SUBJECT ABILITIES TO USE METRIC MDS: EFFECTS OF VARYING THE CRITERION PAIR[1]

Thomas F. Gordon

As Schramm (1973) noted:

> The main effect communication has on us is on the pictures in our heads, our cognitive maps of environment, our images of ourselves, the beliefs and values we have accepted and are prepared to defend, the evaluations we have made of our relationships to individuals and groups ...(p. 194).

Multidimensional scaling (MDS) provides a tool for the creation and exploration of these cognitive maps and their related communication/cognitive processes. For mass communication research, the power of MDS rests with its ability to provide quantitative multidimensional models of cognitive processes which can be applied to a variety of mass media effects areas. In essence, examining an MDS plot representing relevant audience conceptions is like looking into the average mind of the listener/reader/viewer. Studying, in a precise quantitative sense, the interrelationships of these conceptions appears to hold exceptional potential for explanation, prediction, and understanding of communication related phenomena.

Approaches to MDS

Although a variety of scaling techniques are available for MDS, Shepard (1972a) notes that:

> The unifying purpose that these techniques share, despite their diversity, is the double one (a) of somehow getting hold of whatever pattern or structure may otherwise lie hidden in a matrix of empirical data and (b) of representing that structure in a form that is much more accessible to the human eye--namely, as a geometrical model or picture (p. 1).

The two major approaches to MDS are referred to as the "classical" or metric" approach and the "nonmetric" approach. These terms characterize the level of measurement of the data used in the scaling procedures. In the metric case interval or ratio data are required whereas in the nonmetric case ordinal data are sufficient. Building on the mathematical groundwork of Young and Householder (1938) and Richardson (1938), Torgerson (1951; 1952) is credited with developing the metric formulation of MDS and his book published in 1958 has become the classic statment. The first computer routines for nonmetric MDS of ordinal data were introduced by Shepard (1962a; b) and later refined by Kurskal (1964a; b).

Nonmetric vs. Metric: Advantages & Disadvantages

Nonmetric. In nonmetric MDS ordinal judgments of the similarity of a set of elements are obtained (this is the most common form of data collection although others are possible, see Shepard, 1972b for other types). These elements are then modeled by systematically changing their structural relationships (an iterative process) until they "fit" (show a linear or monotonic relationship with) the original ordinal data. The resulting positions of the elements are thus transformed from ordinal information to interval and, thru an additive constant, to ratio representations.

There are a variety of nonmetric MDS routines available. Of those using aggregate data (most relevant for mass media) Shepard (1972a, p. 8) notes:

> ...when they are applied to the same set of data, the various methods of this general type...usually yield virtually indistinguishable results. After all, monotonicity is a rather well-defined notion and any reasonable way of evaluating it could be expected in practice to lead to similar results.

Nonmetric advantages (Shepard, 1972a):

1. Metric representations (models) are derived from ordinal information.
2. The judgement task for the respondent is easier than in the metric case.
3. The method can be applied to situations where collection of metric information may be questionable.

Nonmetric disadvantages:

1. The model is produced through iterative (repeated) adjustment of the structure for a best fit with the original data. Since these data are ordinal, the model is still constrained to fit only the <u>orderings</u> of the judgements (Woelfel, 1975).
2. There should be at least two to three times as many elements to be structured as there are dimensions if an optimum fit with the original data is to be obtained (see Sherman, 1972; Spence, 1972; Young, 1970). Thus, to decide how many elements to use, some prior knowledge of the dimensionality of the structure would be beneficial (Danes and Woelfel, 1975).
3. "Stress" (lack of fit with original data) increases as the number of elements being scaled increases (Young, 1970); yet, the metric retrieval becomes more precise (Danes and Woelfel, 1975).
4. The iterative and algorithmic transformations require more time and expense than do the metric

routines.

Metric MDS. Metric MDS procedures elicit interval or ratio judgments of the degree of similarity of a set of elements. Since, in the aggregate case, the distances bwtween these elements are obtained as mean distances, they can be scaled directly with no need to go thru the iterative process of nonmetric MDS in order to obtain the final structure. In the interval data case, the additive constant transformation would be needed to obtain ratio data for the final structuring. If ratio judgments are initially elicited, no data transformations are required. Thus, the preferred case would be ratio judgments of the similarity in terms of distance. The following advantages and disadvantages are based on this paired comparison ratio metric approach. This approach begins by giving the subject a criterion or standard which can be used to make all other comparisons. For example, the subject is told that elements A and B are X units apart. Knowing this, they judge A from C, C from B, etc. for all pairs (see the methods section of this paper for complete instructions). The absolute zero point is established by instructing the subject to regard complete identity between any two elements (zero dissimilarity) as zero distance apart.

Metric advantages (Woelfel, 1974):

1. No information is lost in the creation of the structural whole from the original ratio data.
2. Since no iterations or transformations are needed, a definite cost-efficiency benefit is gained.
3. Comparison of structures across samples and/or over time is facilitated if a consistent ratio scaled standard of judgment is presented each time judgments are requested.
4. Relative to the aggregated ratio judgments, as opposed to interval or ordinal judgments:
   a. The scale is unbounded at the high end, continuous, and possesses an absolute zero point.
   b. As an aggregate or sample mean, these distances between elements are the best estimates of the true population conceptions.
   c. Error in mean estimates can be expected to be random, normally distributed about the mean, and inversely proportional to sample size.

Metric disadvantages:

1. Ratio scaled metric judgments are more difficult to make than ordinal judgments.
2. If judgments are more difficult, reliability should suffer. To overcome this larger sample

sizes may be needed.

3. In the ratio technique, criteria for the choice of a standard or criterion pair to which all other judgments are relative have yet to be empirically determined. Also, information about subject abilities to apply a given criterion standard in mapping out their cognitions is minimal.

On face value it would appear that the advantage of the metric approach would make it a clear favorite. However, the disadvantages should not be taken lightly. In the aggregate metric case disadvantages one and two above can be overcome by use of a sufficiently large sample size. The third disadvantage above is central to the application of the technique. The present study deals with this problem area.

The Present Study

The selection of a suitable standard or criterion pair on which all other judgments are based in the ratio metric approach appears crucial to the technique since this criterion becomes the ruler used by the respondent to measure their congitions for the researcher. The criteria available for the selection of this standard are minimal. Woelfel (1974a, p. 16) suggests the following:

> First, the standard should be relatively stable. Changes standard over time can confound time series measurements and prevent meaningful comparisons of measurements made at different times. Secondly, the standard should be the same for all observers regardless of reference point, i.e., two independent observers must both agree on the length, for example, of a meter or a kilometer. Less important, but nonetheless worthy of consideration, good practice for minimum error suggests using a standard approximately midway between the largest and smallest measurement likely to be encountered, (measurement of astronomical distances in miles, for example, is cumbersome, as would be measurement of terrestrial distances in fractions of light-years).

The present study builds upon the last of the above suggestions in assessing the respondent's ability to use standards of differing unit lengths (10, 25, 50, 100) involving pairs from the extremes of the concept domain or pairs closer together in that domain. At the same time, the findings of an earlier study by Gordon and De Leo (1975) are replicated.

