PSYCHOMETRIKA—VOL. 50, NO. 3, 275–300.
SEPTEMBER 1985

THREE-WAY METRIC UNFOLDING VIA ALTERNATING WEIGHTED LEAST SQUARES

WAYNE S. DESARBO

MARKETING DEPARTMENT
WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA

J. DOUGLAS CARROLL

AT&T BELL LABORATORIES

Three-way unfolding was developed by DeSarbo (1978) and reported in DeSarbo and Carroll (1980, 1981) as a new model to accommodate the analysis of two-mode three-way data (e.g., nonsymmetric proximities for stimulus objects collected over time) and three-mode, three-way data (e.g., subjects rendering preference judgments for various stimuli in different usage occasions or situations). This paper presents a revised objective function and new algorithm which attempt to prevent the common type of degenerate solutions encountered in typical unfolding analysis. We begin with an introduction of the problem and a review of three-way unfolding. The three-way unfolding model, weighted objective function, and new algorithm are presented. Monte Carlo work via a fractional factorial experimental design is described investigating the effect of several data and model factors on overall algorithm performance. Finally, three applications of the methodology are reported illustrating the flexibility and robustness of the procedure.

Key words: multidimensional scaling, unfolding, preference analysis.

1. Introduction

DeSarbo (1978) has developed three-way unfolding initially as an unfolding model (Coombs, 1964) for three-mode three-way preference data (DeSarbo & Carroll, 1980, 1981). The primary objective in the development of this new method was to portray or scale the interrelationships and underlying dimensions for three-way arrays, assuming a metric ideal point distance model. Basically, a generalization of INDSCAL (Carroll & Chang, 1970) to both nonsymmetric proximity and dominance data was created. (Note, ALSCAL, developed by Takane, de Leeuw, and Young, 1977, can also handle the unfolding of such three-way preference data.) Consider the example where subjects render preferences for various stimuli (e.g., products or brands) over different situations (time, physical surroundings, etc.). Then individuals, represented as ideal points, and stimuli, represented as points, are portrayed in a T-dimensional joint space configuration, and the situation weights are estimated and plotted in the positive orthant of a separate "weights" space. These weights reflect the salience or importance of each derived dimension for each situation. This three-way unfolding model assumes a common stimulus space with differential weighting of uniquely oriented axes (up to permutation or reflection) for each situation. The situation weights or saliences on each dimension are idiosyncratic and can be used to estimate the respective perceptual spaces relevant to each particular situation

We wish to thank the editor and reviewers for their insightful comments.

Requests for reprints should be sent to Wayne S. DeSarbo, Marketing Department, Wharton School, University of Pennsylvania, Philadelphia, PA 19104.

0033-3123/85/0900-6050$00.75/0

(similar to the method employed in INDSCAL for handling subjects). Thus, in the example discussed above, the methodology solves for both the common joint space for stimuli and ideal points, and the dimension weights for each situation. In addition, a "floating" ideal point model was developed as an option for those applications where it is assumed that ideal points (preferences) change or vary over situations. This situational analysis application will be used as an example to assist in the explanation of the model. Later, other applications will be mentioned.

2. Three-Way Unfolding

The Model

The underlying model assumption is that dispreference values, i.e. values inversely related (e.g., via some non-increasing monotonic function) to preference values, are measured on interval (or ratio) scales and are linearly related to squared weighted Euclidean distances:

$$\delta_{ijk} \cong d^2_{ijk} + \alpha_i = \sum_{t=1}^{T} w_{it}(y_{jt} - x_{kt})^2 + \alpha_i ; \tag{1}$$

where:

$i = 1, \ldots, I$ (situations),

$j = 1, \ldots, J$ (judges),

$k = 1, \ldots, K$ (stimuli), and

$t = 1, \ldots, T$ (dimensions);

$\delta_{ijk} \equiv$ the obtained dispreference value of judge j for stimulus k in situation i;

$\cong$ indicates "approximately equals" in a least-squares sense;

d^2_{ijk} = squared Euclidean distance between judge j and stimulus k in situation i;

w_{it} = situation salience (weights for dimension t in situation i);

$y_{jt} \equiv$ the t-th coordinate of judge j's ideal point;

$x_{kt} \equiv$ the t-th coordinate of stimulus point k; and

$\alpha_i \equiv$ an additive constant for situation i.

This model can be viewed as a direct three-way generalization of Schönemann's (1970) two-way metric unfolding model. (The approach to data analysis we outline in Appendix A similarly generalizes the analytic procedure Schönemann proposed to the three-way case.) The assumption of dispreference values being linearly related to squared Euclidean distances indicates that preference diminishes quickly as one deviates (in any direction) from the subject's ideal point.

Also note that the model in (1) can be used to summarize two-mode (or three-mode), three-way nonsymmetric proximities data (δ_{ijk}) where row and column objects are portrayed in the joint space. Here, distances between row object y_j and column object x_k for slice i in the T dimensional joint space are fit to δ_{ijk} via weighted least-squares.

The Objective Function

Originally, DeSarbo (1978) proposed the following objective or loss function which was to be minimized in estimating the parameters w_{it}, y_{jt}, x_{kt}, and α_i given $\Delta = \| \delta_{ijk} \|$

and T:

$$\text{Min } Z_1 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (\delta_{ijk} - \hat{\delta}_{ijk})^2, \tag{2}$$

where:

$$\hat{\delta}_{ijk} = \sum_{t=1}^{T} w_{it}(y_{jt} - x_{kt})^2 + \alpha_i. \tag{3}$$

The original algorithm, presented in detail in Appendix A, consisted of a complex procedure entailing use of the Carroll-Chang CANDECOMP procedure followed by an alternating least-squares procedure (Wold, 1966) which cyclically estimates w_{it}, y_{jt}, x_{kt}, and α_i (an option also existed for estimating "floating ideal points", y_{ijt}), holding the other three sets of parameters fixed at their current values. In addition to being computationally expensive, often times the procedure would render degenerate solutions, typical of most unfolding schemes. For example, one set of points (e.g., the y_{jt}'s) would be quite separate from the other set of points (e.g., the x_{kt}'s) in the resulting joint space representations. At other times the degeneracy took the form (in two dimensions) of points in two concentric circles where, for example, the ideal points might be clustered in a tight small circle and the stimuli would be scattered in a larger circular region around the ideal points.

There has been some research performed attempting to cure unfolding of its predisposition toward degenerate solutions. Heiser (1981) suggests imposing configuration restrictions in the two-way case on the $\mathbf{X} = \| x_{kt} \|$ and $\mathbf{Y} = \| y_{jt} \|$ joint space. The restriction Heiser recommends is one where each stimulus is constrained to be at the centroid of the location of the subjects (ideal points) for whom it was most preferred (first chosen), or, more generally, among the first $R \leq K$ most preferred stimuli. One limitation with the Heiser approach concerns the basic assumption regarding the reliability of the data. Heiser's (1981) approach tends to place most of the emphasis in estimating parameters on the first R choices (most preferred stimuli), while tending to deemphasize the remainder of the data. This roughly assumes that a subject can only reliably render preference information about the most preferred brand, or the first R preferred, while the other judgments may be too "noisy" or unreliable to carry any great amount of weight in estimating model parameters. While this weighting scheme may be viable for some particular applications, it is not realistic for others. Depending upon such factors as the nature of the study, questionnaire design, number of stimuli, discriminability between stimuli, type of subjects, knowledge of the stimuli by subjects, etc., different assumptions may be appropriate concerning how much error the preference judgments contain. For example, in consumer research for new product concepts, a consumer may be able to render reliable judgments on the first R preferred items. Or, it may be the case that the typical subject could provide reliable ratings for a favorite R_1 items and for a least favored R_2 items. Thus, a subject could reliably tell the interviewer what is liked and what is disliked, but may have trouble in rating products that he/she is indifferent towards. Such applications may require a quite different weighting scheme, if the aim is to weight more heavily the more reliable responses.

We propose a different approach to the general degeneracy problem in unfolding, and to three-way unfolding as a specific case (a similar approach to two-way unfolding is presented in DeSarbo and Rao, 1984). Our approach involves explicitly altering the loss function in (2) to incorporate data weights γ_{ijk}:

$$\text{Min } Z_2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \gamma_{ijk}(\delta_{ijk} - \hat{\delta}_{ijk})^2, \tag{4}$$

where the γ_{ijk} are defined by the user to weight the δ_{ijk} values differentially. We believe, like Heiser (1981), that a possible cause for degeneracy is the error in the data, and provide for the flexibility of the user specifying γ_{ijk} differently. For example, one may define these weights as:

$$\gamma_{ijk} = \left(\frac{1}{\delta_{ijk}}\right)^p, \tag{5}$$

where p is some exponent. Assuming the δ_{ijk} are dispreference values, such a weighting scheme would weight smaller values (more preferred) higher than larger values (less preferred).

Where some form of preprocessing is involved (e.g., subtracting means) and negative δ_{ijk}'s appear as interval scale data, such a weighting function may not be meaningful. Here, for example, one could use:

$$\gamma_{ijk} = \left(\frac{1}{r(\delta_{ijk})}\right)^p, \tag{6}$$

where p is an exponent and $r(\delta_{ijk})$ represents the row ranks (from smallest = 1 to largest = K) of the δ_{ijk}. Options also exist to define $r(\delta_{ijk})$ over the slice instead of the row for matrix conditional (Takane et al., 1977) analyses. At any rate, as $p \to \infty$, this (row) weighting scheme in (6) resembles (but is not exactly equivalent to) Heiser's (1981) configuration restrictions ($R = 1$) since only first choices would be significantly weighted (as $\gamma_{ijk} = 1$ for first choices), while the rest of the γ_{ijk} would tend to zero.

Other weighting options are also available in three-way unfolding. For example, one could specify $\gamma_{ijk} = 1$, $\forall i, j, k$, so that the "weighted" loss function reduces to the (equally weighted) nonweighted one. Or, the user can specify a bimodal or step weighting function where, for example, the first four and last two choices would be highly weighted, while all others would receive low weights.

The psychometric literature provides some discussion concerning the reliability of proximity judgments. Isaac (1980) states that large proximity judgments are the most reliable, or are at least as reliable as the other (smaller) ones. Graef and Spence (1979) show that the larger distances are the most important in achieving quality recovery of configurations. However, unfolding analysis and analyses of nonsymmetric proximities typically estimate two sets of points. And the relationships between these two sets of points are often the most important aspects of these analyses. Degeneracies that occur where the two sets of points are separated quite far from one another (larger distances) rarely offer any insight into the structure of the data. Thus, altering the loss function to weight smaller distances more heavily (as one option) tends to counteract typical tendencies toward degenerate solutions and wide separations between the two sets of points.

