

A STOCHASTIC MULTIDIMENSIONAL SCALING VECTOR THRESHOLD MODEL FOR THE SPATIAL REPRESENTATION OF “PICK ANY/*N*” DATA

WAYNE S. DESARBO

GRADUATE SCHOOL OF BUSINESS
UNIVERSITY OF MICHIGAN

JAEWUN CHO

MARKETING DEPARTMENT
ARIZONA STATE UNIVERSITY

This paper presents a new stochastic multidimensional scaling vector threshold model designed to analyze “pick any/*n*” choice data (e.g., consumers rendering buy/no buy decisions concerning a number of actual products). A maximum likelihood procedure is formulated to estimate a joint space of both individuals (represented as vectors) and stimuli (represented as points). The relevant psychometric literature concerning the spatial treatment of such binary choice data is reviewed. The nonlinear probit type model is described, as well as the conjugate gradient procedure used to estimate parameters. Results of Monte Carlo analyses investigating the performance of this methodology with synthetic choice data sets are presented. An application concerning consumer choices for eleven competitive brands of soft drinks is discussed. Finally, directions for future research are presented in terms of further applications and generalizing the model to accommodate three-way choice data.

Key words: binary data analysis, multidimensional scaling, nonlinear probit model.

I. Introduction

The analysis of binary choice data has been the focus of a number of various methodologies in psychometrics and econometrics. The econometric literature primarily deals with modeling “pick 1/*n*” data in a nonspatial manner. The corresponding econometric models describing choices from a limited number of alternatives attempt to relate the conditional probabilities of a particular choice being made to various explanatory factors that include the attributes of the choice alternatives as well as the characteristics of the decision maker. Conditional logit (McFadden, 1976), conditional probit (Hausman & Wise, 1978), and so on, are typical approaches utilized in applied economic studies to analyze such data (see Chow, 1983; Maddala, 1986, for a review of such techniques).

The psychometric literature provides a different perspective to the analysis of such two-way binary choice data where interest typically lies in deriving a spatial representation of the elements of the rows and columns of such data (see Bock & Lieberman, 1970; Bartholomew, 1980; Christofferson, 1975; Coombs, 1964; Kruskal & Shepard, 1974; Muthén, 1978; 1981; Torgerson, 1958). Perhaps the most popular methodology traditionally utilized to spatially analyze such data is correspondence analysis (Benzécri, 1973) and its many variants. Correspondence analysis is typically applied to aggregated choice data in the form of a frequency matrix, although it can also accommodate the analysis of the raw binary data. It derives a joint space of row and column elements based

Requests for reprints should be sent to Wayne S. DeSarbo, Marketing and Statistic Departments, Graduate School of Business, University of Michigan, Ann Arbor, MI 48109.

on an eigenstructure analysis of the normalized frequency matrix. Greenacre (1984) and Lébart, Morineau, and Warwick (1984) discuss extensions of the procedure to multiple correspondence analysis. Other related optimal scaling approaches such as dual scaling methods (Nishisato, 1980) and homogeneity analysis (de Leeuw, 1973; Gifi, 1981a, 1981b; Heiser, 1981) have been shown (Tenenhaus & Young, 1985) to be equivalent forms of correspondence analysis. In fact, Levine's (1979) centroid method for "pick any" data can also be viewed as a special case of correspondence analysis where an alternative normalization is employed.

Recently, two unfolding methodologies have been introduced into the psychometric literature which have purposely been devised for the analysis of "pick any/ n " data. Takane's (1983) unfolding procedure treats such binary data as a special type of successive categories data in which there are only two response categories. Each item (stimulus), rather than each category of an item, is represented as a separate point, and individuals are assumed to choose or not choose the item according to its proximity to their respective ideal points. Takane develops an EM algorithm (see Dempster, Laird, & Rubin, 1977) to maximize a marginal likelihood in order to estimate the joint space. DeSarbo and Hoffman (1986, 1987) develop simple and weighted unfolding threshold models for the spatial representation of such "pick any/ n " data. Their methodology, in the simple unfolding case, estimates a joint space of ideal points for individuals, coordinate points for stimuli, and individual specific threshold values such that when the squared distance between individual i and stimulus j exceeds a threshold value c_i^* , the model predicts no choice. Similarly, when this squared distance is less than this threshold value, a choice is predicted. Thus, the proximity of a stimulus to an individual's ideal point indicates the degree or the magnitude of the probability of choice. DeSarbo and Hoffman devise a nonlinear logit formulation utilizing maximum likelihood procedures to estimate the relevant set of parameters. The vector threshold methodology to be presented in the next section can be viewed as a nonlinear probit, vector model extension of the DeSarbo and Hoffman unfolding methodology.

The goal of this research is to develop a new MDS based choice model for the estimation of a joint space of individuals and stimuli from pick any/ n binary choice data. Most of the techniques described thus far are data analytic procedures which posit no explicit theory on how choice is made. Few of these models provide statistical tests to identify dimensionality of "best" model representations. Only one allows for the possible reparameterization of individuals and/or stimuli to aid in the interpretation of the resulting solution and allow for predictions of individuals/stimuli (within the range of the calibration data) not included in the analysis. The new spatial model of choice developed in the next section can be described as a stochastic model utilizing a vector or scalar products representation (see also Bechtel, Tucker, & Chang, 1971; Cooper & Nakanishi, 1983; DeSarbo & Keramidas, 1983; DeSarbo, Oliver, & DeSoete, 1986; and DeSoete & Carroll, 1983; for other types of stochastic vector MDS models mostly applied to paired comparisons data). Based on utility theory via an indirect utility function (McFadden, 1976) which posits choice as a function of a latent variable and an error function, this formulation allows for (theoretically speaking) an asymptotic statistical test for identifying the appropriate dimensionality and for nested models testing. It can accommodate virtually any type of binary data, but is especially appropriate for "pick any/ n ", and has reparameterization options for individual and/or stimulus coordinates which allow for the prediction of individuals and/or stimuli not utilized in the analysis. We feel it is an improvement over most existing procedures in that its structure is more related to the actual choice process (rather than merely another data analytic procedure), the reparameterization options allow for the direct and simultaneous investigation of features, at-

tributes, demographic, psychographic, and other variables; also, the proposed methodology allows for predictive validation in a simple manner via the reparameterization options.