The Gordon and De Leo study determined that providing a 10 unit criterion pair composed of the extremes from a homogeneous concept domain produced a structure which was statistically identical to using the concepts "red & white" as a criterion (10 units apart) and/or using only a ten unit scale base with no anchor concepts (with the option to make judgments larger than 10

units). The interpretation of those results suggested that for the red-white and no concept treatments, the only usable information for the subject was the 10 unit scale base since color was irrelevant to the concepts being judged. Thus, judgments were made with that scale base in mind and identical structures resulted. With the extreme concepts as the criterion, although these concepts were meaningful, the fact that they represented the extremes produced the same result by essentially limiting the judgments to that base.

The present study was designed to replicate the finding that the criterion pair consisting of the extremes from the concept domain will produce a structure identical to providing the same scale base with no concept anchors. As well, the study extends this to vary the distance between the concepts to further explore the subject's ability to use the criterion given.

Overall, it was expected that the same relationship as evidenced in the earlier study by Gordon and De Leo (1975) would hold true. Thus the prediction was:

> H_1: Given the same unit distance, the no criterion pair and the extreme concepts criterion pair will produce statistically identical structures.

Also, given the evidence concerning the reliability of metric MDS with aggregate data (see Barnett, 1972; Gillham and Woelfel, 1975), it was predicted that:

> H_2: As the distance between the criterion pair is increased, the resulting judgments of distances among concepts will increase but the pattern of concept interrelationships will remain the same.

In the Gordon and De Leo study a fourth criterion pair treatment was employed using two concepts somewhat closer together in the concept domain. The results of that comparison suggested that a criterion pair using concepts that are close together will produce an expanded structure since most of the judgments must be made outside (larger than) that distance base. Thus, the prediction was:

> H3: A criterion pair using concepts close together in the concept domain will produce an expanded structure (as compared to the extremes), yet the interrelationships of the concepts will remain the same.

Methodology

Concepts

As a methods study, the particular concepts used for the scaling comparisons were of secondary interest. The concepts selected were types of television programs and particular titles of shows related to each program type. These choices were made

on the reasoning that most subjects would be able to judge these concepts and, as such, maximum judgements would be obtained. Also, the clusterings of program titles with program types could be examined for internal content validity of the resulting structures. Six general program types and six related shows plus the concept "me" (self) resulted in a total of 13 concepts requiring 78 paired judgements. The particular TV shows were selected on the basis of having high ratings in a recent period (Broadcasting, p. 19).
A listing of the concepts follows:

1. Children's Comedy
2. Adult Situation Comedy
3. Soap Opera
4. Family Drama
5. Medical Drama
6. Crime Drama
7. Fat Albert
8. All in the Family
9. General Hospital
10. The Waltons
11. Medical Center
12. The Streets of San Francisco
13. Me

Treatment Conditions

To explore the effects of criterion pair variation and to partially replicate the findings of our earlier study (Gordon and De Leo, 1975), three major treatment variations were employed. First, the extremes of the concept domain were used and these were specified as differing distances apart for different treatment groups. Second, two concepts closer together in the domain were used, varying their specified distance. Third, no concepts were used but instead subjects were simply told, "As you judge the distances, keep a ten point scale in mind -- some shows may be less than ten units apart and others may be more." The nine specific treatment conditions (independent groups) were as follows:

1. None (no anchor concepts, only a 10 point scale base)
2. Children's Comedy - Crime Drama = 10 (CC10)
3. " " " = 25 (CC25)
4. " " " = 50 (CC50)
5. " " " = 100 (CC100)
6. Family Drama - Medical Drama = 10 (FM10)
7. " " " = 25 (FM25)
8. " " " = 50 (FM50)
9. " " " = 100 (FM100)

Subjects

A total of 863 students were the subjects in this experiment. The number in each treatment ranged from 92-112. The large number of subjects was deemed necessary so that the comparisons would be based on stable structures. Table 1 provides the n's for each treatment. The students were randomly assigned (by classes) to treatments. The departments sampled included

Anthropology, Education, Journalism, Psychology, Radio-TV-Film, Sociology, Speech, and Theatre.

Relative to the distribution of subjects to treatments (see Appendix A), Chi Square analyses on the media related variables of (1) average hours of TV/day, (2) average hours TV/week, and (3) preferred types of TV programs, the distributions of subjects in all treatment groups were equal. This is true also for the demographic variables of age, sex, and family income. On year in school, the random assignment of classes to treatments resulted in treatment CC50 having upper class students overrepresented--all other treatments were equal. On race, CC25 and CC10 had fewer Blacks than did the other treatments.

Procedures

The data were collected December 1-12, 1975. Subjects in classrooms were given one of the nine treatment variations and the following instructions were read with them:

> This form asks you to tell us how different (or in other words, how "far apart") TV shows are from each other. Difference between shows can be measured in units, so that the more different two shows are, the more units apart they are. To help you know how big a unit is, _____ and _____ are _____ units apart. You are supposed to tell us how many units apart the shows on the next few pages are. Remember, the more different the shows are from each other, the larger the number of units apart they are. Some shows may be more than _____ units apart and some may be less.
>
> Note that:
>
> -- "me" on the questionnaire means yourself. Judgements involving "me" should indicate how close you feel to that TV show or type of program.
>
> -- Zero can be used as a distance; if you see two things as identical, they would be zero distance apart.
>
> -- If you are not familiar with a TV show or type of program, leave that pair blank.
>
> Please work quickly. Judge the shows as pairs rather than trying to relate each judgement to all others.

Blanks in the above instructions were filled by the criterion pair used in a particular treatment. For the no criterion pair treatment the last sentence in the first paragraph of the instructions read, "As you judge the distances, keep a ten point scale in mind -- some concepts may be less than ten units apart and others may be more."

On the average, the items were completed in 15-20 minutes. Most of the subjects were able to judge the 78 pairs with the average number of subject judgements ranging from 83.87 to 106.74.

Results

Individual Treatments

Using version 3.0 of the GALILEO metric MDS program, treatments were first processed individually. Table 1 provides the percent of real distance accounted for by the three factors, the imaginary distance of the total solution, the trace values, and the average distance judgment for each treatment.

To avoid the effect of extreme values on the means, maximum values were set for each treatment using the maximum value option of the GALILEO program. These values were determined by successive runs in which the extreme values were gradually reduced while observing the minimum-maximum descriptive statistics as related to (a) the criterion pair given; (b) the means; and (c) the standard deviations. The maximum values finally used are provided in Table 1 along with the number of judgements that value excluded, and the average number of observations remaining per cell.

Given the scaled concepts, plots of the first three orthogonal factors were obtained using the plot option of the program. These plots are presented in Figures 1_a - 1_i. The high degree of similarity of inter-concept locations across the treatments is visually evident in these figures. Keeping in mind the mean distances, these plots show that the concept locations are highly similar for each treatment, while the actual mean distances between concepts differs considerably across treatments.