Concerning the reliability of such judgments, our methodology offers the user the option of specifying a number of different types of weighting functions reflecting his belief concerning the reliability of the data. The choice of the "appropriate" weighting function depends upon such factors as preprocessing options and scale assumptions of the data, assumptions of the conditionality of the data, assumptions concerning the reliability of the different data values, and trial and error. If the data (Δ) were treated as interval scale and preprocessed accordingly (e.g., by taking out particular row, column, and/or "slice" means), then specifying γ_{ijk} as in (5) may not make sense. Depending upon the assumptions made concerning the conditionality of the data, it may or may not make sense to define γ_{ijk} via (6) over rows (slices). For example, if δ_{ijk} were three-way nonsymmetric proximities, it might make more sense to define γ_{ijk} in (6) over each slice of the array rather than over rows within each slice. Different γ_{ijk} could be specified depending upon the assumptions made concerning the reliability of the δ_{ijk} collected. If one believes that

only highly preferred judgments are reliably given, one could specify a unimodal γ_{ijk} function weighting smaller values highly. Alternatively, if one believes that the δ_{ijk} can be reliably given for both highly preferred and highly non-preferred judgments, then a bimodal weighting function could be specified. (Note, there may be instances where the preference judgments elicited for the least preferred stimuli are more reliable than those for the most preferred stimuli, e.g., when all stimuli are onerous or negatively toned. In such cases, the idea of weighting the more preferred stimulus judgments more heavily via (5) or (6) may prove counterproductive. Here, the user may wish to utilize equal weights or even weight the least preferred judgments higher.)

These criteria may provide insight into the *general* form of γ_{ijk} (e.g., unimodal vs. bimodal, or (5) vs. (6)), but specific decisions, such as what value p should have, can be made by trial and error, although as will be demonstrated in the applications that follow, in (6), $p = 2$ or 3 appears to work well. We believe that the specification of p is an empirical issue depending upon the data. One approach is to run the analyses with a different sequence of p's ($p = 0, 1, 2, 3, 4$) and examine at what point degenerate solutions disappear. Another approach would be to attempt to estimate p itself via another stage in the algorithm. While this is feasible, constraints would have to be placed on p to avoid possible "degenerate" values (large). In addition, this would add to the computational complexity of the algorithm.

The Algorithm

Appendix B presents the numerical details of the new cyclical weighted least-squares algorithm utilized to estimate w_{it}, y_{jk}, (or y_{ijk} for the floating ideal point case), x_{kt}, and α_i, given δ_{ijk}, γ_{ijk}, and T. Basically, estimates of one set of parameters occur at a particular stage holding the other three sets of parameters fixed at their current levels. One has the option of performing an external ($\mathbf{X} = \| x_{kt} \|$ given) or internal (estimating $\mathbf{X}$) analysis; of accommodating interval or ratio scale Δ_{ijk}; and, of estimating fixed (y_{jt}) or "floating" (y_{ijt}) ideal points. The new algorithm is more computationally efficient and, as mentioned, provides a reduction of risk of degenerate solutions. Note that unlike INDSCAL (Carroll & Chang, 1970), $\sum_{t=1}^{T} w_{it}^2$ is not in *general* an indicator of goodness of fit by slice in three-way unfolding. One can show that $\sum_{t=1}^{T} w_{it}^2$ is a rough approximation to a sums of squares accounted for measure (by slice) if $\gamma_{ijk} = 1, \forall i, j, k$, and if one imposes the normalization $\sum_{j=1}^{J} \sum_{k=1}^{K} y_{jt} x_{kt} = 1$ for fixed t.

3. Monte Carlo Results

In order systematically to examine the performance of the new three-way unfolding algorithm, a Monte Carlo analysis was performed where some ten factors, displayed in Table 1, were experimentally varied: T, I, J, K, type of weighting functions γ_{ijk}, error in Δ_{ijk}, type of starting configuration, type of ideal points (fixed vs. floating), scale of Δ_{ijk}, and type of analysis (external vs. internal). These ten factors and their respective levels shown in Table 1 were initially hypothesized to have a potential effect on the performance of the new three-way unfolding algorithm. These factors were combined via an asymmetric fractional factorial ($3^8 2^3$) design (Addelman, 1962) for main-effects only estimation. Twenty-seven experimental trials were devised where the ten factors were varied according to the fractional factorial design shown in Table 2.

Each trial of the experimental design defined a particular level for each of the factors investigated in the study. Based on these stipulated levels, $\mathbf{W}$, $\mathbf{X}$, $\mathbf{Y}$ and $\boldsymbol{\alpha}$ were generated randomly from a uniform distribution, and the data array $\boldsymbol{\Delta}$ was created according to (1). Error was added to $\boldsymbol{\Delta}$, where the design stipulated such, as a function of the variance in the $\mathbf{Y}$ configuration (see bottom of Table 1). Types of solution, starting values (see Appendix B), and weighting function were also established by the particular trial in the design.

TABLE 1

Independent Factors and Respective Levels Defined for Three Way Unfolding Monte Carlo Analysis

FACTOR	LEVELS (CODES)		
1. #Dimensions (T)	1(0)	2(1)	3(2)
2. #Slices (I)	2(0)	3(1)	4(2)
3. #Subjects (J)	10(0)	15(1)	20(2)
4. #Stimuli (K)	10(0)	15(1)	20(2)
5. γ_{ijk} Weighting Scheme	None(0)	$\frac{1}{r(\delta_{ijk})}$(1)	$\left(\frac{1}{r(\delta_{ijk})}\right)^2$(2)
6. Error to Data*	None(0)	15%(1)	30%(2)
7. Start	Random(0)	Close(1)	CANDECOMP(2)
8. Ideal Points	Fixed(0)	Floating(1)	
9. Scale	Interval(0)	Ratio(1)	
10. Analysis	Internal(0)	External(1)	

* Error percentages reflect the percent of the variance in the Y configuration used to define the variance parameter σ^2 of the uniformly distributed error introduced in the analyses.

TABLE 2

$3^7 2^3$ Fraction Factorial Design for Three-Way Unfolding Monte Carlo Analysis

	TRIAL																									
Factor	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
2	0	0	0	1	1	1	2	2	2	0	0	0	1	1	1	2	2	2	0	0	0	1	1	1	2	2	2
3	0	0	0	1	1	1	2	2	2	1	1	1	2	2	2	0	0	0	2	2	2	0	0	0	1	1	1
4	0	0	0	2	2	2	1	1	1	1	1	1	0	0	0	2	2	2	2	2	2	1	1	1	0	0	0
5	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
6	0	2	1	0	2	1	0	2	1	1	0	2	1	0	2	1	0	2	2	1	0	2	1	0	2	1	0
7	0	1	2	1	2	0	2	0	1	0	1	2	1	2	0	2	0	1	0	1	2	1	2	0	2	0	1
8	0	0	1	1	0	0	0	1	0	1	0	0	0	1	0	0	0	1	0	1	0	0	0	1	1	0	0
9	0	1	0	0	0	1	1	0	0	1	0	0	0	1	0	0	0	1	0	0	1	1	0	0	0	1	0
10	0	0	1	0	1	0	1	0	0	1	0	0	0	0	1	0	1	0	0	1	0	1	0	0	0	0	1

The dependent (performance) measures collected were: the number of major iterations required for convergence; the average (over dimensions) sum of squares accounted for between $\mathbf{W}$ and $\hat{\mathbf{W}}$; the average (over dimensions) variance accounted for between $\mathbf{X}$ and $\hat{\mathbf{X}}$; the average (over dimensions) variance accounted for between $\mathbf{Y}$ and $\hat{\mathbf{Y}}$; and, the average (over dimensions) variance accounted for between the joint space

$$\mathbf{L} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{L}} = \begin{bmatrix} \hat{\mathbf{X}} \\ \hat{\mathbf{Y}} \end{bmatrix}.$$

Note, because of the impact of the γ_{ijk} on the overall goodness-of-fit measure, it was decided not to include the goodness-of-fit in the dependent measure group. (C.P.U. time would have also been a good dependent measure to examine, but was not available for each trial).

Table 3 presents the dependent measure results for the twenty-seven trials designated by the fractional factorial design. Note that recovery of $\mathbf{W}$ appears to be more successful than of $\mathbf{X}$, $\mathbf{Y}$, or $\mathbf{L}$ over the twenty-seven trials. This is probably a result of the fact that the sums of squares accounted for measure may be less sensitive than the variance accounted for measures and/or the fewer number of parameters in $\mathbf{W}$. Table 4 presents the results of the five regression analyses performed, one for each of the dependent measures. Table 4 presents the various main effects regression coefficients for the nonzero coded levels of the experimental factors (the design in Table 2 was converted to dummy variables where factors with r levels were coded into $r - 1$ dummy variables), standard error, R^2, adjusted R^2, and F-statistics. Because dependent measures 2–5 are proportions of variance accounted for, ranging between 0 and 1 (analogous in some ways to probabilities), additional multiple regression analyses were performed on two transformations of these six measures. One transformation was the logistic transformations $\{\log [p_{ij}/(1 - p_{ij})]\}$ and the other was the arc sine transformation of $[(p_{ij})^{1/2}]$ (Snedecor & Cochran, 1981). The results of these regressions were nearly identical to those presented in Table 4 and will thus not be presented.

All five of these multiple regressions were *not* significant according to the associated F-tests. The "close" start (defined in Appendix B) significantly detracted from configuration recovery for dependent measures 3, 4, and 5. In fact, in returning to the data in Table 3, one notes the asterisks for trials 4, 9, 10, and 18 which denote poor local minimum solutions (Trial 11 also uses a close start and fails to recover the joint space adequately). Trials 4, 9 and 18 were designated using the "close" starting option, while Trial 10 used a random start. These four analyses were redone using the CANDECOMP start and Table 5 presents the improved results for these four trials.

Returning to Table 4 and the multiple regression analyses, the external analysis variable was significantly related to the recovery of $\mathbf{X}$. This trivially makes sense since one would expect to recover $\mathbf{X}$, given that $\mathbf{X}$ was fixed in the external analysis.