II. The Spatial Representation of Pick Any/ N Data

A. The Model

Let:

$t = 1, \dots, T$ dimensions (extracted in an MDS context);

$i = 1, \dots, I$ individuals (or subjects);

$j = 1, \dots, J$ stimuli (or products);

$$y_{ij} = \begin{cases} 1 & \text{if individual } i \text{ chooses stimulus } j, \\ 0 & \text{otherwise;} \end{cases}$$

p_{ij} = the probability individual i chooses stimulus j ;

a_{it} = the t -th vector terminus coordinate for individual i ;

c_i = an additive constant for individual i ;

b_{jt} = the t -th coordinate for stimulus j ,

We assume that the choice process of individual i choosing stimulus j is Bernoulli (independence is assumed across individuals and stimuli), with probability of choice given by p_{ij} . Unlike the conditional logit (McFadden, 1976) and conditional probit (Hausman & Wise 1978), there need not be the constraint that $\sum_j p_{ij} = 1$ since the sum of the probabilities across stimuli is the expected number of picks for individual i which in many cases exceeds 1. By not requiring this constraint, the model can be utilized to accommodate choice situations, where, say, complementary or multiple-purchased products/brand are bought with correspondingly high probabilities. For example, large families may consistently buy two or more different brands of breakfast cereals on the same purchase occasion. This is why the methodology is more appropriate for pick any/ n data. The independence assumption across individuals and stimuli presupposes that each choice is an independent "coin toss" with given probability p_{ij} . This independence assumption may be questionable for certain applications. For example, in marketing research, one often encounters/designs questionnaires where the order of presentation for products is not randomized across subjects. There can be order effects where the probability of a subject choosing the j -th product presented may be a function of the previous $j - 1$ products presented and the respective choice decisions made. The effects of some violations of this independence assumption are examined in the next section of the paper.

Let us define a latent, unobservable utility variable u_{ij} as:

$$u_{ij} = \sum_{t=1}^T a_{it} b_{jt} + c_i + e_{ij}, \quad (1)$$

where e_{ij} is an error term assumed to have a $N(0, \sigma_{ij}^2)$ distribution where:

$$\text{Cov}(e_{rs}, e_{tu}) = \begin{cases} \sigma_{ij}^2 & \text{if } r = t = i, s = u = j, \\ 0 & \text{else.} \end{cases}$$

The right-hand side of Equation (1) contains the scalar product of the i -th individual's

vector coordinates with the j -th stimulus's coordinates. It is identical to a vector model (Tucker, 1960) of utility where individuals are typically represented by vectors and stimuli by points. The projection of a stimulus onto an individual vector indicates the degree of magnitude of utility—the larger the scalar products (i.e., the higher the projection of a stimulus onto an individual's vector), the higher is the utility of that stimulus for that individual (Slater, 1960).

Here, u_{ij} is specified such that if $u_{ij} \leq c_i^*$ then we observe $y_{ij} = 0$ (no choice), and if $u_{ij} > c_i^*$ we observe $y_{ij} = 1$ (a choice). Thus, c_i^* is a threshold parameter which varies by individual. Therefore:

$$\begin{aligned} P(y_{ij} = 0) &= P(u_{ij} \leq c_i^*) \\ &= P\left(\sum_{t=1}^T a_{it} b_{jt} + c_i + e_{ij} \leq c_i^*\right) \\ &= P\left(e_{ij} \leq -\sum_{t=1}^T a_{it} b_{jt} + \delta_i\right) = \Phi(s), \end{aligned} \quad (2)$$

where $\delta_i = c_i^* - c_i$, and $\Phi(s)$ represents the standard normal distribution function evaluated at:

$$s = \frac{\left(\delta_i - \sum_{t=1}^T a_{it} b_{jt}\right)}{\sigma_{ij}}. \quad (3)$$

Similarly,

$$\begin{aligned} P(y_{ij} = 1) &= P(u_{ij} > c_i^*) \\ &= P\left(e_{ij} > -\sum_{t=1}^T a_{it} b_{jt} + \delta_i\right) \\ &= 1 - \Phi(s) = p_{ij}. \end{aligned} \quad (4)$$

Thus, one can assume that a latent utility variable exists which, after reaching an individual specific threshold value, "produces" the observed choice $y_{ij} = 1$. This general specification is quite common in the econometrics literature (Chow, 1983; Maddala, 1986) where discrete choice models are tied into latent, indirect utility scores and threshold values. The theoretical justification for such a specification can be found in Lewin (1944), Siegel (1957), Simon (1959), and other works on aspiration levels and decision making. According to Simon (1959, 1978), economic agents engage in satisfying behavior rather than maximizing behavior. He claims that economic agents form thresholds or aspiration levels which "defines a natural zero point in the scale of utility (Simon, 1959; p. 264)". When the economic agent has alternatives to it that are at or above its aspiration level, this theory predicts that the agent will choose amongst these alternatives as opposed to those alternatives below this level. This appears to be congruent with multistage decision making/choice processes which combine compensatory and conjunctive rules (Coombs, 1964; Dawes, 1964; Einhorn, 1970; and Green & Wind, 1973).

The likelihood function can be expressed as:

$$L^* = \prod_{i=1}^I \prod_{j=1}^J [1 - \Phi(s)]^{y_{ij}} [\Phi(s)]^{(1-y_{ij})}, \quad (5)$$

and the log likelihood as:

$$K^* = \ln L^* = \sum_{i=1}^I \sum_{j=1}^J [y_{ij} \ln (1 - \Phi(s)) + (1 - y_{ij}) \ln \Phi(s)]. \quad (6)$$

Given $\mathbf{Y} = ((y_{ij}))$ and T , one wishes to estimate $\mathbf{A} = ((a_{it}))$, $\mathbf{B} = ((b_{jt}))$, $\boldsymbol{\delta} = ((\delta_{ij}))$, and σ_{ij} to maximize K^* in (6).