Comparison of Treatments

Correlations. To statistically verify the consistency of concept locations across treatments, the mean distances for the 78 pairs in each treatment were entered as scores into a standard Pearson product moment correlation. Thus, the pattern evident in the inter-concept distances could be compared across treatments. Table 2 presents the results of these intercorreclations. In all cases the correlations are extremely high, ranging from .933 to .988.

Of course, these exceptionally high correlations and the high degree of linearity evident from the plots suggest high predictability. For example, a simple two variable linear regression of the criterion pair distances as independent variables (within a CC or FM treatment set) on (1) the average distance judgment, (2) the trace values for each treatment, (3) the overall sum of squares for all matrix values in each treatment and (4) the average distance moved for the concepts in each treatment, as dependent variables produces r^2 (variance explained) values ranging from .966 to .999.

Figure 3 presents the comparison of spaces plot for only the CC10 and NONE treatments, both having a 10 unit base. It is obvious from this plot that hypothesis one (predicting statitically identical structures for CC10 and NONE) is confirmed. This result replicates the findings of our earlier

TABLE 1

Descriptive characteristics of treatments

Treatment	N	Max Value	Judge.* Excluded	Ave Obs. Per Cell	3 Factor % Real Dist. Acct. For	Cum. Percent Imaginary Dist.	Trace	Ave. Dist. Judgment
None	93	50	19	86.96	79.19	11.04	255.59	6.11
CC10	93	50	8	88.37	82.21	10.05	238.98	5.91
CC25	93	79	133	87.94	79.61	11.42	1479.64	14.58
CC50	92	250	44	86.91	81.25	11.22	6441.19	30.45
CC100	94	400	27	83.87	83.25	12.69	23137.37	57.00
FM10	112	100	33	106.74	79.71	17.32	951.15	11.60
FM25	96	145	135	89.59	81.43	17.66	4712.67	25.82
FM50	93	400	68	87.60	83.53	12.77	13772.84	44.43
FM100	97	600	111	90.32	82.03	10.67	45100.99	80.11
Total	863							

*The number of possible judgements in the treatments ranged from 7166 in CC50 to 8736 in FM10.

Concepts

1. Children's Comedy
2. Adult Situation Comedy
3. Soap Opera
4. Family Drama
5. Medical Drama
6. Crime Drama
7. Fat Albert
8. All in the Family
9. General Hospital
10. The Waltons
11. Medical Center
12. Streets of San Francisco
13. Me

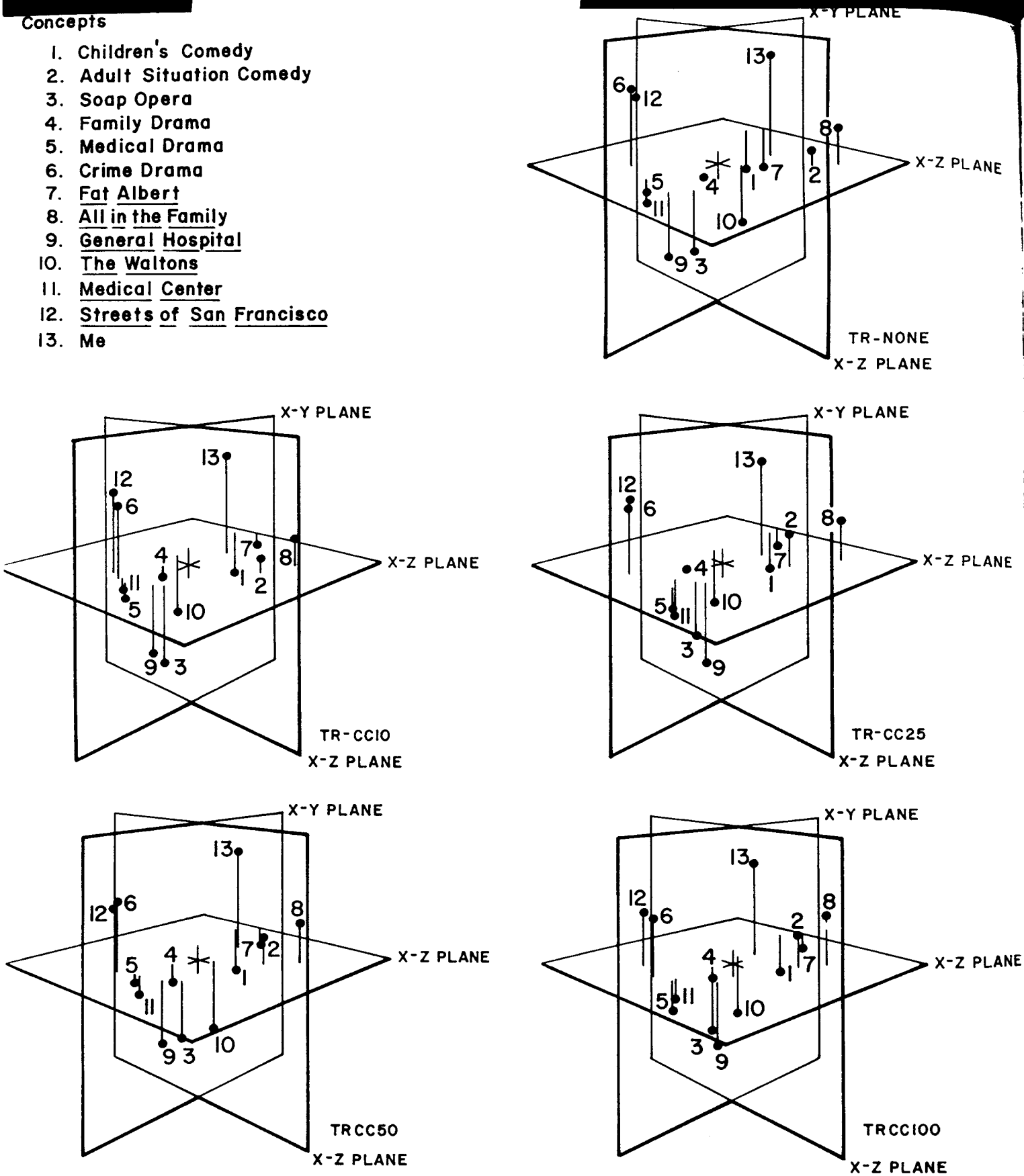


FIGURE I. INDIVIDUAL PLOTS OF EACH TREATMENT.