Thus, the Monte Carlo analysis displays positive evidence concerning the robustness of the three-way unfolding algorithm to variations in T, I, J, K, amount of error in Δ, type of ideal point estimation, scale type, and type of analysis. However, evidence does appear to suggest that the procedure is somewhat sensitive to the type of start one employs, suggesting that the CANDECOMP start may provide the most reduction in risk against poor local optimum solutions.

While the Monte Carlo analysis has provided evidence as to the robustness of the methodology, several limitations must be noted. The use of the fractional factorial design does not allow the flexibility of measuring possible interaction effects between the factors studied in the analysis. Clearly, assuming computational expense was not a limitation, a full factorial design would have been a more comprehensive design to use in order to estimate higher order interaction terms. In addition, the design should have been repli-

TABLE 3

Results for Dependent Measures

Trial	No. of Iterations	W	X	Y	X,Y
1	9	1.000	1.000	1.000	1.000
2	15	1.000	.789	1.000	.891
3	12	1.000	1.000	.758	.879
4*	27	1.000	.011	.056	.034
5	20	1.000	1.000	.915	.956
6	14	1.000	1.000	.833	.909
7	17	1.000	1.000	1.000	.999
8	8	1.000	1.000	.941	.971
9*	2	.858	.059	.141	.088
10*	2	.998	1.000	.007	.501
11	48	.984	.913	.225	.098
12	25	.998	.800	.777	.808
13	18	.998	.879	.892	.921
14	8	1.000	1.000	.801	.900
15	27	1.000	1.000	.998	.997
16	41	1.000	1.000	.795	.965
17	28	1.000	1.000	.737	.879
18*	50	.903	.319	.012	.005
19	24	.951	.806	.665	.828
20	31	.989	1.000	.666	.833
21	34	.991	.819	.760	.781
22	3	1.000	1.000	.744	.881
23	26	.931	.916	.998	.956
24	48	.999	.998	.881	.940
25	23	.994	.891	.521	.706
26	50	.956	.667	.634	.441
27	10	1.000	1.000	.889	.923

* represents local minimum solutions due to poor initial starting configurations

cated in order to improve the degrees of freedom for estimation. Finally, more levels for each of the factors should be investigated in future work, especially concerning the different types of weighting functions tested in order to obtain better guidelines as to when to use a particular type of weighting function.

4. Applications

Three different applications of three-way unfolding are presented to illustrate the various types of three-way data that can be analyzed. In the first application, we examine three-way, three-mode preference data. In the second, an analysis of three-way, three-mode profile data is presented with use of the floating ideal point option. Finally, an

TABLE 4

Multiple Regression Results for Three-Way Unfolding Monte Carlo Analysis

	Dependent Measures				
	1	2	3	4	5
$T=2$	13.67	0.003	0.093	-0.132	-0.073
$T=3$	13.89	-0.005	0.114	0.036	0.062
$I=3$	-1.00	0.002	-0.059	0.163	0.097
$I=4$	3.22	-0.022	-0.156	0.003	0.071
$J=15$	-1.44	0.011	-0.106	-0.206	-0.224
$J=20$	-7.00	-0.005	-0.074	0.017	-0.009
$K=15$	0.78	-0.020	-0.083	-0.174	-0.157
$K=20$	10.78	-0.013	-0.165	-0.205	-0.163
$\gamma_{ijk} = \frac{1}{r(\Delta_{ijk})}$	7.78	-0.009	0.101	0.114	0.010
$\gamma_{ijk} = \left(\frac{1}{r(\Delta_{ijk})}\right)^2$	6.44	-0.021	-0.066	0.041	0.056
15% error to Δ	-3.67	-0.027	-0.024	-0.069	-0.007
30% error to Δ	-3.78	-0.014	0.008	0.001	0.054
Close Start	-0.67	-0.019	-0.254*	-0.254*	-0.310**
CANDECOMP Start	-0.44	0.001	-0.005	0.070	0.054
Floating Y_{ijk}	0.39	0.006	-0.079	-0.210	-0.155
Ratio Scale	-2.28	-0.000	0.019	-0.094	-0.065
External Analysis	-9.44	0.023	0.218*	0.095	0.192
Intercept	13.89	1.021	1.027	0.964	1.018
S.E.	17.11	0.04	0.30	0.27	0.32
R^2	0.54	0.55	0.59	0.74	0.66
adj R^2	0.00	0.00	0.00	0.24	0.03
F	0.62	0.66	0.76	1.48	1.05

*$p < .10$
**$p < .05$
***$p < .01$

TABLE 5

Redone Trials 4, 9, 10 and 18 with CANDECOMP Start

	Dependent Measure				
New Trial	1	2	3	4	5
4	10	1.000	1.000	1.000	.978
9	5	1.000	1.000	.804	.875
10	15	1.000	1.000	.872	.941
18	30	.986	.476	.445	.460

analysis of three-way, two-mode nonsymmetric proximities is illustrated with the Miller-Nicely (1955) data. In all cases, only the weighting function in (6) was applied with various values of p.

DeSarbo and Rao Pain Reliever Preference Data

DeSarbo and Rao (1983) report the collection of intention to use judgments (measured on a ten point scale) for ten brands of over-the-counter (OTC) pain relievers by some 61 MBA students at Fairleigh Dickenson University in Madison, New Jersey for three different maladies. The ten brands tested were: Anacin, Ascriptin, Bayer Aspirin, Bufferin, Cope, Datril, Excedrin, Hudson Aspirin, Tylenol, and Vanquish. The three common maladies were: headache, fever, and muscle aches. Students were given color photographs of each of the ten brands with corresponding information on price, ingredients, and package remedy claims. Thus, the three-way, three-mode array of subjects (rows) rendering intention to use judgments for ten brands of OTC pain relievers (columns) for three different maladies (slices) comprised the input data.

The data were treated as interval scale and the grand mean was removed. The analysis was conducted in 1–4 dimensions using (6) as the weighting function for various values of p (0, 1, 2, 3, 4). As one reviewer aptly pointed out, one could also run this application with a bimodal weighting function arguing that subjects could reliably render such judgments for brands that were most and least preferred. Degenerate solutions stopped arising with $p = 2$. $T = 2$ dimensions appeared to best describe the structure in the data (explaining 51% of the variance) as judged by examination of the goodness-of-fit measure for each of the dimensionalities estimated. Figure 1 presents the two-dimensional joint space of brands and subjects. The numbers in the plot represent the ideal points for subjects 1–61. The first dimension appears to separate heavily advertised, high market share brands (Bayer, Datril, Tylenol, Bufferin, Anacin, and Excedrin) from the less known, low market share brands (Ascriptin, Cope, Vanquish, and Hudson). The second dimension separates aspirin (Ascriptin, Cope, Vanquish, Hudson, Excedrin, Bufferin, Anacin, and Bayer) from aspirin substitutes (Tylenol and Datril). Note that the ideal points are more prevalent in Quadrants III and IV of the joint space nearer the heavily advertised brands for both aspirin and aspirin substitutes. Figure 2 presents a plot of $\mathbf{W} = \| w_{it} \|$ which shows some similarity of the pattern of weights for each of the three maladies. In all cases, the second dimension possesses almost twice the salience as the first, indicating that the

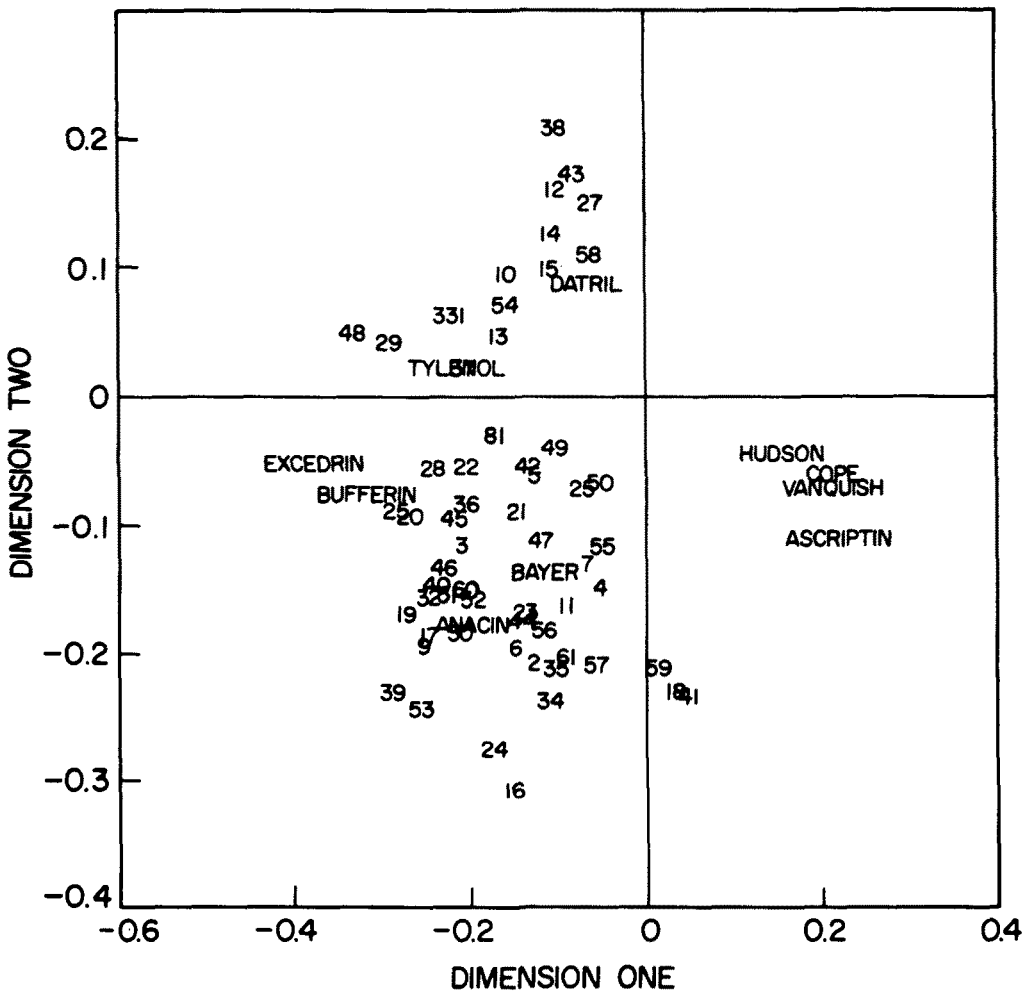

FIGURE 1
Joint Space Plot for Pain Reliever Data.

aspirin-nonaspirin dimension is more important than the market share/advertising dimension in all three situations for the students.