B. Program Options

The spatial model can accommodate external analysis where \mathbf{A} and/or \mathbf{B} are given/specified from, say, some previous analysis, or it can estimate both sets of coordinates (internal analysis). Several options also exist with respect to estimating σ_{ij} and $\boldsymbol{\delta}$. Clearly, one quickly runs out of degrees of freedom if one had to estimate all IJ σ_{ij} parameters, plus \mathbf{A} , \mathbf{B} , and $\boldsymbol{\delta}$. Given the form of (3), estimating a general σ (a constant) or σ_i (by individual) is not appropriate since it can be absorbed in the numerator. Therefore, without much loss of generality, we set $\sigma_{ij} \equiv 1, \forall i, j$, and assume a form of homoscedasticity. Similarly, an option exists to set $\boldsymbol{\delta} = \mathbf{0}$, or $\boldsymbol{\delta}$ can be freely estimated.

Finally, the model defined in (2) and (4) can be generalized to incorporate additional data in the form of individual and/or stimulus background variables. The coordinates for individuals (vector termini) and/or stimuli, as the case might be, can be reparameterized as linear functions of background variables (Bentler & Weeks, 1978; Bloxom, 1978; de Leeuw & Heiser, 1980; and Noma & Johnson, 1977, for constraining MDS spaces). If stimulus attribute data is available, then b_{jt} can be reparameterized as

$$b_{jt} = \sum_{k=1}^K x_{jk} \gamma_{kt}, \quad (7)$$

where x_{jk} is the value of feature k ($k = 1, \dots, K$) for stimulus j and γ_{kt} is the impact of feature k on dimension t . As in CANDELINC (Carroll, Pruzansky, & Kruskal, 1979), Three-Way Multivariate Conjoint Analysis (DeSarbo, Carroll, Lehmann, & O'Shaughnessy, 1982), and GENFOLD2 (DeSarbo & Rao, 1984, 1986), one models the location of stimuli to be a direct function of their respective features. Thus, the x_{jk} are quantified features which are related to subjective attributes (Lancaster, 1966, 1979). Similarly, when individual background data is available, a_{it} can be reparameterized as

$$a_{it} = \sum_{r=1}^R z_{ir} \alpha_{rt}, \quad (8)$$

where z_{ir} is the value of characteristic r ($r = 1, \dots, R$) for individual i and α_{rt} is the impact of the r -th individual characteristic on dimension t . When both stimuli and individual background data are available, both sets of coordinates can be so reparameterized. (Note that one always has the option of performing general property fitting analyses in the non-reparameterized model with \mathbf{A} and \mathbf{B} if, for example, one did not have the "full set" of background variables to describe \mathbf{A} or \mathbf{B} fully.) An option in our methodology also exists to estimate a stretching/shrinking parameter, τ_i , when a_{it} is so reparameterized to avoid potential problems with placing constraints on individual vectors as discussed in Carroll, Pruzansky, and Kruskal (1979) and DeSarbo, et al. (1982). This parameter, τ_i , would appear as a multiplicative term on the right-hand side of (8).

C. The Algorithm

Maximum likelihood methods are utilized to maximize $K^* = \ln L^*$ in (6) with respect to the given set of unknown parameters specified in the particular model of interest. The

method of conjugate gradients (Fletcher & Reeves, 1964) with automatic restarts (Powell, 1977) was utilized for this optimization problem. The specific steps or phases of the algorithm are as follows:

Phase I: Input options. The user must specify T (the number of dimensions for the analysis), IT^* (the maximum number of major iterations allowed), TOL (the convergence tolerance), whether or not $\mathbf{A} = ((a_{it}))$ is reparameterized via $\mathbf{A} = \mathbf{Z}\boldsymbol{\alpha}$, whether or not $\mathbf{B} = ((b_{jt}))$ is reparameterized via $\mathbf{B} = \mathbf{X}\boldsymbol{\gamma}$ and method of obtaining starting estimates. The matrices $\mathbf{Y} = ((y_{ij}))$, $\mathbf{X} = ((x_{jk}))$, and $\mathbf{Z} = ((z_{ir}))$ must also be input, depending upon which model is to be fit.

Phase II: Starting Estimates. The user has the choice of one of three methods to obtain starting estimates of \mathbf{A} (or $\boldsymbol{\alpha}$), \mathbf{B} (or $\boldsymbol{\gamma}$), and $\boldsymbol{\delta} = ((\delta_i))$: either random starting estimates, given starting estimates, or estimates obtained from an MDPREF (Carroll, 1980) metric analysis on \mathbf{Y} (with $\delta_i = 0$). Options exist in the methodology to perform external analyses with \mathbf{A} and/or \mathbf{B} is given and held fixed throughout the analyses. Set $IT = 0$.

Phase III: Estimate \mathbf{A} (or $\boldsymbol{\alpha}$), \mathbf{B} (or $\boldsymbol{\gamma}$), and $\boldsymbol{\delta}$. Set $IT = IT + 1$. This phase of the algorithm estimates \mathbf{A} (or $\boldsymbol{\alpha}$), \mathbf{B} (or $\boldsymbol{\gamma}$), and $\boldsymbol{\delta}$. Estimates of these parameters are sought to maximize the ln of the likelihood function in (6). The method of conjugate gradients (Fletcher & Reeves, 1964) is utilized to solve this unconstrained optimization procedure. Assuming the ln of the likelihood function in (6) obeys typical continuity and regularity conditions, one can take partial derivatives of expression (6) with respect to the desired set of parameters:

$$\frac{\partial K^*}{\partial a_{it}} = \sum_j \frac{\phi(s)b_{jt}}{\sigma_{ij}} \left(\frac{y_{ij}}{1 - \Phi(s)} - \frac{(1 - y_{ij})}{\Phi(s)} \right), \quad (9)$$

$$\frac{\partial K^*}{\partial b_{jt}} = \sum_i \frac{\phi(s)a_{it}}{\sigma_{ij}} \left(\frac{y_{ij}}{1 - \Phi(s)} - \frac{(1 - y_{ij})}{\Phi(s)} \right), \quad (10)$$

$$\frac{\partial K^*}{\partial \delta_i} = \sum_j \frac{\phi(s)}{\sigma_{ij}} \left(\frac{(1 - y_{ij})}{\Phi(s)} - \frac{y_{ij}}{1 - \Phi(s)} \right), \quad (11)$$

