to be dichotomized by the subject as a zero distance perceived between pairs of stimuli in the same class and equal non-zero distances perceived between pairs of stimuli in different classes.

Moreover, responses of subjects under the Rating Scale method are easily dichotomized, with pairs of stimuli containing only one class being rated in one half of the scale and those pairs of stimuli containing two classes being rated in the other half.

Therefore, when stimuli scaled by the Rating Scale method exhibit both continuous variation and strong class variation, as they do in this study, it is possible that a tendency to dichotomize the psychological distance, due to the strong class variation, may be reinforced by the ease of dichotomization of the scale. The result may be a tendency for the subject to place more weight on class differences than on continuous variation than would be the case in assessing similarity under other data collection methods studied.

The rationale presented above might apply equally well to any situation where there is a dimension present on which the stimuli take just two values, even if it is a dimension in terms of a property

which could vary continuously. There is no evidence in this specific study to confirm or deny such a hypothesis, although the findings here suggest the need for further investigation.

In addition, the hypothesized explanation for the emphasis on class structure dimensions when using Rating Scales may appear only when class structure is very striking, as in this study. It would be of interest, therefore, to investigate different types of objects, in which class structure is weaker in relation to the continuously varying attributes which are present, to determine if different results might be obtained under those conditions.

Patterns of Variation

In light of the results above, it is interesting to look at the patterns of average total variation explained by each dimension of the group stimulus space as a proportion of average total variance explained by each of the data collection methods. These proportions are shown in Table 6.

The values in the first column of Table 6 show that the average total variance explained is highest for the Rank Order of Pairs method. However, the difference between the highest and lowest values is only .0245. Therefore, the differences among methods should have little operational impact on nonmetric MDS results.

However, the differences in reliance on shape between the Rating Scale method and the other three methods are both statistically significant and of large relative magnitude. This finding may have important implications for the researcher who must choose among data collection methods. Many researchers rely on measures of stress or proportion of explained variance to determine the dimensionality of the solution to be interpreted when they are investigating stimuli whose dimensions are not known a priori. The use of a

TABLE 6
Proportion of Average Total Variance Explained

| Method | Average Total Variance Explained | Proportion by Dimension For Each Method* | | |
		Shape (1)	Brightness (2)	Size (3)
Conditional Rank Order (CR)	.9569	.6193	.2268	.1540
Rank Order of Pairs (RP)	.9765	.6294	.2558	.1148
Rating Scale (RS)	.9511	.8053	.1089	.0858
Triads (T)	.9523	.5549	.2336	.2115

* Each proportion is calculated by dividing the average proportion of variance explained by dimension k of method j by the average proportion of total variance explained under method j.

TABLE 7
Factors Affecting Choice of Data Collection Methods

Method	Preparation of Stimuli Required	Average Response Time (12 stimuli)	Preprocessing Required	Response Pattern
Conditional Rank Order	N sets of N stimuli	23 min.	YES	
Rank Order of Pairs	$N(N-1)/2$ stimuli pairs	37 min.	NO	Greatest explained variance
Triads	$N(N-1)(N-2)/6$ stimuli triads	37 min.	YES	
Rating Scale	$N(N-1)/2$ stimuli pairs	18 min.	NO	Most reliance on class difference

method in which significantly more weight is placed on class differences, with the resultant lower proportion of variation accounted for by continuously varying attributes, may result in interpretable dimensions of difference being excluded from the analysis.

Conclusions

The choice of data collection method for a specific application is governed by a consideration of several factors including ease of administration, cost of obtaining the data (in both time and money), amount of preprocessing that is required, validity of the assumptions underlying the method, reliability and accuracy of subjects over different methods, and the effect, if any, that use of the method may have on the scaling solution.

In this study, no statistically significant differences were found among methods in immediate test-retest reliability. However, the results concerning response pattern differences, to the extent that they can be generalized to other settings, imply that the researcher is faced with tradeoffs in selecting an appropriate data collection method. Some of the major considerations in this regard are summarized in Table 7.