The Grambsch, Clark, DeSarbo, and Rothkopf Reading Profile Data

Grambsch, Clark, DeSarbo, and Rothkopf (1983) report an experiment and subsequent data analysis to investigate differences among 39 subjects in horizontal eye movement patterns elicited in a variety of reading tasks. Subjects were asked to engage in six different reading tasks on two occasions separated by at least one week. On each occasion, readers inspected nearly 20,000 words, half drawn from relatively easy and half from difficult text, spending roughly 100 minutes on the total task. The collection and analysis of eye-movement data from such an extensive observation period was done via a special computer-based technique (see Fisher and Rothkopf, 1982) in order to measure and analyze the very large quantities of data that were so produced. Electrodes placed near the eyes relayed data to a computer during reading. The six reading tasks varied in the degree of text interpretation required for successful execution and in the number of text elements that had to be considered at any given time.

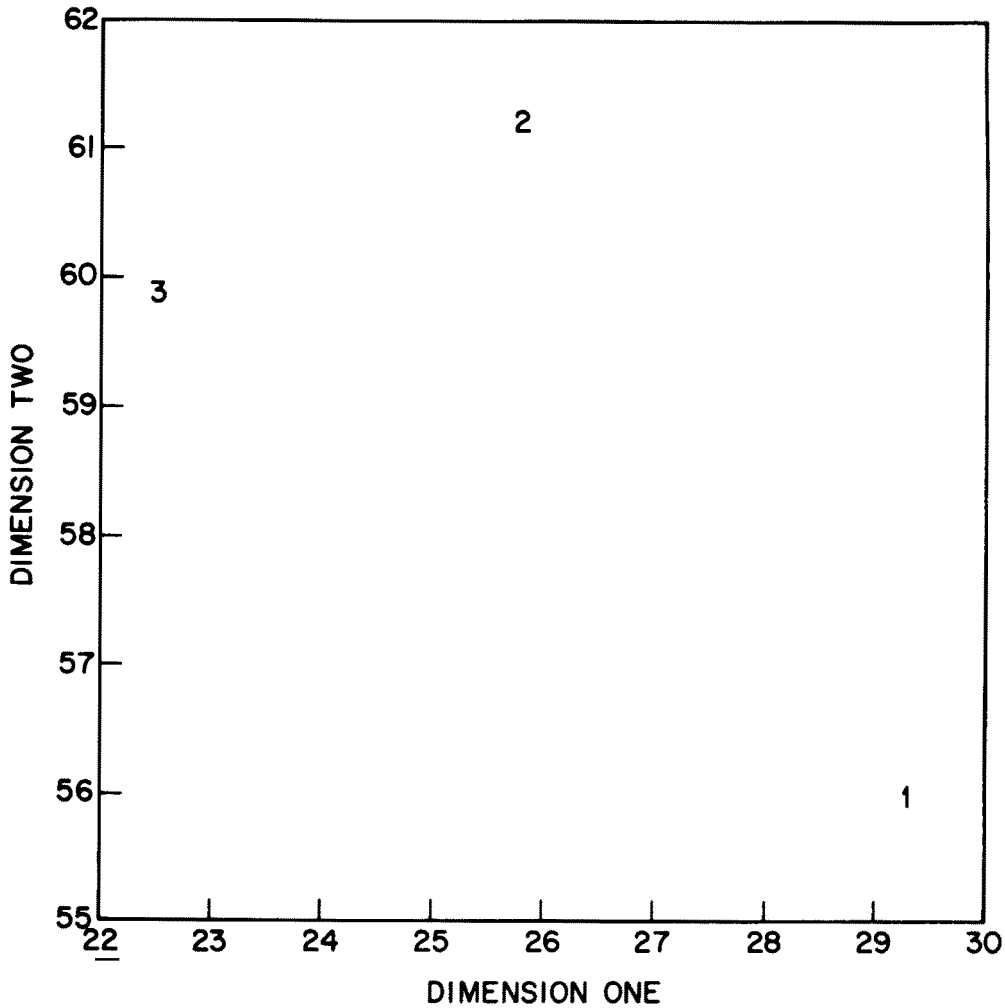

FIGURE 2
Weights Plot for Pain Reliever Data.

The method that was used to investigate individual differences in reading had two important ingredients.

1. Reading was characterized by several dependent measures. The following reading characteristics were tabulated: (a) total reading time (RT), (b) number of left-right saccades (LRS), (c) number of right-left saccades other than backsweeps (RLS), (d) number of backsweeps (BSW), (e) duration of each fixation (FIX), (f) amplitudes of left-right saccades (LRA), and (g) amplitudes of right-left saccades (RLA). Reading time was defined as the total recorded time in milliseconds from the first backsweep to the last backsweep per 1000 letter spaces. Left to right saccades were defined as the number of left to right saccades (any left to right eye movement from first to last backsweep) per 1000 letter spaces. Right to left saccades are the number of right to left saccades (any right to left eye movement less than one-half line excursion) per 1000 letter spaces. Left to right amplitude is the sum of the peak velocities of left to right saccades for experimental treatment divided by the total number of left to right saccades for that treatment. Right to left amplitude is the sum of the peak velocities of right to left saccades for an experimental treatment divided by the total number of right to left saccades for that treatment. Peak velocity

values used in calculations were actually digitized voltages. These digitized voltages were approximately linear with saccadic excursions. Backsweeps were defined as the total number, per 1000 letter spaces, of right to left saccades of greater than one-half line excursion. Fixation time was defined as the total time that the eye exhibited no movement (in milliseconds) divided by the number of fixations.

2. Several different reading tasks were used. This was done to find out (a) how average eye movement measures differed among tasks, and (b) whether performance differences among tasks tended to be the same from reader to reader. There were four search tasks and two comprehension tasks.

The Nonword Search Task (Task 1) required subjects to search a slide for the presence of a pronounceable one or two syllable nonword, such as HESH. When a nonword was present on a slide, it replaced a content word for the original text.

The Verbatim Search Task (Task 2) required subjects to search a slide for the presence of a particular one or two syllable word, such as SOFA. Special care was taken in order to achieve a degree of similarity between Verbatim Search target items and Synonym Search target: the class of Verbatim Search items was restricted to words that each had close synonyms.

The Synonym Search Task (Task 3) required subjects to search a slide for a one or two syllable synonym of a particular one or two syllable word, such as SMALL (LITTLE).

In the Information Task (Task 4), subjects were shown a detailed question, and were instructed to read the following slide in order to determine whether the information on that slide contained an answer to the question. One question, for example, was "What kinds of plants were probably the first ones to exist on land?"

The Information Memory Task (Task 5) was similar to the Information Search task except that subjects began a trial by reading a single slide for comprehension. After they finished reading, and after the slide was out of sight, they were shown a detailed question and were instructed to determine whether the question could have been answered from the information contained on the slide.

A final task, Learning (Task 6), required subjects to read a passage of eight slides for general comprehension. These passages were approximately 1200 words long, and were extracted in their entirety from contiguous text in the original source material. Each passage had a title and dealt with a uniform topic. Subjects read these passages in an uninterrupted fashion, proceeding at their own pace to the next slide when they had finished the previous one. At the end of a passage subjects were given an eight-item, short-answer comprehension test. A single Passage Comprehension task included an eight slide passage of easy reading material and an eight slide passage of difficult reading material.

Thus, the three-way, three-mode profile data collected was a subject (39 rows) by measure (7 columns) by task (6 slices) array. The variables (measures) were standardized across the entire array to zero mean and unit variance prior to analysis since the variables were measured in different units and similar normalizations were performed in Grambsch, et al. (1983). These analyses also suggested considerable differences in reading style over reading task suggesting the need for the floating ideal point option in three way unfolding analysis.

The analysis was conducted in 1–4 dimensions using the floating ideal point option with γ_{ijk} defined in (6) for a sequence of p values (0, 1, 2, 3, 4, 5). Based upon goodness-of-fit values and interpretation, $p = 2$ and $T = 1$ provided a parsimonious, nondegenerate solution accounting for 74.4% of the variance. The 39×6 matrix of floating ideal points was clustered using Johnson's (1967) complete linkage method to examine if any groups of subjects were found with similar reading style. Three clusters were identified on the basis of the resulting dendogram and an investigation of the respective group means and stan-