$$\frac{\partial K^*}{\partial \gamma_{kr}} = \sum_i \sum_j \frac{\phi(s)a_{it}x_{jk}}{\sigma_{ij}} \left(\frac{y_{ij}}{1 - \Phi(s)} - \frac{(1 - y_{ij})}{\Phi(s)} \right), \quad (12)$$

$$\frac{\partial K^*}{\partial \alpha_{rt}} = \sum_i \sum_j \frac{\tau_i \phi(s)b_{jt}z_{ir}}{\sigma_{ij}} \left(\frac{y_{ij}}{1 - \Phi(s)} - \frac{(1 - y_{ij})}{\Phi(s)} \right), \quad (13)$$

$$\frac{\partial K^*}{\partial \tau_i} = \sum_j \frac{\phi(s)b_{jt}a_{it}}{\sigma_{ij}\tau_i} \left(\frac{y_{ij}}{1 - \Phi(s)} - \frac{(1 - y_{ij})}{\Phi(s)} \right), \quad (14)$$

where $\phi(s)$ is the evaluation of the standard normal density function at s defined in (3). For external analyses, or analyses where $\tau_i = 1$, the relevant set of partial derivatives are set equal to zero.

For sake of convenience, let's assume that the relevant parameters to be estimated are contained in the vector $\boldsymbol{\theta}$ and that ∇K is the vector of relevant partial derivatives for this desired set of parameters. Then the complete conjugate gradient procedure utilized

can be summarized as follows:

- (i) Start with initial parameter estimate $\theta^{(1)}$; set $IT = 1$.
- (ii) Set the first search direction $S^{(1)} = \nabla K^{(1)}$.
- (iii) Find $\theta^{(2)}$ via:

$$\theta^{(2)} = \theta^{(1)} + \lambda^{(1)}S^{(1)}. \quad (15)$$

where $\lambda^{(1)}$ is the optimal step length (obtained through quadratic interpolation) in the direction $S^{(1)}$. Set $IT = 2$.

- (iv) Calculate $\nabla K^{(IT)}$ and set

$$S^{(IT)} = \nabla K^{(IT)} + \frac{(\nabla K^{(IT)})(\nabla K^{(IT)})}{(\nabla K^{(IT-1)})(\nabla K^{(IT-1)})} S^{(IT-1)}. \quad (16)$$

- (v) Compute the optimal step length $\lambda^{(IT)}$ in the direction $S^{(IT)}$, and find:

$$\theta^{(IT+1)} = \theta^{(IT)} + \lambda^{(IT)}S^{(IT)}. \quad (17)$$

- (vi) If $\theta^{(IT+1)}$ is optimal and/or satisfies the iteration tests below, stop. Otherwise set $IT = IT + 1$ and go to step (iv) above (i.e., undertake another iteration). A number of convergence tests are performed in this estimation cycle to test whether additional iterations of the conjugate gradient are required:

$$\begin{aligned} &\text{---If } IT \geq IT^*, \text{ stop;} \\ &\text{---If } \|\nabla K^{(IT)}\| \leq \underline{\text{TOL}}, \text{ stop;} \\ &\text{---If } (K^{*(IT)} - K^{*(IT-1)}) \leq \text{TOL}, \text{ stop.} \end{aligned} \quad (18)$$

It has been shown that conjugate gradient procedures can avoid the typical "cycling" often encountered with steepest descent algorithms. In addition, they demonstrate valuable quadratic termination (Himmelblau, 1972) properties, that is, conjugate gradient procedures will typically find the globally optimum solution for a quadratic loss function in H steps, where H is the number of parameters to solve for. This is relevant for our application only if a Taylor expansion of $\ln L^*$ around the optimum is dominated by its second order term.

This conjugate gradient method is particularly useful for optimizing functions of several parameters since it does not require the storage of any matrices as is necessary in Quasi-Newton and second derivative methods. However, as noted by Powell (1977), the rate of convergence is linear only if this iterative procedure is "restarted" occasionally, that is, cycles back through previous search directions. Restarts have been implemented in this algorithm automatically depending upon successive improvement in the objective function. Also note that since K^* in (6) has an upper bound of zero, and since each estimating phase (iteration) of the likelihood maximization can be shown to conditionally increase K^* , one can use a limiting sum argument (Courant, 1965) to prove convergence to at least a locally optimum solution.

Phase IV: Output. Several related goodness-of-fit measures are computed for the spatial representation:

1. The \ln likelihood function: $K^*(A, B, \delta)$;
2. The Deviance Measure (Nelder & Wedderburn, 1972; McCullagh & Nelder, 1983):

$$D = -2 \left[\sum_{i=1}^I \sum_{j=1}^J y_{ij} \ln(\hat{p}_{ij}) + (1 - y_{ij}) \ln(1 - \hat{p}_{ij}) \right] = -2K^*, \quad (19)$$

where \hat{p}_{ij} , the estimated probability of $y_{ij} = 1$, is expressed as in (4) using estimated values for a_{it} and b_{jt} . Note, one can (theoretically) asymptotically test nested models as the difference between respective deviance measures which is χ^2 distributed with the difference in model degrees of freedom as the appropriate χ^2 test degrees of freedom. This is appropriate in testing dimensionality and model structure because of the obvious nesting. However, one question that arises with the use of maximum likelihood procedures concerns the validity of the statistical properties of estimators and associated tests. This question is of paramount importance with respect to the general choice model presented here due to the presence of incidental parameters (e.g., \mathbf{A}_i , \mathbf{B}_j , δ_j) whose number varies according to the order of \mathbf{Y} . According to Andersen (1980), maximum likelihood estimators in such cases may not be consistent. Andersen suggests the use of conditional maximum likelihood estimators in such cases. Takane (1983) also reports of such difficulties in his unfolding-type item response model. DeSarbo and Hoffman (1986) cite similar difficulties with an unfolding version of this choice model. We will present a small Monte Carlo analysis which will attempt to shed some light on this problem. Clearly, massive simulations are needed to test this asymptotic property properly.

3. The Akaike information criterion (AIC) statistic. The AIC statistic (Akaike, 1974) can also be utilized as a method of model selection. Here it is defined as:

$$\text{AIC} = D + 2n, \quad (20)$$

where n is the effective number of independent parameters estimated in the model and D is the deviance measure defined in (19). Typically, one selects the model/solution with the minimum value of AIC. However, as pointed out by Bozdogan (1987), one potential problem with the use of this model selection criterion is that the AIC statistic is "dimension inconsistent" in that it typically over estimates the true dimensionality in such spatial models. Thus, caution must also be given here with use of such a test statistic.