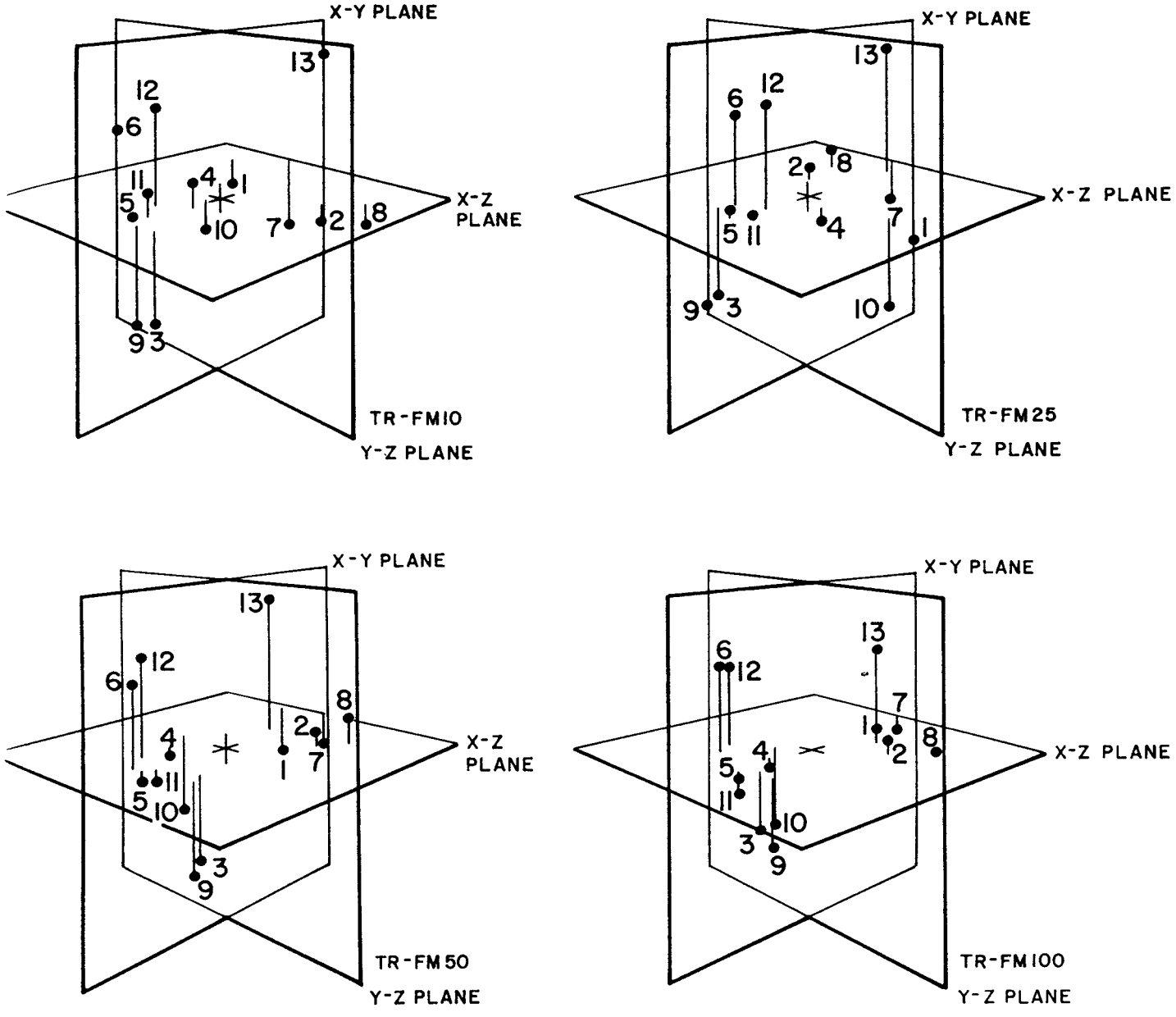


FIGURE I (cont.) INDIVIDUAL PLOTS OF EACH TREATMENT.

study (Gordon and De Leo 1975).

Comparison of spaces. Next, the treatments were compared using the Comparison of Spaces option of the GALILEO program. This option uses a least squares rotation of the axes of the individual treatment to one treatment specified as the mainspace. For this plot, the treatments were organized by trace to keep the plot as neat as possible, given the complexity of comparing nine data sets. Thus, FM100 was defined as the mainspace (this decision was made, rather than using CC10 as the mainspace, because the concept numbers are placed with the first plot and the CC10 plot would have placed 13 numbers into a very small space). The trace ordering was: CC10, None, FM10, CC25, FM25, CC50, FM50, CC100, FM100. Figure 2 shows the resulting plot and confirms the intercorrelation interpretation that the relative concept locations are highly similar for each treatment but the space expands as the criterion pair distance increases. As predicted, for the same given distance the criterion pair which is closer together produces a larger space than does the extreme pair.

Figure 3 presents the comparison of spaces plot for only the CC10 and None treatments, both having a 10 unit base. It is obvious from this plot that hypothesis one (predicting statistically identical structures for CC10 and None) is confirmed. This result replicates the findings of our earlier study (Gordon and De Leo 1975).

Comparison of spaces. Next, the treatments were compared using the Comparison of Spaces option of the Galileo program. This option uses a least squares rotation of the axes of the individual treatments to one treatment specified as the mainspace. For this plot, the treatments were organized by trace to keep the plot as neat as possible, given the complexity of comparing nine data sets. Thus, FM100 was defined as the mainspace (this decision was made, rather than using CC10 as the mainspace, because the concept numbers are placed with the first plot and the CC10 plot would have placed 13 numbers into a very small space). The trace ordering was: CC10, None, FM10, CC25, FM25, CC50, FM50, CC100, FM100. Figure 2 shows the resulting plot and confirms the intercorrelation interpretation that the relative concept locations are highly similar for each treatment but the space expands as the criterion pair distance increases. As predicted, for the same given distance, the criterion pair which is closer together produces a larger space than does the extreme pair.[2]

Figure 3 presents the comparison of spaces plot for only the CC10 and None treatments, both having a 10 unit base. It is obvious from this plot that hypothesis one (predicting statistically identical structures for CC10 and None) is confirmed. This result replicates the findings of our earlier study.

Distance Judgments Relative to Criterion Distance

The correspondence of the average (observed distances in a resulting structure) relative to the criterion pair used to

TABLE 2

Intercorrelations of Mean Distances
Among Concepts for all Treatments

	None	CC10	CC25	CC50	CC100	FM10	FM25	FM50	FM100
None	----								
CC10	.972	----							
CC25	.977	.978	----						
CC50	.975	.983	.982	----					
CC100	.972	.978	.977	.982	----				
FM10	.953	.960	.955	.963	.945	----			
FM25	.963	.979	.964	.972	.959	.959	----		
FM50	.974	.987	.976	.983	.973	.967	.979	----	
FM100	.964	.972	.970	.971	.973	.933	.946	.970	----

The n in each cell is 78.
All correlations are significant p .0001.