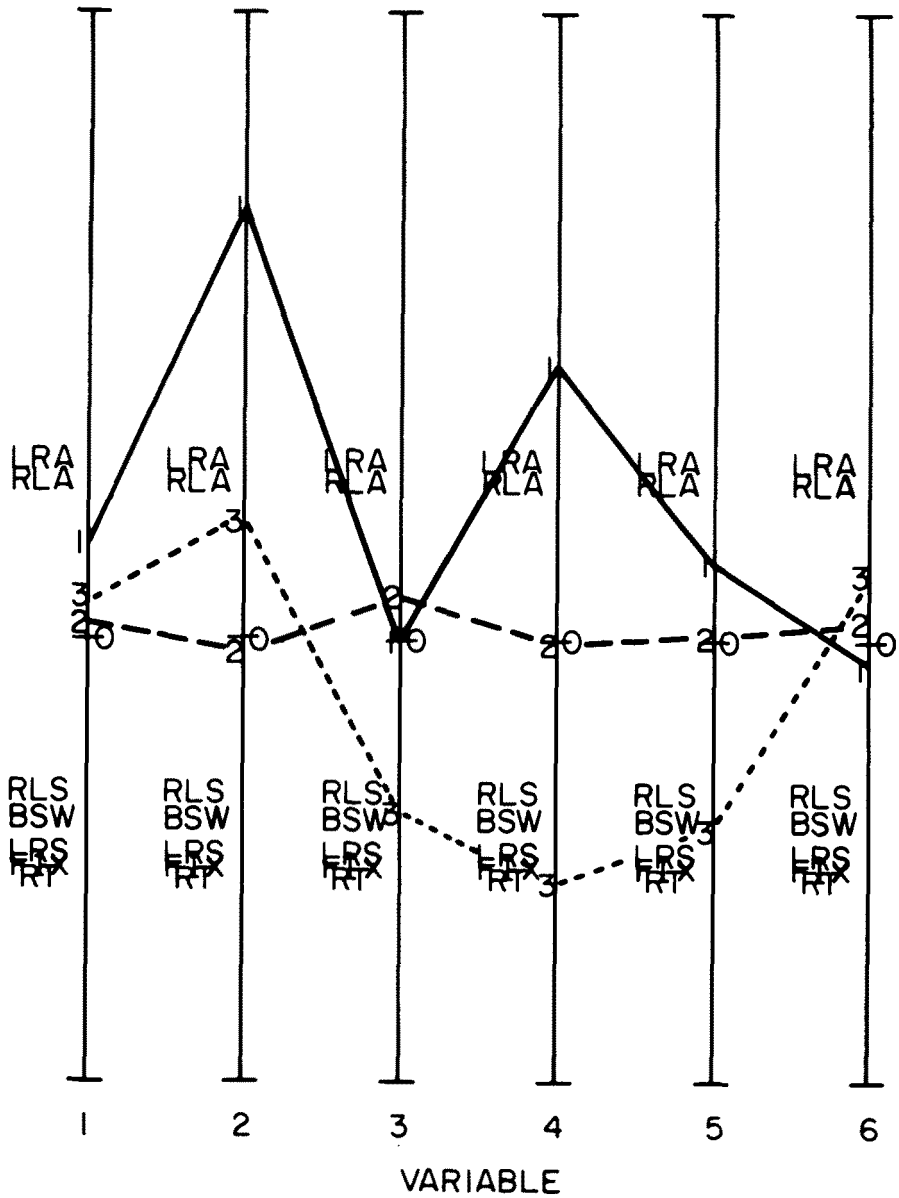

FIGURE 3
Joint Space Plot with Floating Ideal Points for Reading Data.

dard deviations, and their mean ideal point values are presented in Figure 3 with the stimuli (variable) locations for the one-dimensional solution for six reading tasks.

The one dimension can be easily interpreted as a reading effort dimension where the saccadic amplitude measures (velocity) is on the upper or positive end of the dimension while the gaze time measures (reading time, number of saccades, number of backsweeps, and fixation time) are on the lower or negative end of the dimension. The mean ideal point scores for each of the three groups are connected by their corresponding lines for each of the six tasks. A group's style of reading can thus be characterized by the proximity of the group's mean ideal point with the stimulus/variable location.

Group 1 appears to be the slowest reading group of the three, especially in the Verb-

atim and Information Search (Task 2 and 4) where they have lower velocities and higher gaze times and rereading. Group 3 appears to be the fastest reading group, especially in Synonym Search, and Information Memory Tasks (Tasks 3, 4, and 5). Their styles here are characterized as high velocity and low gaze times. Group 2 appears to be unaffected by any of the experimental tasks, displaying a relatively constant reading effort over all six tasks.

Thus, the three-way unfolding results render insights into individual differences in reading styles providing a method for graphically examining how aspects of reading effort differ for different subjects over different reading tasks.

The Miller-Nicely Data

Miller and Nicely (1955) collected data on confusions among 16 English consonants under each of several conditions of noise (varying the signal-to-noise ratio), low-pass filtering (filtering out acoustical energy in the higher frequencies), and high-pass filtering (filtering out acoustical energy in the lower frequencies). The subjects listened to speakers read c-v syllables (each syllable consisted of one of the 16 consonants followed by the vowel *a* as in f*a*ther), and tried to identify the consonant they heard after each syllable was spoken. In each of the 17 experimental conditions the speech was acoustically degraded in a different manner. A matrix of frequencies of stimulus-response confusions was derived for the data in each experimental condition. Note all 17 nonsymmetric matrices are presented in the original Miller and Nicely paper.

We chose to sort the 17 proximity (confusions) matrices into three general conditions: added noise, low-pass filtering, and high-pass filtering, and average the entries of the matrices in the three conditions producing a 3 (degradation condition) × 16 (consonants) × 16 (consonants) averaged proximity array. The scale was reversed to transform them into distance-like numbers and a grand mean was removed from the array.

The analysis was conducted in 1–6 dimensions with γ_{ijk} defined in (6) for each matrix slice (not over rows) for a sequence of $p = (0, 1, 2, 3, 4)$. Based upon goodness-of-fit measures and interpretation, two dimensions appeared to render a parsimonious description of the structure of the data with $p = 3$. Figure 4 presents the joint space of row and column consonants, where respective identical row and column consonants are connected by line segments. The length of the line segment indicates the degree of confusability of the consonant.

Dimension 1 separates the voiced consonants (those which, when spoken, produce vocal cord vibration) such as d, g, dz, z, b, v, and th from their voiceless cognates t, k, sh, s, p, f, and θ. The nasals (m, n), which do not have voiceless cognates, are grouped with the other voiced consonants on this dimension. The second dimension represents a manner of articulation dimension where nasals (m, n), unvoiced fricatives (θ, f, s, sh), voiced fricatives and stops (d, g, dz, z, b, v, th) and unvoiced stops (p, k, t) are separated. The two dimensional space is quite similar to that found by Shepard (1972) in applying a two-way multidimensional scaling method to a confusions matrix averaged over the first six conditions.

Since our joint space represents both row and column consonants (connected by line segments), we can examine Figure 4 to investigate how accurately each transmitted consonant is received. We find that the nasals m and n and fricatives z and dz are least confused. Voiced consonants d and b appear to be the most confused.

Finally, Figure 5 presents the plot of the condition weights for this analysis. The high pass filtering condition (3) weights the manner of articulation dimension slightly more heavily than the voiced-unvoiced dimension, while the noise (1) and low pass filtering (2) condition (especially) weight the voice-unvoiced dimension more heavily.

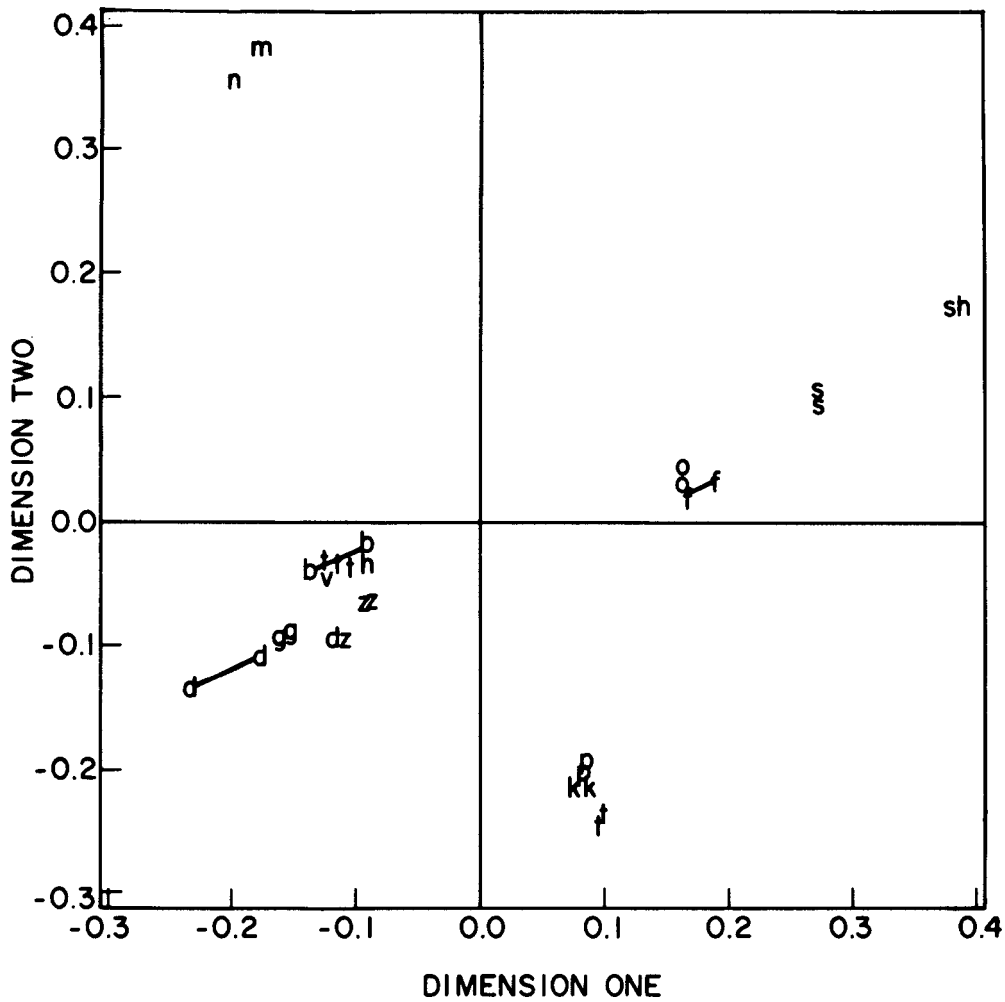

FIGURE 4
Joint Space Plot for Miller-Nicely Data.

5. Discussion

We have presented a detailed description of the three-way unfolding model and the weighted least-squares algorithm for fitting it. Monte Carlo results were also presented illustrating the apparent robustness of the performance of the algorithm when some eleven factors were experimentally varied. Finally, three applications were presented with parsimonious, nondegenerate, and interpretable solutions. Still, many unanswered questions remain as opportunities for further research.

Perhaps the most important research question that is still left unanswered is "what really causes degenerate solutions in unfolding?". While the proposed approach of utilizing a weighted loss function does appear tentatively to render nondegenerate solutions, additional research is required to understand how and why. Are degenerate solutions due to particular error structures in the data, poorly determined parameter estimates, erratically shaped loss functions, or inadequate algorithms?