4. A sum of squares statistic:

$$\text{SSQ} = \sum_{i=1}^I \sum_{j=1}^J [y_{ij} - \hat{p}_{ij}]^2. \quad (21)$$

5. The point biserial correlation between \mathbf{Y} and $\hat{\mathbf{P}}$: $r_{pb}(\mathbf{Y}, \hat{\mathbf{P}})$, where $\hat{\mathbf{P}} = ((\hat{p}_{ij}))$.

6. The proportion of correct predictions via the simple matching coefficient between \mathbf{Y} and $\hat{\mathbf{Y}}$.

7. The phi coefficient calculated between \mathbf{Y} and $\hat{\mathbf{Y}}$.

Note that the last three related measures are also given for row and column elements to examine goodness of fit for each individual and stimulus.

The appropriateness of the dimensionality of a solution or the nesting of models can be evaluated by a simultaneous examination of *all* these measures. Theoretically, as previously stated, the difference between deviance measures is asymptotically distributed as χ^2 with degrees of freedom equal to the difference between the corresponding degrees of freedom of the two dimensionalities or the nested models. Yet with the presence of incidental parameters (except with complete reparameterization with $\delta = \mathbf{0}$), it is also a good idea to always examine the other various measures. This central χ^2 distribution asymptotically holds (theoretically) when the "restricted" model is the true model (vs. the "full" model). In practice, the user rarely knows what the "true model" really is, especially concerning the appropriate dimensionality. In fact, one can argue for the use of a non-central χ^2 if the full model is the true model and it is compared against a restricted model. For all these reasons, little credence is placed in such a test per se and a strong recommendation is made to examine *all* of these measures simultaneously.

III. Monte Carlo Analyses

A. *The Independence Assumption*

The relationship between a set of choice data and its spatial representation can be illustrated by a small synthetic data set. Such a data set was designed so that all possible choice patterns for $J = 3$ stimuli (except for the two patterns “choosing every stimulus” and “choosing no stimuli”) would occur. The number of stimuli was purposely limited with the expectation that the solution space would be of sufficiently low dimensionality to facilitate visual examination and interpretation. Several choice patterns were repeated in order to check the robustness of the methodology to violations of row and column independence.

Synthetic data were constructed for $I = 12$ individuals evaluating $J = 6$ stimuli by repeating the patterns:

		Stimulus		
		1	2	3
	A	1	0	0
	B	0	1	0
Pattern:	C	0	0	1 = M
	D	1	1	0
	E	1	0	1
	F	0	1	1

four times to create a 12×6 matrix \mathbf{Y} :

$$\mathbf{Y} = \begin{bmatrix} \mathbf{M} & \mathbf{M} \\ \mathbf{M} & \mathbf{M} \end{bmatrix}.$$

Each (hypothetical) individual's choices can be represented by one of these 6 patterns, where a “1” indicates that a stimulus was chosen by that individual and a “0” that it was not chosen. It was desired to examine the performance of the algorithm in: determining the dimensionality of the space, producing a reasonable spatial representation of the structure in the data, examining whether individuals with the same choice pattern had the same coordinates (local optimum solutions are possible with such an algorithm which might render different coordinates for individuals with the same choice values), examining whether stimuli with the same choice patterns had the same coordinates, and investigating what happens when the independence assumption is seemingly violated across *both* i and j . Table 1 presents the statistical summary for the general model without reparameterization for $T = 1$, $T = 2$, and $T = 3$. Note that the deviance measure for going from $T = 1$ to $T = 2$ is significant, but the deviance measure for going from $T = 2$ to $T = 3$ is not, indicating that two dimensions appears to be the appropriate dimensionality. The AIC statistics also confirm this result. Even if the asymptotic χ^2 test were completely invalid, the other measures reported also clearly verify that two dimensions are the appropriate representation for such data.

Figure 1 portrays the rotated joint space for such data (see Kruskal, 1981, for the indeterminacies of such two-way bilinear models). Individual vectors are labeled by capital letters ($A-L$) and stimulus points by numbers (1–6). The configuration in Figure 1 appears to be a parsimonious spatial representation of the underlying structure in this

TABLE 1

Summary of the analysis for the simulated data

T	d.f.	$K^* = \ln L^*$	Deviance	SSQ	rpb	ϕ	Matching Coefficient	Diff. D	A.I.C.
1	29	-16.33	32.66	5.43	0.83	.78	.89	-	90.66
2	44	-0.00	.00	0.00	1.00	1.00	1.00	32.65**	88.00 ⁺
3	57	-0.00	.00	0.00	1.00	1.00	1.00	0.002NS	114.00

** $p < .01$
⁺ min A.I.C.

synthetic choice data set. Each pattern can be identified by a cluster of the individuals who indeed have that choice pattern [(A, G), (B, H), (C, I), (D, J), (E, K), (F, L)]. Individual vectors with the same pattern have grouped very tightly together in the space. Similarly, stimuli with the same patterns [(1, 4), (2, 5), and (3, 6)] are located close to one another. The positioning of the stimuli relative to the vector groups clearly reflects the choices made by an individual. Note how nicely the vectors are aligned so that stimuli 1(4), 2(5), or 3(6) project highest on those individuals that solely chose 1(4), 2(5), or 3(6), respectively. Those individuals that chose multiple stimuli in Y (stimuli with truly different patterns) are represented by vectors in between the stimuli they chose. For example, those individuals (F, L) that chose stimuli 2(5) and 3(6) have vectors between stimuli 2(5) and 3(6) indicating high positive projections of 2(5) and 3(6) on their choice vector and are reflections of those individuals (A, G) that did not choose 2(5) and 3(6), but chose solely 1(4). Thus, individual choices can be assessed quickly and comprehensively.

The analysis for the synthetic data was also performed using a different random start for initial estimates. The goal was to examine the stability of the resulting solutions obtained from the algorithm and to check for potential locally optimum solutions. Canonical correlation was used as an approximate configuration matching technique to examine the similarity of the two solutions. Substantial congruence was found between the two solutions with canonical correlations of .997 and .996 for A, .999 and .999 for B, and .999 and .991 for the joint space.