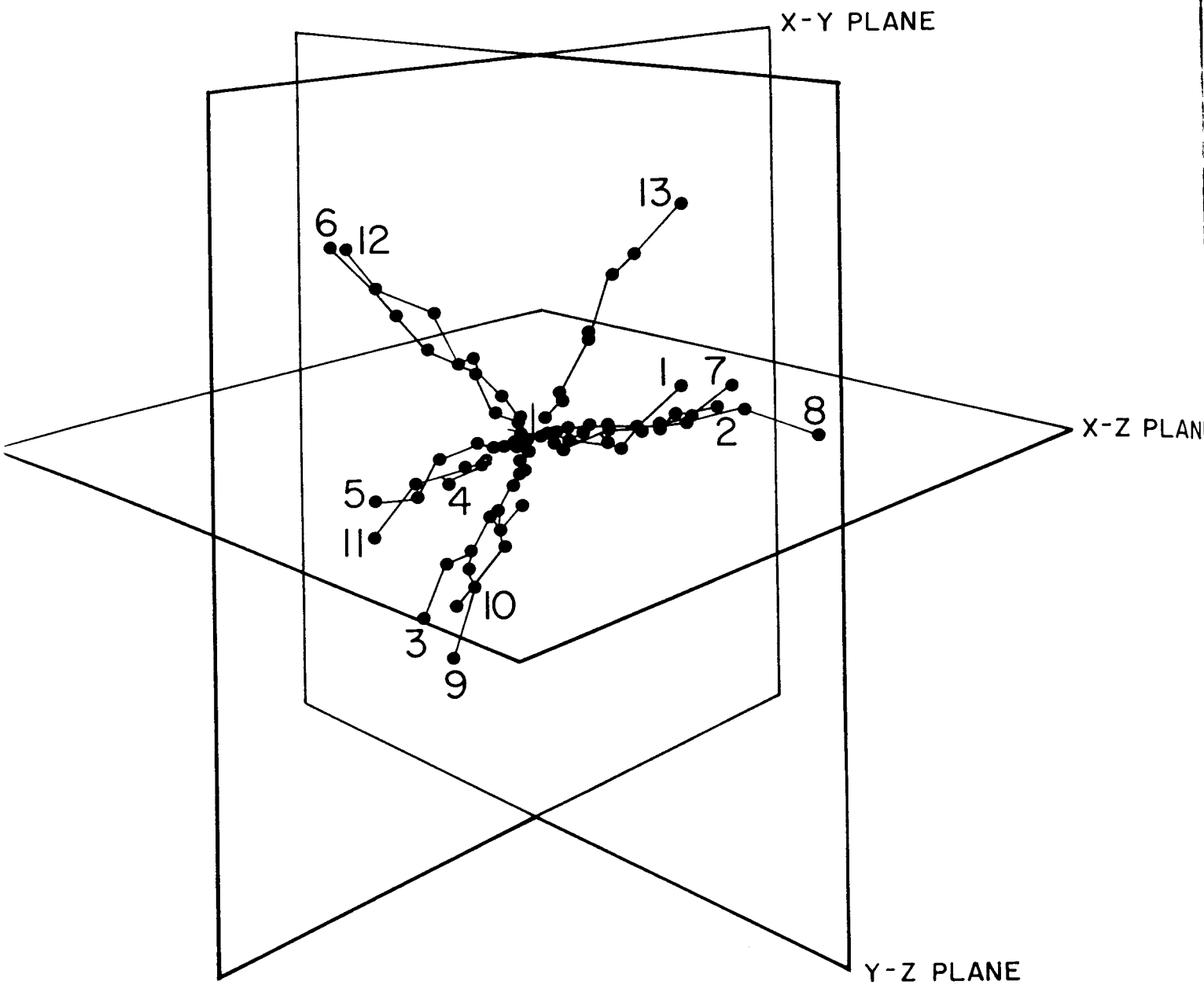
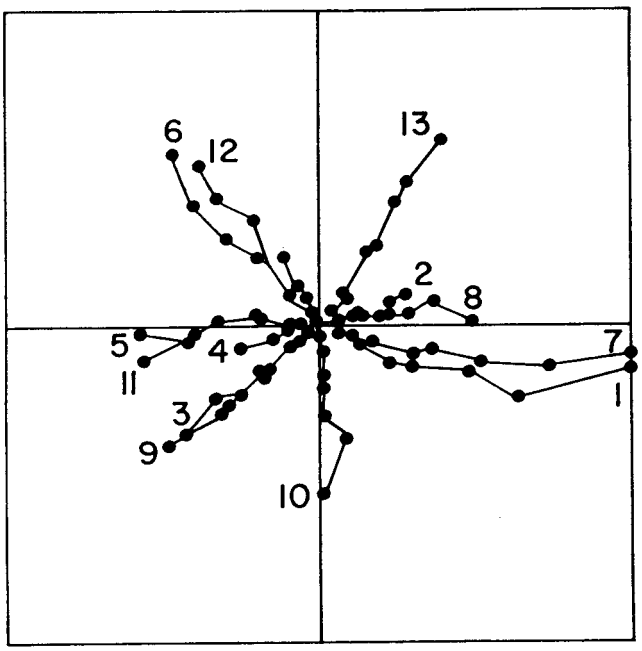
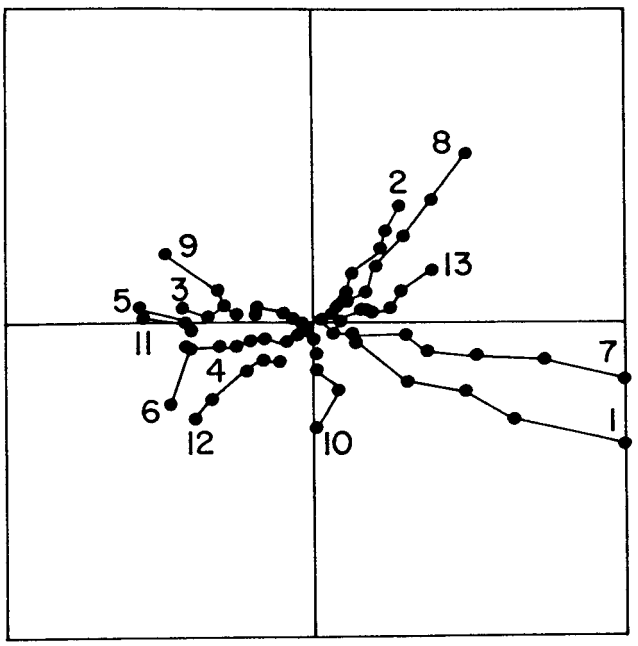


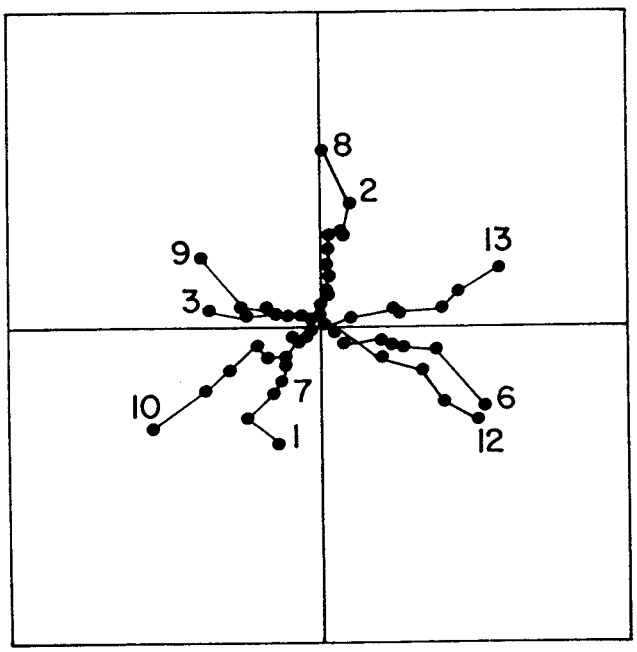
FIGURE 2. COMPARISON OF TREATMENTS. BEGINNING AT CONCEPT NUMBER, EACH POINT REPRESENTS THE JUDGEMENT OF THAT CONCEPT USING DIFFERENT CRITERION PAIR.
THE ORDER OF TREATMENTS FROM OUTER TO INNER IS: FM100, CC100, FM50, CC50, FM25, CC25, FM10, CC10 ("NONE" TREATMENT NOT INCLUDED, SEE FIG. 3).