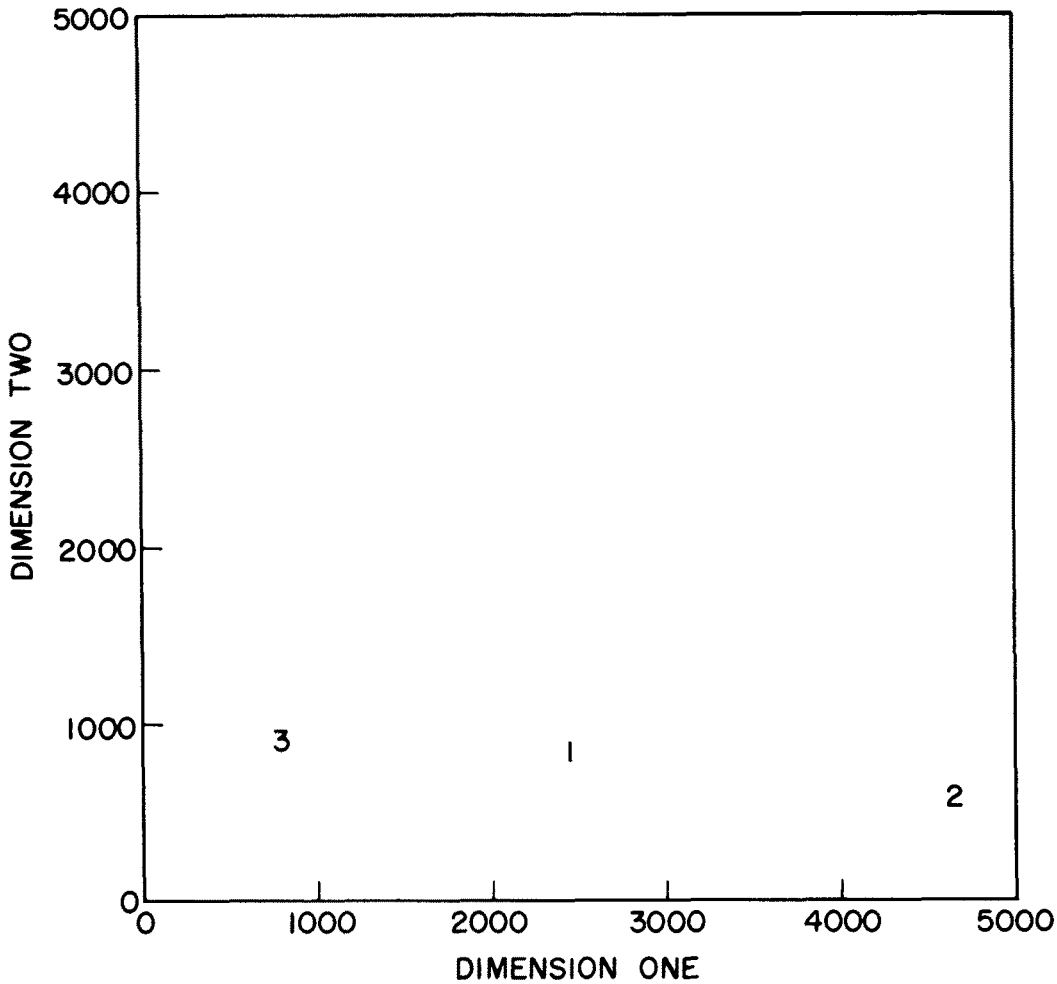

FIGURE 5
Weights Plot for Miller-Nicely data.

Another related question concerns the choice of the weighting function γ_{ijk}. Some guidelines can be established to rule out certain *general* forms of γ_{ijk} based on conditionality, reliability assumptions, preprocessing, etc. Yet, the choice of a *specific* γ_{ijk} (especially p) still remains a trial and error procedure. While the applications described in the paper suggest a p of 2 or 3 for γ_{ijk} defined in (6), more experience with the procedure must be obtained with more data sets before this can be a general recommendation. Further Monte Carlo research must also be performed to examine what the effect the choice of γ_{ijk} has on configuration recovery.

Obviously, this three-way unfolding technique can be applied to other substantive areas, especially in marketing. For example, this new methodology could be employed to examine children's preferences for various types of television shows over different situations (time of day, which family members are present, etc.). Or the model could be utilized to analyze people's preferences for various competing retail stores/outlets over different situations (Christmas gifts, spouse's birthday present, etc.). Still another possible application involves the analysis of dyadic relationships in personal selling where one might be

interested in whether certain types of situations lend themselves to a particular type of exchange.

In addition, this three-way unfolding method need not be restricted to merely situational analysis. For example, the technique could be successfully employed in testing product-advertising congruence where subjects could evaluate the most "appropriate" advertisement displayed for different products tested. Similarly, one may utilize such a model to investigate people's perception of particular celebrities advertising various products (e.g., one could attempt to select which of many products tested Arnold Palmer could most effectively endorse). In pricing research, one could utilize such a model to examine evaluations/preferences of various brands of products under different pricing structures or rate plans. Applications to product management or product positioning involve the possible analysis of consumer preferences for various products with various combinations of features/attributes. Another research application is the use of three-way unfolding as a *dynamic mapping tool* for "before and after" experimental treatments, to examine the treatments' effects on elicited preferences. One could also use three-way unfolding to examine the brand-switching behavior in longitudinal brand-switching matrices.

Applications also exist outside of psychology and marketing in such other social and behavioral sciences as economics, sociology, political science, anthropology, etc. This methodology should be a rather useful approach for analyzing the data obtained from such applications.

A number of interesting areas for future research appear to follow as a consequence of the three-way metric unfolding model:

1. The development of a nonmetric version of this technique to handle ordinal scaled data. (The results reported on potential degeneracies in two-way nonmetric unfolding by Kruskal and Carroll, 1969, should be carefully attended to, however.) Closely related to this nonmetric development is the development of extensions enabling the handling of (row or column) conditional data in an appropriate manner.
2. The development of three-way mixed models that could accommodate mixtures of vector and unfolding (ideal point) representations.
3. Development of viable statistical inference (via distribution theory or bootstrapping) for such scaling procedures.
4. Generalization of unfolding to N-way arrays.
5. Development of three-way unfolding subject to constraints on any or all of the data modes (stimuli, subjects, and/or situations), thereby generalizing DeSarbo and Rao's (1983) constrained approach to unfolding.
6. Exploration of other possible numerical procedures to fit the data in a more efficient manner.
7. Further empirical and numerical research into specifying a "best" γ_{ijk} weighting function. We have merely suggested a number of possible different weighting schemes in analyzing three-way data. One particular form was successful in use with the three applications presented. Further research is required into investigating the performance of other equally justifiable weighting functions.

Appendix A
The Candecomp Based Algorithm for Three-Way Unfolding

Scalar Products and CANDECOMP

The first step entails double centering each Δ_i matrix of raw dispreference values, and then converting this into scalar products by use of a nonsymmetric generalization of Torgerson's (1958) standard procedure:

$$s_{ijk} = -\tfrac{1}{2}(\delta_{ijk} - \delta_{i\cdot k} - \delta_{ij\cdot} + \delta_{i\cdot\cdot}), \tag{A1}$$

where:

$$\delta_{i\cdot k} = \frac{1}{J}\sum_{j=1}^{J}\delta_{ijk},$$

$$\delta_{ij\cdot} = \frac{1}{K}\sum_{k=1}^{K}\delta_{ijk},$$

$$\delta_{i\cdot\cdot} = \frac{1}{JK}\sum_{j=1}^{J}\sum_{k=1}^{K}\delta_{ijk},$$

s_{ijk} = scalar product between subject j and stimulus k for situation i.

It can be shown that, after that transformation, the model can be expressed in scalar products form as:

$$s_{ijk} \cong \sum_{t=1}^{T} w_{it}^{*} y_{jt}^{*} x_{kt}^{*} \tag{A2}$$

(where w_{it}^{*}, y_{jt}^{*}, and x_{it}^{*} will be discussed below). Equation (A2) is the form of the three-way CANDECOMP model (Carroll & Chang, 1970), and one may therefore analyze $\mathbf{S} = \| s_{ijk} \|$ by such a procedure to estimate **X**, **Y** and **W** via the nonlinear iterative least squares algorithm utilized in CANDECOMP. However, because there are indeterminacies with respect to origin and allowable diagonal (scale) transformations associated with this three-way decomposition, one must then solve for scale parameters a_t, b_t, for the "stimulus" and "subject" parameters, and a set of parameters defining a shift of origin of one of the two sets of points, c_t. Thus denoting w_{it}^{*}, y_{jt}^{*}, and x_{kt}^{*} as the parameters obtained from the CANDECOMP model on the derived scalar products s_{ijk}, one obtains:

$$\delta_{ijk} \cong d_{ijk}^{2} + \alpha_i = \sum_{t=1}^{T} w_{it}(a_t(y_{jk}^{*} - c_t) - b_t x_{kt}^{*})^2 + \alpha_i, \tag{A3}$$

where one set of points (the x_{kt}^{*}'s) is fixed at its centroid, and the origin of the remaining set (the y_{jt}^{*}'s) must be solved for. Note that in (A3) above, the w_{it}, y_{jt} and x_{kt} parameters are related to the w_{it}^{*}, y_{jt}^{*} and x_{kt}^{*} estimates by the following:

$$y_{jt} = a_t(y_{jt}^{*} - c_t),$$

$$x_{kt} = b_t x_{kt}^{*},$$

$$w_{it} = \frac{w_{it}^{*}}{a_t b_t}. \tag{A4}$$

In the initial estimation scheme for three-way unfolding, the CANDECOMP procedure is utilized to estimate the w_{it}^{*}, y_{jt}^{*} and x_{kt}^{*} parameters. Then, further numerical optimization methods, to be discussed below, are used to estimate the a_t, b_t, and c_t parameters thus yielding estimates of y_{jt} and x_{kt}. The w_t's are estimated separately by regression techniques. (In the case in which the δ_{ijk}'s are assumed to be ratio scale estimates of squared distances, the α_i's are set equal to zero. In the case of interval scale δ_{ijk}'s, the α_i's are estimated by simply adding an extra term to the regression equation involved in estimating the w_{it}'s.) The reason for the need to estimate the a_t, b_t, and c_t parameters is that, in the transformation to scalar products form in (A1), the double centering operation arbitrarily shifts the centroids of the two sets of points (the x's and the y's) to the origin of the coordinate system. While the overall origin of the coordinate system is arbitrary (since we are dealing with a Euclidean distance model), the relative positions of these two sets of points vis à vis one another are quite relevant. We can solve for this relative placement of the centroids by fixing one set (the x's) at the origin, and solving for that of the other set (the y's). This is the function of the c_t parameters. The scale parameters (a_t, b_t) are relevant

because in the derived scalar products model we are dealing with, after double centering each Δ_i, the coordinates of each set of points are defined only up to a diagonal transformation (as are the w_{it}^*'s). The **W** matrix gets re-estimated, as we shall see shortly, because the CANDECOMP procedure uses a different loss function than does the Newton-Raphson procedure for estimating the translation and scale parameters. The procedures for the estimation of the a_t, b_t, and c_t parameters are a three-way analogue of procedures used in Schönemann's (1970) two-way metric unfolding for estimating both linear transformations of the stimulus and subject spaces. In the two-way case, an analytic solution was available as described in Schönemann (1970) for estimating these transformations (an affine transformation of one set combined with a linear transformation of the other, the linear transformation components of the two transformations being linked by an "inverse-adjoint" relationship). In the three-way case unfortunately, no such analytic solution is available. Thus, we were required to rely on a less elegant numerical procedure based on iterative gradient procedures. However, a partially compensating advantage is afforded by the fact that the linear transformation in the three-way case can be restricted to diagonal transformations, which greatly simplifies that component of the estimation scheme. While the two diagonal transformations should ideally be inversely related (the inverse adjoint of a diagonal matrix is identical to its inverse), the procedure for estimating these parameters utilizes a different loss function and we do not explicitly constrain them to be so related.