Other analyses with versions of this data set involving the introduction of other forms of dependency such as the dropping of certain choice patterns, having different number of replications for each choice pattern, adding complementary (e.g., if brand A is chosen, so is brand B) and substitute (e.g., if brand A(B) is chosen, B(A) is never chosen) stimuli, etcetera, were also performed. In every case, the methodology produced a parsimonious spatial representation which accurately described the respective choice patterns with similar results as presented in Table 1. While not every form of row/column dependency was tested, it does appear that this methodology is somewhat robust to these obvious violations to its independence assumptions.

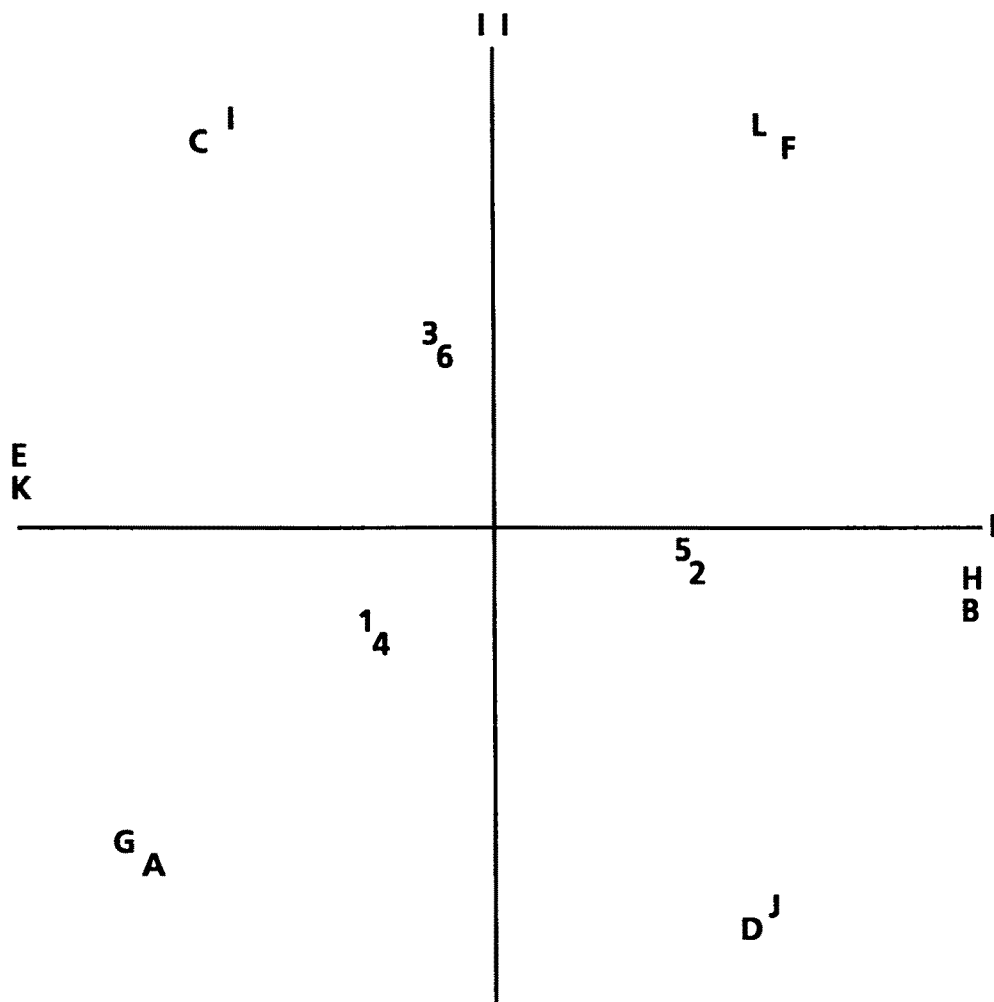


FIGURE 1
Rotated Two-Dimensional Joint Space Plot for the Synthetic Data Analysis.

B. Algorithm Performance

In order to examine the performance of the methodology more rigorously, a Monte Carlo analysis was performed where some eight independent factors were experimentally manipulated to create synthetic data. Table 2 presents the eight factors and levels investigated in this modest Monte Carlo analysis. A $2^5 3^3$ factorial design (Addelman, 1962) was utilized to generate appropriate combinations of these levels of factors allowing for main-effects-only estimation (see DeSarbo, 1982; and DeSarbo & Carroll, 1985, for similar types of Monte Carlo analyses). While this is clearly a limitation, it was done to reduce the number of runs from a total of 864 (complete enumeration of all possible combinations) to 27. Table 3 lists the specific $2^5 3^3$ asymmetric fractional factorial design utilized here in this analysis.

Three general areas of methodological performance were measured: (a) the amount of computer usage required; (b) a comparison between the actual binary data in Y and the model predicted values; and, (c) the recovery of the actual parameters of the model.

TABLE 2

Independent factors for Monte Carlo analysis

<u>Factor</u>	<u>Levels</u>	<u>Code</u>
A. T	2	0
	3	1
B. I	15	0
	30	1
C. J	7	0
	12	1
D. Threshold Parameters	$\delta_i=0$	0
	Estimate δ_i	1
E. Starting Procedure	Random	0
	SVD of \underline{Y}	1
F. Reparameterize B	No	0
	K=2	1
	K=4	2
G. Reparameterize A	No	0
	L=3	1
	L=5	2
H. Error Distribution	N(0,1)	0
	Type I Extreme Value	1
	Cauchy	2

Computer usage for each trial run was operationalized in terms of the number of iterations required for convergence (CPU time was not available). Overall goodness of fit was measured for each trial via $\ln L^*$, the point biserial correlation coefficient between \underline{Y} and $\hat{\underline{P}}$, the phi coefficient between \underline{Y} and $\hat{\underline{Y}}$, the sums of squares statistic between \underline{Y} and $\hat{\underline{P}}$, and the simple matching coefficient between \underline{Y} and $\hat{\underline{Y}}$. As previously discussed, these measures are the standard goodness of fit indices that are output with this methodology. Finally, parameter recovery for each trial was assessed by comparing \underline{A} vs. $\hat{\underline{A}}$, \underline{B} vs. $\hat{\underline{B}}$, and the actual versus predicted joint space $\underline{N}' = (\underline{A}, \underline{B})'$ vs. $\hat{\underline{N}}' = (\hat{\underline{A}}, \hat{\underline{B}})'$. Basically each trial or row in the fractional factorial design shown in Table 3 stipulated how each synthetic data set was to be constructed and analyzed. Dimensionality, the number of individuals and stimuli, whether the threshold parameters were to be fixed at zero or whether they were to be estimated, the type of start (random vs. a singular value decomposition of \underline{Y} (MDPREF; Carroll, 1980)), whether to estimate \underline{B} or $\underline{\gamma}$, whether to estimate \underline{A} or $\underline{\alpha}$, and the type of error assumed (normal, extreme value, or Cauchy), were all varied.