X-Y PLANE



X-Z PLANE



Y-Z PLANE

FIGURE 2 (cont.). INDIVIDUAL PLANES FROM THREE-DIMENSIONAL PLOT.

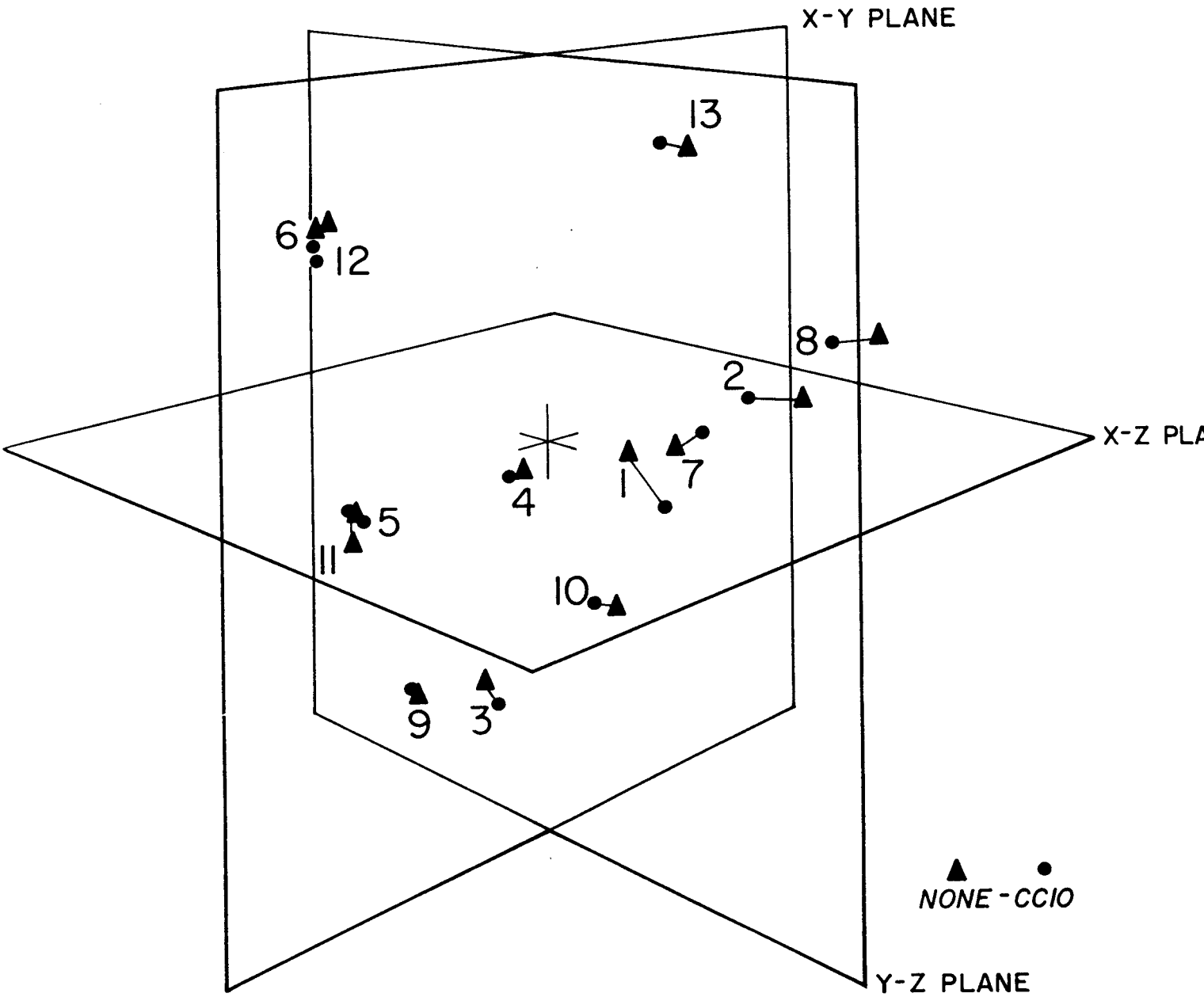


FIGURE 3. COMPARISON OF SPACES FOR TREATMENTS "NONE" AND "CCIO".

elicit those structures should provide useful information for the selection of the criterion pair. To examine this relationship the concepts involved in the pair were also included within the concept list. The actual average judgment of the concepts involved in the criterion pair, for each treatment are given below:

Treatment	Ave. Judgment CC concepts	Difference (Crit.-Observed dist.)
None	8.02	1.98
CC10	9.54	.46
CC25	23.09	1.91
CC50	48.83	1.17
CC100	97.16	2.84
FM 10	6.47	3.53
FM 25	17.15	7.85
FM 50	27.60	22.40
FM 100	50.97	49.03

It is clear from the above that the pair involving the extremes given as 10 units apart (CC10) produced the structure (on that pair) which most closely corresponds to the criterion distance given. The question of whether this single pair is representative since it was given to the subject is answered by the recognition that it was given to all treatments yet there are gross differences across treatments; and, at the same time, the intercorrelations of the structures as wholes are estremely high. Since the FM concepts were quite close together in the space, it would be expected that any movement in that direction by choosing other than the extremes would reduce the distance averages overall. Thus, it can be concluded that if the absolute distances in the structure are of interest, a criterion pair involving the extremes would be the best choice.

Variability of Judgments

The variability of the concept pair judgments across treatments was compared thru the coefficient of variation. This coefficient, which is the ratio of the standard deviation over the mean was used rather than a direct comparison of variances because of the differences in scale bases across treatments. Thus, ratios were calculated for each concept pair and these were averaged across the 78 pairs for each treatment. Table 3 presents the mean and standard deviation for each treatment, the one-way ANOVA comparing all nine treatments, and the selected comparisons (Scheffe's procedure).

Following from the significant overall ANOVA, the selected comparisons demonstrated that the only significant differences ($p < .05$) were between CC10 vs. FM10 and CC10 vs. FM100. Thus, variability of judgments relative to the mean of a given scale base revealed that CC10 produced the smallest variance ratio with FM10 and FM100 producing the largest. In general, this ratio was larger for all FM treatments as compared to the CC

TABLE 3

ANOVA COMPARISON OF TREATMENTS ON COEFFICIENT OF VARIATION, SKEWNESS AND KURTOSIS OF JUDGMENTS

Treatments*	Coef. of Var. Mean	Coef. of Var. S.D.	Skewness Mean	Skewness S.D.	Kurtosis Mean	Kurtosis S.D.
NONE	.922	.459	3.480	1.576	20.203	13.083
CC10	.640	.468	.522	.895	1.869	4.187
CC25	.817	.698	1.363	1.054	4.701	5.124
CC50	.756	.555	.945	.764	2.242	2.822
CC100	.775	.496	1.462	1.186	6.042	7.745
FM10	1.339	.418	3.632	1.013	16.196	10.608
FM25	.974	.732	1.865	1.049	5.029	8.720
FM50	.933	.661	2.103	1.039	7.630	6.907
FM100	1.018	.593	2.664	1.001	10.302	8.358

ANOVA: $p < .001$

Source	D.F.	SS	MS	F
Between	8	25.24	3.16	9.56
Within	693	228.69	.33	
Total	701	253.93		

Selected Comparisons:**

--CC10 vs. FM10 ($p < .05$)
--CC10 vs. FM100 ($p < .05$)
--All other comparisons n.s.