Estimation Procedure for Translation Parameters

In this routine, we fix all parameters (using CANDECOMP estimates) except the c_t's, at their current estimates, and we solve for these entities via:

$$\delta_{ijk} - \alpha_i \cong d_{ijk}^2 = \sum_{t=1}^{T} w_{it}^* [(y_{jt}^* - c_t) - x_{kt}^*]^2, \tag{A5}$$

where estimates of c_t minimize the following sum of squares:

$$R = \sum_i \sum_j \sum_k [\delta_{ijk} - \sum w_{it}^* (y_{jt}^* - x_{kt}^*)^2 - \sum_t c_t^2 w_{it}^* + 2 \sum_t c_t w_{it}^* (y_{jt}^* - x_{kt}^*) - a_i]^2. \tag{A6}$$

Since the objective function (R) in (A6) is continuous and twice differentiable, a modified gradient technique, the Newton-Raphson method (Himmelblau, 1972), appears to be quite efficient, especially since R "resembles" a quadratic sum of squares function, and since such gradient estimation procedures converge in only one iteration when dealing with quadratic unconstrained objective functions ("quadratic convergence"). Letting $\boldsymbol{\theta}$ be the vector of parameters to be estimated, the transition from $\boldsymbol{\theta}^{(l)}$ to $\boldsymbol{\theta}^{(l+1)}$ for the Newton-Raphson method is:

$$\boldsymbol{\theta}^{(l+1)} = \boldsymbol{\theta}^{(l)} - \frac{\lambda^{(l)} [\nabla^2 f(\boldsymbol{\theta}^{(l)})]^{-1} \nabla f(\boldsymbol{\theta}^{(l)})}{\| [\nabla^2 f(\boldsymbol{\theta}^{(l)})]^{-1} [\nabla f(\boldsymbol{\theta}^{(l)})] \|}, \tag{A7}$$

where:

$\boldsymbol{\theta}^{(l+1)}$ = value of $\boldsymbol{\theta}$ parameters at $(l+1)$-th transition,

$\boldsymbol{\theta}^{(l)}$ = value of $\boldsymbol{\theta}$ parameters at l-th transition,

$\lambda^{(l)}$ = step size parameter,

$\nabla f(\boldsymbol{\theta}^{(l)})$ = gradient evaluated at $\boldsymbol{\theta}^{(l)}$,

$\nabla^2 f(\boldsymbol{\theta}^{(l)})$ = Hessian matrix of second derivatives evaluated at $\boldsymbol{\theta}^{(l)}$.

Because a positive definite Hessian matrix is required for convergence, the Green-

stadt approximation subroutine (Greenstadt, 1967) is utilized to force the Hessian to be positive definite at each stage of the minimization. The step size parameter λ is obtained by either a standard Fibonacci (Cooper & Steinberg, 1970) line search procedure or a quadratic line search routine (Carroll & Pruzansky, 1980) by option of the user. Note, the Newton-Raphson modified gradient procedure is applied iteratively either until some convergence criterion is satisfied or a preset maximum number of iterations has been attained. Once this stage's estimates ($\hat{c}_t$) have been obtained via this procedure, we define new y_{jt}^{**} by

$$y_{jt}^{**} = y_{jt}^{*} - \hat{c}_t. \tag{A8}$$

Estimation of Scale Parameters

In this procedure, all estimates except a_t and b_t are kept at their current values, and these scale parameters are solved for so that:

$$\delta_{ijk} - \alpha_i \cong d_{ijk}^2 = \sum_{t=1}^{T} w_{it}^{*}(a_t y_{jt}^{**} - b_t x_{kt}^{*})^2. \tag{A9}$$

Specifically one attempts to find scale estimates so as to minimize the following least-squares loss function:

$$Q = \sum_i \sum_j \sum_k \left[\delta_{ijk} + 2\sum_t a_t b_t w_{it}^{*} y_{jt}^{**} x_{kt}^{*} - \sum_t a_t^2 w_{it}^{*} y_{jt}^{**2} - \sum_t b_t^2 w_{it}^{*} x_{kt}^{*2} - \alpha_i\right]^2. \tag{A10}$$

Again, the Newton-Raphson method is utilized iteratively, with the Greenstadt approximation and Fibonacci or quadratic line search procedures. At the end of this stage, new estimates y_{jt}^{***} and x_{kt}^{***} are defined as follows:

$$y_{jt}^{***} = \hat{a}_t y_{jt}^{**} = \hat{a}_t(y_{jt}^{*} - \hat{c}_t),$$

and

$$x_{kt}^{***} = \hat{b}_t x_{kt}^{*}. \tag{A11}$$

Estimation of Situation Weights and Additive Constants

In this subroutine, all parameters except the w_{it}^{*} are held fixed at their current values. To estimate the w_{it} in (1) (assuming, for now, that $\alpha_i = 0, \forall_i$), we first express:

$$\begin{aligned} \delta_{ijk} \cong d_{ijk}^2 &= \sum_{t=1}^{T} w_{it}(y_{jt}^{***} - x_{kt}^{***})^2, \\ &= \sum_{t=1}^{T} w_{it} p_{jkt}, \end{aligned} \tag{A12}$$

where

$$p_{jkt} = (y_{jt}^{***} - x_{kt}^{***})^2.$$

Now one can estimate w_{it} via OLS where we let $\Delta = \| \delta_{i(jk)} \|$ be an $I \times JK$ matrix and let $\mathbf{P} = \| p_{(jk)t} \|$ be a $(JK) \times T$ matrix. We then obtain:

$$\Delta \cong \mathbf{WP}', \tag{A13}$$

and a new estimate of $\mathbf{W}$ ($\mathbf{W}^{***}$) as obtained via

$$\mathbf{W}^{***} = \Delta\mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}. \tag{A14}$$

Note, if one desires to estimate an additive constant (in dealing with interval scaled preference values), the procedure simply appends a column of 1's onto the $\mathbf{P}$ matrix, thereby

estimating an intercept term corresponding to an estimate of the additive constant (α_i) for a given $J \times K$ (e.g., subjects by stimuli) matrix.

Termination Criterion

At the end of the previous estimation phase (one "major" iteration completed), we return to the phase for estimating translation parameters (replacing the w^*, x^* and y^* values with w^{***}, x^{***} and y^{***} as defined above) and start the cycle over, solving first for $\hat{c}_t$ with all parameters fixed at current values, etcetera, continuing the iterative reestimation until either: (a). the maximum number of preset iterations has been attained; or (b). we have obtained convergence with respect to our variance accounted-for measure (V.A.F.):

$$\text{V.A.F.} = 1 - \frac{\sum_i \sum_j \sum_k (\delta_{ijk} - \hat{d}^2_{ijk})^2}{\sum_i \sum_j \sum_k (\delta_{ijk} - \bar{\delta}_{i..})^2} \tag{A15}$$

for interval scale data, or sums of squares accounted for (S.S.A.F.)

$$\text{S.S.A.F.} = 1 - \frac{\sum_i \sum_j \sum_k (\delta_{ijk} - \hat{d}^2_{ijk})^2}{\sum_i \sum_j \sum_k \delta^2_{ijk}} \tag{A16}$$

for ratio scale data, where:

δ_{ijk} = input data,

$\bar{\delta}_{i..}$ = grand mean of input data for slice i, and

$\hat{d}^2_{ijk} = \sum_t w_{it}(y_{jt} - x_{kt})^2$ = predicted squared distances.

(An earlier version of the procedure utilized $(\text{SQ})^2$ as a sums-of-squares accounted-for measure, where:

$$\text{SQ} = \frac{\sum_{i,j,j} \delta_{ijk} \hat{d}^2_{ijk}}{\left(\sum_{i,j,k} \delta^2_{ijk} \sum_{i,j,k} d^4_{ijk}\right)^{1/2}} \tag{A17}$$

SQ, as defined above, can be viewed as the uncentered correlation function. S.S.A.F., as defined above, $= (\text{SQ})^2$, just as V.A.F. $= r^2$, where r is the (centered) produce moment correlation).

Note, since the objective function to be minimized at each stage (of the gradient procedure) is the same sum of squared differences between observed and estimated values from the model, this least squares fit measure must be monotonically decreasing. We would therefore expect our variance accounted for (or sum of squares accounted for) measure to be monotonically increasing, being an inverse monotonic function of the residual sum of squares, thereby assuming convergence (for formal proof, see Courant, 1965, p. 61).

Normalization

At this point, after convergence, if the user had selected the "fixed" ideal point option, where ideal points do not change by situation but rather are stretched or shrunk by situation weights (w_{it}'s) then **X**, **Y**, and **W** matrices are normalized and plotted. Normalization here amounts to normalizing to unit sum of squares the composite matrix defined

by stacking **X** and **Y** into a single $(J + K) \times T$ matrix and then redefining the matrix of situation weights **W** so that the differential slice variance is reflected in these weights.

Floating Ideal Point Option

If the floating ideal point option had been selected, a generalization in which ideal points are assumed to vary over situations, an additional stage is required to estimate a set of ideal points for *each situation.* In effect, this treats the fixed ideal point model as a special case where the ideal points are the *same* for each situation. The development of this part of the algorithm parallels that of Carroll (1972) PREFMAP two-way unfolding model. We again assume that preference values are linearly related to squared distances:

$$\delta_{ijk} = a_{ij} d^2_{ijk} + b_{ij} + e_{ijk}, \tag{A18}$$

where:

δ_{ijk} = preference value,

d^2_{ijk} = squared weighted distance = $\sum_{t=1}^{T} w_{it}(y_{(i)jt} - x_{kt})^2$,

$y_{(i)jt}$ = the t-th coordinate of ideal point for subject j in situation i,

a_{ij} = scale parameter for subject j in situation i,

b_{ij} = intercept parameter for subject j in situation i, and

e_{ijk} = error term.