Table 4 presents the regression results for three of these various dependent measures for the main-effects only design. For sake of brevity, one specific measure was selected from each of the three groups presented above. For goodness of fit, the phi coefficient was selected for presentation since these results were quite similar to the point biserial and simple matching coefficient results ($\ln L$ and the sums of squares statistics are excessively influenced by the size of \underline{Y} , i.e., I and J). Similarly, the results presented for the joint space

TABLE 3

 $2^5 3^3$ Asymmetric fractional factorial Monte Carlo design

<u>Trial</u>	<u>Factor:</u>							
	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>	<u>E</u>	<u>F</u>	<u>G</u>	<u>H</u>
1	0	0	0	0	0	0	0	0
2	0	0	0	0	1	2	1	2
3	0	0	0	0	0	1	2	1
4	0	1	1	0	0	1	2	2
5	0	1	1	0	1	0	0	1
6	0	1	1	0	0	2	1	0
7	0	0	0	1	0	2	1	1
8	0	0	0	1	1	1	2	0
9	0	0	0	1	0	0	0	2
10	1	0	1	1	0	1	1	1
11	1	0	1	1	1	0	2	0
12	1	0	1	1	0	2	0	2
13	1	1	0	0	0	2	0	0
14	1	1	0	0	1	1	1	2
15	1	1	0	0	0	0	2	1
16	1	0	0	0	0	0	2	2
17	1	0	0	0	1	2	0	1
18	1	0	0	0	0	1	1	0
19	0	0	0	0	0	2	2	2
20	0	0	0	0	1	1	0	1
21	0	0	0	0	0	0	1	0
22	0	1	0	1	0	0	1	1
23	0	1	0	1	1	2	2	0
24	0	1	0	1	0	1	0	2
25	0	0	1	0	0	1	0	0
26	0	0	1	0	1	0	1	2
27	0	0	1	0	0	2	2	1

recovery (the average canonical correlation between \mathbf{N} and $\hat{\mathbf{N}}$) were quite similar to those for \mathbf{A} and \mathbf{B} recovery. (Given that phi and canonical correlations vary between 0 and 1, arc sin and logit transformations were applied to each and the corresponding regressions were also performed on them. The results were not very different than those presented in Table 4 and thus are not discussed.)

From an inspection of this table, it appears that the number of iterations appears to be affected significantly by the number of parameters to be estimated. Here, setting $T = 3$ and estimating the threshold coefficients significantly increase the number of iterations required for convergence. The nearly significant and positive coefficients for $I = 30$, $J = 12$, $K = 4$, and $L = 5$ also support this conclusion. Oddly, the Cauchy error level appears to enhance the speed of convergence (perhaps leading to "quicker" locally optimal solutions). This regression is highly significant as shown by the corresponding F -test.

Concerning the overall goodness of fit as measured by the phi coefficient, better recovery appears to be associated with higher dimensions. The SVD start appears to detract significantly from this goodness of fit statistic suggesting that its use could possibly lead to locally optimal solutions. The most significant variable is the Cauchy error level

TABLE 4

Monte Carlo regression results

<u>Independent Variable</u>	<u>Number of Iterations</u>	<u>Phi Coefficient</u>	<u>Joint Space Recovery</u>
Intercept	18.20	0.90	0.81
T=3	5.89*	0.09*	-0.02
I=30	5.22	-0.02	0.03
J=12	3.56	-0.06	-.03
Estimate δ_i	9.22*	0.06	-0.04
SVD start	-2.61	-0.07*	-0.11*
K=2	-6.00	-0.06	0.02
K=4	3.22	-.00	0.07
L=3	-1.67	-0.03	0.09
L=5	4.89	-0.02	0.05
Extreme Error	-2.22	-0.01	0.09
Cauchy Error	-7.22*	-0.13**	-0.08
S.E	6.59	0.08	0.10
R ²	0.74	0.73	0.64
adj R ²	0.55	0.53	0.37
F	3.87**	3.64*	2.40

* $p \leq .05$
 ** $p < .01$

which significantly detracts from the phi coefficient value as expected since the Cauchy distribution has undefined moments (moments ≥ 1 are infinite). This regression equation is significant at the .05 level.

Finally, concerning the recovery of the joint space, the only significant variable is the SVD start which appears to *detract* significantly from parameter recovery. This was also the case in the analyses performed for A and B separately (not shown) indicating that this starting method might lead to locally optimum solutions. Note also that, while not significant, the Cauchy error variable also possesses a hefty negative coefficient indicating that it has some negative effect on subsequent parameter recovery. However, this regression equation is not significant at the .05 or .01 levels.

The results of this preliminary Monte Carlo analysis appear to demonstrate the

somewhat consistent performance of the methodology. Larger data sets (and more parameters to estimate) will naturally require additional computer demands. Severe misspecification with a distribution with undefined moments should detract from data and parameter recovery. Perhaps the most surprising result concerned the results of using a SVD analysis of Y (MDPREF—a metric vector model; Carroll, 1980) as a possible rational starting method and finding that it can lead to poorer results. Note the consistent fitting over all different model types involving the various reparameterization options. These results should only be considered as preliminary due to the small scope of this reported Monte Carlo analysis. Clearly, a full factorial design with replications would have been preferred had computer expense been unlimited. Also, additional factors and perhaps additional levels per factor might also have been desirable. These are left as avenues for possible future research.