ANOVA: $p < .0001$

Source	D.F.	SS	MS	F
Between	8	726.85	90.86	77.18
Within	693	815.79	1.18	
Total	701	1542.64		

Selected Comparisons:
Nonsignificant groupings
--CC10, CC50
--CC25, CC50, CC100
--CC25, CC100, FM25
--CC100, FM25, FM50
--FM50, FM100
--FM10, NONE

ANOVA: $p < .0001$

Source	D.F.	SS	MS	F
Between	8	24596.88	3074.61	46.99
Within	693	45353.26	65.44	
Total	701	69950.14		

Selected Comparisons:
Nonsignificant groupings
--CC10, CC25, CC50, CC100, FM25
--CC25, CC100, FM25, FM50
--CC100, FM50, FM100
--FM10, NONE

*N for each treatment is 78 (number of pairs).
**All selected comparisons by Multiple Range Test, Scheffé procedure.

treatments, though not significantly so in most cases.

Skewness of Judgments

The skewness of the distributions of judgments within each treatment were compared by averaging the skewness values of each concept pair within each treatment. Thus, since there were 78 concept pairs in each treatment, 78 skewness values were averaged. The mean skewness values were then compared by (1) comparison to a normal distritution, (2) a one-way ANOVA across the nine treatment groups and (3) selected comparisons using the Scheffe procedure. The mean skewness values and related standard deviations appear in Table 3.

The skewness values demonstrate first that all treatments, on the average, were skewed significantly to the right as compared to a normal curve. This skewness is due primarily to the effect of the larger distance judgments in extending the right tail of the distribution. Thus, as the criterion pair distance increases, the extent of the skewness to the right also increases. Overall, CC10 was the least skewed though still significant at the .05 level. The skewness of all other treatments was at or beyond the .01 level.

The ANOVA comparing the nine treatment groups on skewness was signficant($p < .001$). The selected comparisons demonstrated that all other treatments differed signficantly from CC10 ($p < .05$) with the single exception of CC50. In terms of general consistency the criterion pairs involving the extremes (CC) were less deviant than were the closer pairs (FM). Here, CC25, CC50, and CC100 did not differ significantly from each other.

Kurtosis of Judgments

The degree of kurtosis of the distributions of judgments on the concept pairs were compared thru analyses similar to those for skewness. The kurtosis values for each concept pair were averaged for the 78 pairs in each treatment. The mean kurtosis values were then examined by (1) comparison to a normal distribution, (2) a one-way ANOVA across the nine treatment groups and (3) selected comparisons (Scheffe's). (See Table 3)

Relative to a normal distribution, with the exception of CC10 and CC50 all treatments evidenced leptokurtosis (peaked more than normal) beyond the .01 significance level. Treatment CC10 evidenced platykurtosis (flatter than normal) beyond the .01 significance level. Only CC50 produced a normal distribution.

The ANOVA comparing the nine treatments was significant ($p < .0001$). The selected comparisons (Scheffe's) demonstrated, again, that the criterion pairs involving the extremes produced the most consistency -- with no differences evident among the CC treatments on kurtosis (see Table 3).

Validity checks. Two elements were built into the design of the instrumentation to allow for validity checks on the structuring of the concepts. First, as a type of content validity check the particular television program titles were expected to cluster in the space near their general program type.

It is obvious from the plots that this is the case. Second, as a form of construct validity check, and open ended item was included asking the subject to indicate, in general, what types of TV programs they prefer. A summary of responses to this question and the significance test for differences between treatments appears in Appendix A. The significance test shows that the treatment groups are equal in terms of program preferences of the subjects in the study. Also, of the program types used, the preferred types were adult situation comedy and crime drama. As can be seen in the plots, the "me" or self-concept is located between and nearest these two types. The fact that the self-concept is somewhat distant from all programs, in general, is evidenced by the fact that a substantial number of subjects indicated preferences for types of shows not used in this study (see Appendix A). Thus, although the validity checks are minimal, the structures are highly interpretable in terms of both the content and construct checks.

Conclusions and Discussion

Hypotheses Conclusion

The results of this study confirmed all of the initial hypotheses:

> First, the no concept criterion treatment produced a structure which, relative to mean distances, was essentially identical to the structure produced by the extreme concepts criterion pair (see Figure 3).
>
> Second, as the distance between the criterion pair was increased from 10 to 25, 50, and 100 units, the resulting structural space was increased--yet the interrelationships of the concepts to each other remained the same (see Figure 2).
>
> Third, given the same distance between criterion pairs, the pair close together in the concept domain will produce a larger overall structural space than will the extreme concept pair. This results from the fact that more judgments must be outside or larger than that unit base.

Subanalyses Conclusions

Further analyses comparing the judgments resulting from the extreme criterion pairs (as opposed to the closer pairs) showed (A) more consistently high intercorrelations among all structures, (B) less variability of judgments, (C) less skewness, (D) less kurtosis, (E) more interpretable absolute distances relative to the reiterion pair distance given.

Within the extreme pair set the least deviant judgment distributions resulted from the CC10 and CC50 conditions. CC10 produced the lowest variability of all pairs (though not

significantly different from any other CC pair), the least skewness, and the smallest absolute difference between the criterion distance and the resulting observed distance on that pair. CC50 however was just as low in variability and skewness, and produced the only distribution not significantly different from normal on kurtosis. The CC10 distribution was the only treatment significantly flatter (platykurtosis) than normal.

The 25 unit distance, as an odd number, appeared to be more difficult to use. This distance, for the extreme pairs, produced the most excluded scores on the maximum value limits and the second highest variability, skewness, and kurtosis values. The 100 unit base produced the most skewness and kurtosis of the extreme pairs. The frequency distribution of judgements for CC100 revealed that almost all judgments were made in 10 unit jumps with some 5 unit choices between. This continued up to the 130 unit distance and from there the judgments were in 50's and 100's. Thus, with 100 units the full scale is rarely used and the larger jumps at the extremes contribute to the distortion of the distribution.

Discussion

The results of this study, relative to the ability of metric MDS to reproduce a consistent structure as operationalized thru the GALILEO program and the subject's ability to adapt to a given criterion are very impressive. The odds that nine independent groups of people, each using different criteria to judge 78 different pairs of concepts, would produce statistically identical structures must be very low. Thus, these results shed light on both the subject's ability to adapt to differing measurement critieria and the method's ability to precisely reflect those judgments. Error on either side would have decreased the likelihood that similar structures would be evident across treatments.