Note, w_{it} and x_{kt} are as previously defined for the fixed ideal point case. Substituting for d^2_{ijk} in (A18) (dropping the error term and substituting "$\cong$" for "$=$"), one obtains:

$$\begin{aligned}\delta_{ij} &\cong a_{ij} \sum_{t=1}^{T} w_{it}(y_{(i)jt} - x_{kt})^2 + b_{ij}, \\ &\cong a_{ij} \sum_{t} w_{it} y^2_{(i)jt} - 2a_{ij} \sum_{t} w_{it} y_{(i)jt} x_{kt} + a_{ij} \sum_{t} w_{it} x^2_{kt} + b_{ij}.\end{aligned} \tag{A19}$$

Defining a constant term z_{ij} as:

$$z_{ij} = a_{ij} \sum_{t} w_{it} y_{(i)jt} + b_{ij}, \tag{A20}$$

we are left with

$$\delta_{ijk} \cong a_{ij} \sum_{t} (w_{it} x^2_{kt}) + \sum_{t} q_{ijt}(w_{it} x_{kt}) + z_{ij}, \tag{A21}$$

where:

$$q_{ijt} = -2a_{ij} y_{(i)jt}, \tag{A22}$$

which is now in the form of a quadratic regression of the x's scaled by the w's to estimate the floating ideal points via:

$$y_{(i)jt} = -1/2(q_{ijt}/a_{ij}). \tag{A23}$$

Again, we can consider the fixed ideal point model as a special case of the floating ideal point case where the ideal points are the same for each situation (i.e., $y_{(i)jt} \equiv y_{jt}$ for all i). Also, because we are dealing with considerably more parameters in the floating case compared to the fixed case, we would expect the floating ideal point model to account for an

equal or greater amount of variance (than the fixed case). Similarly, after the estimation procedure, normalizations and subsequent plotting comprise the final stages.

Appendix B
The Weighted Least Squares Algorithm for Three-Way Unfolding

1. Starting estimates of **W**, **X**, **Y**, and α are obtained randomly, user-given, via use of the CANDECOMP based algorithm described in Appendix A, or via a "close" start which averages the three-way array Δ into a two-way array and performs two-way unfolding (an unrotated PREFMAP2 model 3 analysis with internally generated stimulus configuration) on it to obtain **X** and **Y**, with $\mathbf{W} = \| 1 \|$. Let IT (initially = 0) = IT + 1;
2. Estimate **X** (for internal analyses) holding all other parameter values fixed via a conjugate gradient method (Fletcher & Reeves, 1964) where the partial derivatives of the weighted loss function with respect to x_{kt} are:

$$\frac{\partial Z_2}{\partial x_{kt}} = 4 \sum_i \sum_j \gamma_{ijk} w_{it} (\delta_{ijk} - \hat{\delta}_{ijk})(y_{jt} - x_{kt}), \tag{B1}$$

where we shall use $\nabla\mathbf{X}$ as the partial derivative in (B1) above, whose elements are strung out in a large $KT = 1$ vector. Then, the conjugate gradient method used in this phase can be summarized as follows:

a. Set MIT = 1.
b. Set the final search direction $\mathbf{S}^{(1)} = -\nabla\mathbf{X}^{(1)}$ in this first "minor" iteration.
c. Find $\mathbf{X}^{(2)}$ according to:

$$\mathbf{X}^{(2)} = \mathbf{X}^{(1)} + u^{(1)}\mathbf{S}^{(1)}, \tag{B2}$$

where $u^{(1)}$ is the optimal step length in the direction $\mathbf{S}^{(1)}$ obtained by a quadratic interpolation line search procedure (Carroll & Pruzansky, 1980). Set MIT = 2.
d. Calculate $\nabla\mathbf{X}^{(\mathrm{MIT})}$ and set:

$$\mathbf{S}^{(\mathrm{MIT})} = -\nabla\mathbf{X}^{(\mathrm{MIT})} + \frac{(\nabla\mathbf{X}^{(\mathrm{MIT})})'(\nabla\mathbf{X}^{(\mathrm{MIT})})}{(\nabla\mathbf{X}^{(\mathrm{MIT}-1)})'(\nabla\mathbf{X}^{(\mathrm{MIT}-1)})}\mathbf{S}^{(\mathrm{MIT}-1)}. \tag{B3}$$

e. Compute the optimal step length $u^{(\mathrm{MIT})}$ in the direction $\mathbf{S}^{(\mathrm{MIT})}$, and find

$$\mathbf{X}^{(\mathrm{MIT}+1)} = \mathbf{X}^{(\mathrm{MIT})} + u^{(\mathrm{MIT})}\mathbf{S}^{(\mathrm{MIT})}.$$

f. If $\mathbf{X}^{(\mathrm{MIT})}$ is optimal, stop. Otherwise set MIT = MIT + 1 and go to step (d) above (i.e., undertake another minor iteration). Note that one would stop at this stage (f) if either:

1). MIT $\geq$ maximum number of minor iterations stipulated by the user,
2). $\| \nabla\overline{\mathbf{X}}^{(\mathrm{MIT})} \| <$ gradient convergence tolerance TOL, or
3). $(Z_2^{(\mathrm{MIT}-1)} - Z_2^{(\mathrm{MIT})}) <$ TOL.

It has been demonstrated empirically that conjugate gradient procedures can avoid the typical "cycling" often encountered with steepest descent algorithms. In addition, these procedures demonstrate valuable quadratic termination (Himmelblau, 1972) properites—i.e., conjugate gradient procedures will find the globally optimum solution for a quadratic loss function in n steps, where n is the number of parameters to be estimated.

3. Estimate **Y** via a similar conjugate gradient scheme holding values of all other parameters fixed.

4. Estimate $\mathbf{W}$ and α as in Phase 4 of the algorithm in Appendix A using weighted least-squares to accommodate the γ_{ijk}.
5. Termination criterion—we go to Step 6 if either:

   (a) $(Z_2^{(IT-1)} - Z_2^{(IT)}) < TOL$, or
   (b) IT $\geq$ maximum number of major iterations stipulated by the user.
   Otherwise, set IT = IT + 1 and go to Step 2.
6. Normalization—same as Step 6 in Appendix A.
7. Option for estimating floating ideal points—similar to Phase 7 of Appendix A but here using weighted least-squares to accommodate the γ_{ijk}.

References

Addelman, S. (1962). Orthogonal main-effect plans for asymmetrical factorial experiments. *Technometrics, 4,* 21–46.

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

Carroll, J. D. (1972). Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. Nerlove (Eds.), *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences* (Vol. I). New York: Academic Press.

Carroll, J. D., & Pruzansky, S. (1980). Discrete and hybrid scaling models. In E. D. Lantermann & H. Feger (Eds.), *Similarity and choice.* Bern: Hans Huber.

Coombs, C. H. (1964). *A Theory of Data.* New York: Wiley.

Cooper, L., & Steinberg, D. (1970). *Introduction to Methods of Optimization.* Philadelphia, PA: W. B. Saunders Co.

Courant, R. (1965). *Differential and Integral Calculus* (2nd ed., Vol. I), New York: Wiley.

DeSarbo, W. S. (1978). *Three-way unfolding and situational dependence in consumer preference analysis.* Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.

DeSarbo, W. S., & Carroll, J. D. (1980). Three-way unfolding and situational dependence in consumer preference analysis. In K. Bernhardt, I. Dolich, M. Etzel, T. Kinnear, W. Perreault, & K. Roering (Eds.), *The changing marketing environment*: New theories and applications (pp. 321–325). Chicago: American Marketing Association.

DeSarbo, W. S., & Carroll, J. D. (1981). Three-way metric unfolding. *Proceedings of the ORSA/TIMS Market Measurement Conference.* New York, NY.

DeSarbo, W. S., & Rao, V. R. (1983). *A constrained unfolding model for product positioning.* Unpublished manuscript, Bell Laboratories, Murray Hill, NJ.

DeSarbo, W. S., & Rao, V. R. (1984). *GENFOLD2: A Set of Models and Algorithms for the GENeral UnFOLDing Analysis of Preference/Dominance Data. Journal of Classification, 1,* 147–186.

Fisher, D. G., & Rothkopf, E. Z. (1982). An automated eye movement laboratory for on-line electro-oculography. *Behavior Research Methods and Instrumentation, 14,* 113–120.

Fletcher, R., & Reeves, C. M. (1964). Function minimization by conjugate gradients. *Computer Journal, 7,* 149–154.

Graef, J., & Spence, I. (1979). Using distance information in the design of large multidimensional scaling experiments, *Psychological Bulletin, 86,* 60–66.

Grambsch, P., Clark, L., DeSarbo, W. S., & Rothkopf, E. (1983). An analysis of individual differences in reading. Unpublished manuscript, Bell Laboratories, Murray Hill, NJ.

Greenstadt, J. (1967). Eigenvalue analysis for gradient procedures. *Mathematical Computation, 21,* 360.

Heiser, W. J. (1981). *Unfolding Analysis of Proximity Data.* Leiden, The Netherlands: Reprodienst Psychologie RUL.

Himmelblau, D. M. (1972). *Applied Nonlinear Programming.* New York: McGraw-Hill.

Isaac, P. D. (1980). Considerations in selection of stimulus pairs for data collection in MDS. In R, B. Golledge & J. N. Rayner (Eds.), *Multidimensional Analysis of Large Data Sets* (pp. 326–341). Columbus: Ohio State University Press.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika, 32,* 241–254.

Kruskal, J. B., & Carroll, J. D. (1969). Geometrical models and badness-of-fit functions. In P. R. Krishnaiah (Ed.), *Multivariate Analysis* (Vol. 2, pp. 639–671). New York: Academic Press.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusion among some English consonants. *Journal of the Acoustical Society of America, 27,* 338–352.

Schönemann, P. H. (1970). On metric multidimensional unfolding. *Psychometrika, 35*, 349–366.

Shepard, R. N. (1972). Psychological representation of speech sounds. In E. E. David & P. B. Denes (Eds.), *Human Communication: A Unified View*. New York: McGraw Hill.

Snedecor, G. W., & Cochran, W. G. (1981). *Statistical Methods* (7th ed.). Ames: Iowa State University Press.

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

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

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

Manuscript received 12/14/83
Revision received 6/1/84
Final version received 12/1/84