It would seem that the method of Triads is the least desirable method of collecting similarity judgments, requiring the preparation of more stimuli, the largest average response time[3], the greatest

---

[3] A Henry and Stumpf study (1975) shows that the average time necessary for the Triads method rises much more rapidly than for the Rank Order of Pairs method when the stimuli number more than about twelve.

amount of preprocessing time to convert the responses to a proximity matrix[4], and producing no more variance accounted for in scaling than the other three methods.

On the other hand, the Rating Scale method, which is easy to administer, is less costly than the other methods in terms of time and cost of obtaining data, and requires no preprocessing of data to obtain the complete proximity matrix, may not be the preferred method. The findings of this study imply that if class structure is very striking, respondents providing Rating Scale data may place a heavy reliance on class differences in determining their similarity responses. Since this tendency may result in interpretable dimensions being excluded from analysis, the researcher may wish to consider the use of either the Conditional Rank Order or the Rank Order of Pairs methods.

The Conditional Rank Order method might provide a reasonable compromise between time and cost considerations and the desire to have sources of variation more equally represented in the data. However, if subject response time considerations are not prohibitive, the Rank Order of Pairs method might be the most desirable method since it requires preparation of less stimuli than the Conditional Rank Order method, and the similarity judgments result directly in a complete proximity matrix for each subject without preprocessing.

APPENDIX 1
DATA COLLECTION METHODS

Descriptions

Triads Method

In the method of Triads, the respondent is presented with all possible groups of three of the objects in the stimulus set. For each triad, the respondent is asked to indicate which pair is most similar and which pair is least similar.

In this study, in order to assist the respondent in assessing similarity, each triad was presented in the form of three pairs. For example, if the triad contained objects A, B, and C, the respondent was presented with the three pairs: AB AC BC, and asked to designate the most similar and the least similar pair.

---

[4] Triads responses must be converted to a 0-1 matrix before being processed by the TRICON algorithm, unlike Conditional Rank Order responses which can be input directly to the TRICON algorithm.

Conditional Rank Order Method

In the method of Conditional Rank Order, the respondent is presented with one of the objects to be scaled, as the anchor or reference object, and asked to rank the other N-1 objects in order of decreasing similarity to it. Another object from the set is then presented as the reference object, and the respondent repeats the task. N sets of conditional rankings are obtained from the respondent in this manner.

Rank Order of Pairs Method

In the method of Rank Order of Pairs, the respondent is presented with all pairs of the N objects and asked to rank them in order of decreasing relative similarity.

Researchers have reported that respondents expressed a high level of frustration when given this task with no further instructions (see e.g., Shepard and Chipman, 1970). Therefore, in this study a technique described by Taylor and Kinnear (1971) was used in which the respondent was instructed to initially perform a series of sorts of the pairs into piles of differing degrees of similarity before performing the final complete ranking of pairs.

Rating Scale Method

In the method of Rating Scale, the respondent is asked to rate on a scale the relative similarity of each pair of objects in the stimulus set, where the extreme values on the scale represent maximum dissimilarity and maximum similarity.

Criteria for Selection of Data Collection Methods

Two criteria were used in determining what data collection methods to include in this study. First, the data collection methods used should be those with which it is possible to obtain a complete proximity matrix for each respondent with as few ties in the data as possible, in order to reduce the possibility of degenerate solutions due to ties in the data, and in order to avoid confounding the results through averaging responses over respondents or through the presence of incomplete proximity matrices. This criterion eliminated, for example, methods of data collection based solely on sorting the stimuli into unlabelled clusters.

Second, data collection methods incorporating response formats

which are widely used should be selected, so that the results of the study will be applicable to as wide a range of research situations as possible. This criterion eliminated, for example, the Dyads method of data collection. That method is not widely used, primarily because it requires many more judgments than other available methods.