It is clear that for the choice of a criterion pair for use with a homogeneous set of concepts, the extremes from the concept domain are preferable over a pair close together in the set. Given the evidence for interpretation of the absolute distances produced from the judgments, this also appears preferable over the suggestion to chose a standard, "...approximately midway between the largest and smallest measurement likely to be encountered,..."(Woelfel, 1974, p. 16).

The best choice for a unit distance between concepts in the criterion pair, of those tested, appears to be either 10 or 50. On the basis of the distributions, 50 would be preferred. However, we should note that in the methods section it was pointed out that the random assignment of classes to treatments resulted in an over representation of upper class students in the CC50 treatment. Thus, the greater consistency of CC50 in terms of kurtosis may reflect the greater degree of homogeniety of that group and their more advanced cognitive skills.

These findings should be placed in the context of both the nature of the concepts and the characteristics of the sample. The concepts were purposely selected to be homogeneous. A

heterogeneous set of concepts might show less stability. As well, the sample of college students may be better able to adapt judgments to differing criteria than is the general public. These questions are yet to be explored. Overall, putting aside the differences in the distributions resulting from the different pairs, the fact that the structures as relative wholes were so consistent under the variations imposed makes it clear that this procedure is exceptionally robust. Thus, given a sufficient sample size and care with extreme values, it is reasonable to assume that accurate representations of the general public would be obtained.

NOTES

1. An earlier draft of this chapter was presented to the Association for Education in Journalism, College Park, Maryland, 1976. The competent assistance of Michael Fisher and Gary Solarz is gratefully acknowledged. Their efforts in data collection and processing were far above and beyond the call of duty. Thanks go also to Gary Jeffries and Ken Galen for their help in data collection.

REFERENCES

Barnett, G.A. Reliability and metric multidimensional scaling. Unpublished research report, East Lansing, Michigan State University, Communication Department, 1972.

Broadcasting. "ABC-TV Takes Over Second In Ratings." Vol. 89, No. 13, September 29, 1975.

Danes, J.E., and Woelfel, J. An alternative to the "traditional" scaling paradigm in mass communication research: multidimensional reduction of ratio judgments of separation. Paper presented at the annual meeting of the International Communication Association, Chicago, 1975 (chapter 9).

Gillham, J.R. and Woelfel, J. The Galileo system of measurement: preliminary evidence fore precision, stability and equivalence to traditional measures. Unpublished research report, Akron, Ohio, University of Akron, 1975.

Gordon, T.F., and De Leo, H.C. Structural variation in "Galileo" space: effects of varying the criterion pair in metric multidimensional scaling. Paper presented at the annual meeting of the International Communication Association, Portland, Oregon, 1976.

Gregson, R.A.M. Psychometrics of Similarity. New York: Academic Press, 1975.

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

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

Richardson, M.W. Multidimensional psychophysics. Psychological Bulletin, 1938, 35, 659-660.

Schramm, W. Men, Messages, and Media. New York: Harper & Row, 1973.

Shepard, R.N. The analysis of proximities: multidimensional scaling with an unknown distance function. I. Psychometrika, 1962, 27, 219-246. (a)

Shepard, R.N. Introduction to volume I. In R. Shepard, A. Romney, and S. Nerlove (Eds.), Multidimensional Scaling: Theory and Applications in the Behavioral Sciences, Vol. 1. New York: Seminar Press, 1972. (a)

Shepard, R.N. A taxonomy of some principal types of data and of multidimensional methods for their analysis. In R. Shepard, A. Romney, and S. Nerlove (Eds.), Multidimensional Scaling: Theory and Applications in the Behavioral Sciences, Vol. 1. New York: Seminar Press, 1972. (b)

Sherman, C. Nonmetric multidimensional scaling: a Monte Carlo study of the basic parameters. Psychometrika, 1972, 37, 461-486.

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

Torgerson, W.S. A theoretical and empirical investigation of multidimensional scaling. Ph.D. Thesis, Princeton University, 1951.

Torgerson, W.S. Multidimensional scaling: I. Theory and method. Psychometrika, 1952, 17, 401-419.

Torgerson, W.S. Theory and Methods of Scaling. New York: Wiley, 1958.

Woelfel, J. & Saltiel, J. Cognitive processes as motion in a multidimensional space: a general linear model. Unpublished manuscript, Michigan State University, East Lansing, 1975.

Woelfel, J. Procedures for the precise measurement of cultural processes. Unpublished manuscript, East Lansing, Michigan State University, Communication Department, 1974.

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

Young, G. & Householder, A.S. Discussion of a set of points in terms of their mutual distances. Psychometrika, 1938, 3, 19-22.

Sample Demographics by Treatment

Treatment	YEAR Fr.	Soph	Jr	Sr.	AGE 18-25	26-35	36+	SEX Male	Fem.	RACE Bl.	Wh.	Oth.r	FAMILY INCOME 0-4999	5-7999	8-9999	10-14999	15-20	20
None	21	24	28	21	79	10	4	66	27	11	77	5	6	7	8	26	21	23
CC10	14	13	36	28	77	14	1	74	18	7	79	6	11	9	12	23	17	19
CC25	28	15	33	16	80	10	2	59	33	6	85	2	4	5	9	27	22	25
CC50	4	8	46	29	77	13	1	62	29	22	64	5	6	7	9	27	14	28
CC100	16	33	23	20	83	8	2	63	29	13	71	9	8	8	7	31	18	21
FM10	12	43	36	16	102	7	1	73	35	16	88	3	6	10	13	22	29	27
FM25	24	24	28	19	87	10	0	63	28	13	78	5	4	5	13	25	25	22
FM50	33	12	26	18	79	9	0	46	43	11	72	2	7	7	9	22	23	22
FM100	14	33	26	24	89	8	1	66	32	14	78	9	7	7	10	23	18	32

Treatment	AVE HOURS TV/DAY 0-.9	1-1.9	2-2.9	3-3.9	4+	AVE HOURS TV/WEEK 0-3.9	4-8.9	9-13.9	14-18.9	19-23.9	24+	PREFERRED PROGRAM TYPE #2	#6	1/3/4/5	others
None	11	20	32	20	11	8	17	23	20	10	15	46	15	6	26
CC10	18	22	23	19	10	13	21	19	16	11	11	47	14	5	28
CC25	17	25	27	6	15	11	26	18	14	8	15	46	12	7	27
CC50	9	28	30	12	12	4	31	17	17	7	15	36	15	5	35
CC100	16	26	19	16	16	11	23	15	15	9	20	48	15	1	28
FM10	20	20	32	19	19	17	17	21	23	14	13	59	18	6	29
FM25	9	29	30	17	11	10	26	17	17	18	13	47	14	4	39
FM50	11	12	31	21	13	9	15	18	20	18	13	44	15	10	25
FM100	19	33	26	11	8	13	31	18	19	8	8	46	15	2	34

All cross-tabs nonsignificant by Chi Square except Year (X^2=101.18, df=24) and Race (X^2=27.75, df=16).