APPENDIX 2
INTERPRETATION OF "VARIANCE EXPLAINED"

The subject weights in NINDSCAL are simply the regression weights obtained in a multiple regression predicting the subject's data from some linear combination of products of the stimulus space dimensions. Thus, the square of the regression weights can be strictly interpreted as variance components only when the predictor variables are orthogonal. The design of the stimulus space in this study did not provide strict orthogonality because of the selection of only half of the possible triangles and rectangles (i.e., dimension one of the solution). However, as demonstrated below, the design does *not* provide a degree of departure from orthogonality such that the results of this study would be significantly affected.

The total variance in the least squares solution matrix for subject i, V_i, is given by the expression (Carroll and Chang, 1970):

$$V_i = \sum_j \sum_k \left(\sum_{t=1}^{3} W_{it} X_{jt} X_{kt} \right)^2$$

$$i = 1, \cdots, 56; \quad j, k = 1, \cdots, 12; \quad t = 1, \cdots, 3 \qquad (1)$$

where X_{jt} is the projection of stimulus object on dimension t in the group space, and W_{it} is the weight that subject i gives to dimension t of the group space in deriving the subject's "private" space.

An expansion of equation (1) shows that the effect of deviations from orthogonality can be evaluated by estimating the effect of cross product terms involving dimension one of the solution.

$$2W_{i1} W_{i2} \sum_j X_{j1} X_{j2} \sum_k X_{k1} X_{k2} + 2W_{i1} W_{i3} \sum_j X_{ji} X_{j3} \sum_k X_{k1} X_{k3}$$

The sum of products of interest are:

$$\sum X_{ji} X_{j2} = \sum X_{k1} X_{k2} = .04090 \qquad (2)$$

$$\sum X_{j1} X_{j3} = \sum X_{k1} X_{k3} = .04098 \qquad (3)$$

where (2) is the correlation between the projections of stimuli on dimension one (shape) and the projections on dimension two (brightness) of the group stimulus space and (3) is the corresponding correlation between the dimensions of shape and size.

If, in order to place an upper bound on the effect on total variance of these deviations from orthogonality, we assume that the subject's weight on each dimension is .707 (recognizing this to be impossible), the total effect is (substituting into equation 1):

$$2(.5)(.04090)^2 + 2(.5)(.04098)^2 = .0034$$

Thus, an estimated upper bound of the effect of the deviation from orthogonality found in the stimuli set used in this study is that it would change the stated value of the variance accounted for by less than the amount .0034. In no case would that alter any of the findings or conclusions presented in this study.

REFERENCES

Aiken, L. S. Simultaneous processing of typal and dimensional variation among multidimensional events. *Multivariate Behavioral Research,* 1972, 7, 305–316.

Attneave, F. Dimensions of similarity. *American Journal of Psychology,* 1950, 63, 516–556.

Bock, D. R. *Multivariate statistical methods in behavioral research.* New York: McGraw-Hill, 1975.

Carmone, F. K., Green, P. E., and Robinson, P. J. TRICON—An IBM 360/65 FORTRAN IV program for the triangularization of conjoint data. *Journal of Marketing Research,* 1968, 5, 219–220.

Carroll, J. D. An overview of multidimensional scaling methods emphasizing recently developed models for handling individual differences. In W. King and D. Tigert (Eds.) *Attitude research reaches new heights.* Chicago: American Marketing Association, 1970.

Carroll, J. D. and Chang, J. Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckart-Young' decomposition. *Psychometrika,* 1970, 35, 283–319.

Carroll, J. D. and Chang, J. *A quasi-nonmetric version of INDS-CAL, A procedure for individual differences multidimensional scaling.* Unpublished paper, Bell Laboratories, 1971.

Carroll, J. D. and Wish, M. Multidimensional perceptual models and measurement methods. In E. C. Carterette and M. P. Friedman (Eds.) *Handbook of perception.* New York: Academic Press, 1974, 391–447.

Chang, J. and Carroll, J. D. How to use PROFIT, a computer program for property fitting by optimizing nonlinear or linear correlation. Murray Hill, N. J.: Bell Laboratories, (n.d.). (Mimeographed)

Cooley, W. W. and Lohnes, P. R. *Multivariate data analysis.* New York: John Wiley, 1971, 228–230.

Degerman, R. L. The geometric representation of some simple structures. In R. N. Shepard, A. K. Romney, and S. B. Nerlove, *Multidimensional scaling: Theory and applications in the behavioral sciences*. New York: Seminar Press, 1972, 196.

Degerman, R. L. Multidimensional analysis of complex structure: Mixtures of class and quantitative variation. *Psychometrika*, 1970, 35, 475–491.

Finn, J. D. *Multivariance*, Version V, Release 3, June 1976. Chicago: International Educational Services, 1976.

Finn, J. D. *A general model for multivariate analysis*. New York: Holt, Rinehart and Winston, 1974.

Fischer, W. and Micko, H. C. More about metrics of subjective spaces and attention distributions. *Journal of Mathematical Psychology*, 1972, 9, 36–54.

Green, P. E. and Carmone, F. J. Stimulus context and task effects on individuals' similarities judgments. In C. W. King and D. Tigers (Eds.) *Attitude research reaches new heights*. Chicago: American Marketing Association, 1970.

Green, P. E. and Rao, V. R. Configuration invariance in multidimensional scaling: An empirical study. *Proceedings*, Fall Conference, American Marketing Association, 1969, 182–187.

Green, P. E., Maheshwari, A., and Rao, V. R. Dimensional interpretation and configuration invariance in multidimensional scaling: An empirical study. *Multivariate Behavioral Research*, 1969, 4, 159–180.

Henry, W. A. and Stumpf, R. V. Time and accuracy measures for alternate multidimensional scaling data collection methods. *Journal of Marketing Research*, 1975, 12, 165–170.

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

Humphreys, M. A. A methodology for the identification of differences in perception as an aid in market segmentation analysis. *Proceedings of the American Institute for Decision Sciences*, November, 1978, Vol. I, 205–207.

Kinnear, T. C. and Taylor, J. R. The effect of proximity judgment collection procedures on individual differences scaling results: An empirical examination. *Combined proceedings*, Spring and Fall conferences, American Marketing Association, 1974, 162–163.

Rao, V. R. Multidimensional analysis: Discussion. *Combined proceedings*, Spring and Fall Conferences, American Marketing Association, 1974, 164–166.

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

Ross, R. T. Optimum orders for the presentation of pairs in the method of paired comparisons. *Journal of Educational Psychology*, 1934, 25, 375–382.

Shepard, R. N. and Chipman, S. Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology*, 1970, 1, 1–17.

Sherman, R. C. Individual differences in perceived trait relationships as a function of dimensional salience. *Multivariate Behavioral Research*, 1972, 7, 109–129.

Taylor, J. R. and Kinnear, T. C. Empirical comparisons of alternate methods for collecting proximity judgments. *Combined proceedings*, Spring and Fall Conferences, American Marketing Association, 1971, 547–550.

Torgerson, W. S. Multidimensional scaling of similarity. *Psychometrika*, 1965, 30, 379–393.

Whipple, T. W. Variation among multidimensional scaling solutions: An examination of the effect of data collection differences. *Journal of Marketing Research*, 1976, 13, 98–103.

Wish, M. and Carroll, J. Applications of individual differences scaling to studies of human perception and judgment. In E. C. Carterette and M. Friedman (Eds.) *Handbook of Perception*, Vol. II. New York: Academic Press, 1974, 449–491.

Young, F. W. Nonmetric multidimensional scaling: Recovery of metric information. *Psychometrika*, 1970, 35, 445–473.