IV. An Application to Soft Drinks

A. Study Description

A group of $N = 50$ male and female MBA students from Columbia University were asked to indicate, for a variety of popular soft drinks, which soft drinks they purchased and consumed within a given time period. The collected data were coded as 1 to indicate purchase and consumption at least every other week and 0 to indicate purchase and consumption less than every other week. The data for five of the 50 individuals were discarded since they claimed not to drink any soft drinks. Eleven brands of soft drinks were initially tested (see Hoffman & Franke, 1986):

	<u>Plot Code</u>
1. Coke®	<i>A</i>
2. Diet Coke®	<i>B</i>
3. Diet Pepsi®	<i>C</i>
4. Diet 7UP®	<i>D</i>
5. Dr. Pepper®	<i>E</i>
6. Mountain Dew®	<i>F</i>
7. Pepsi®	<i>G</i>
8. Pepsi Light®	<i>H</i>
9. Sprite®	<i>I</i>
10. Tab®	<i>J</i>
11. 7UP®	<i>K</i>

The complete 45×11 data set is presented in Table 5. The associated row and column marginals are also reported. This particular data set appears well suited for analysis via the spatial methodology since most students did consume multiple brands of soft drinks during the time period studied (e.g., only Subject 1 in Table 5 has a row sum of one). One can use arguments of variety seeking and/or situational differences in consumption to justify multiple brand consumption and the independence assumption over brands.

In addition to these choice data, background information about the individuals and brands were also collected. Concerning the individuals, specific demographic and psychographic responses were collected and initially examined for individual differences. Unfortunately, given the fact that these individuals were all students, these responses (outside of gender) were rather homogeneous and not very valuable in explaining the large individual differences in soft drink choice behavior witnessed in Y in Table 5. However, given the

TABLE 5

Soft drink choice data by 45 students
for 11 brands

<u>Subject:</u>	<u>Soft Drink:</u>											<u>Row Totals</u>
	<u>A</u>	<u>B</u>	<u>C</u>	<u>D</u>	<u>E</u>	<u>F</u>	<u>G</u>	<u>H</u>	<u>I</u>	<u>J</u>	<u>K</u>	
1	1	0	0	0	0	0	0	0	0	0	0	1
2	1	0	0	0	0	0	1	0	0	0	0	2
3	1	0	0	0	1	1	1	1	1	0	1	7
4	1	0	0	0	0	0	1	0	0	0	0	2
5	1	0	0	0	0	0	1	0	0	0	0	2
6	1	0	0	0	0	0	1	0	0	0	0	2
7	0	1	0	1	0	0	0	0	1	1	1	5
8	1	0	0	0	0	0	1	0	1	0	1	4
9	1	0	0	0	1	0	1	0	1	0	1	5
10	0	1	1	1	0	0	0	1	0	1	0	5
11	1	1	1	0	0	0	1	1	1	0	1	7
12	1	1	0	0	0	0	1	0	1	1	1	6
13	1	0	0	0	0	0	1	0	1	0	1	4
14	1	0	0	0	0	1	1	1	1	0	1	6
15	1	0	0	0	0	0	1	0	1	0	1	4
16	0	1	0	1	0	0	0	0	1	0	0	3
17	0	1	1	1	1	0	0	1	0	1	0	6
18	0	1	1	1	0	0	0	0	1	0	1	5
19	1	0	0	0	1	0	0	0	1	0	0	3
20	0	1	1	1	0	0	0	1	0	1	0	5
21	1	0	0	0	1	1	1	0	0	0	1	5
22	0	1	0	0	0	0	0	0	0	1	0	2
23	0	0	0	0	1	1	1	0	1	0	1	5
24	0	1	0	0	1	1	0	0	1	1	1	6
25	1	1	1	1	0	0	1	0	1	0	0	6
26	0	0	0	0	1	0	1	0	0	0	0	2
27	1	0	0	0	0	0	0	0	1	0	1	3
28	1	1	1	0	0	0	1	0	0	0	1	5
29	1	1	0	0	0	0	1	0	0	0	1	4
30	1	0	0	0	0	0	1	0	0	0	1	3
31	0	1	0	1	0	0	0	1	0	1	0	4
32	1	1	0	0	0	0	1	0	1	0	1	5
33	1	1	1	1	0	0	1	1	0	1	0	7
34	0	1	0	1	0	0	0	0	0	1	0	3
35	1	1	1	1	0	0	0	0	0	0	1	5
36	0	1	0	0	0	0	0	0	0	1	0	2
37	0	1	0	0	0	0	0	0	0	1	0	2
38	1	0	0	0	0	1	1	0	1	0	1	5
39	1	0	0	0	0	0	1	0	1	0	0	3
40	0	1	1	1	0	0	0	1	0	1	0	5
41	1	0	0	0	0	0	1	0	1	0	1	4
42	0	1	1	1	0	0	0	0	0	1	0	4
43	1	1	0	0	1	1	1	0	0	0	1	6
44	0	1	1	1	0	0	0	1	0	1	0	5
45	0	0	0	0	0	1	1	0	1	0	1	4
<u>Column Totals</u>	27	24	12	14	9	8	26	10	21	15	23	

TABLE 6

X Matrix of brand features

<u>Brand:</u>	<u>Feature:</u>						
	X_1	X_2	X_3	X_4	X_5	X_6	X_7
A	1	0	0	1	0	21.8	1
B	1	0	1	1	0	5.1	1
C	1	0	1	1	1	2.9	1
D	0	1	1	0	1	1.4	0
E	1	0	0	1	0	5.1	0
F	0	1	0	1	1	2.7	0
G	1	0	0	1	1	17.0	1
H	1	1	1	1	1	0.2	0
I	0	1	0	0	0	3.3	0
J	1	0	1	1	0	1.6	0
K	0	1	0	0	1	5.2	0

Coding:

$$X_1 = \begin{cases} 1 & \text{if cola,} \\ 0 & \text{else;} \end{cases}$$

$$X_5 = \begin{cases} 0 & \text{if Coca Cola brand,} \\ 1 & \text{if Pepsi brand;} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{if fruity tasting,} \\ 0 & \text{else;} \end{cases}$$

$$X_6 = \begin{cases} 1985 & \text{Market Share,} \end{cases}$$

$$X_3 = \begin{cases} 1 & \text{if diet,} \\ 0 & \text{else;} \end{cases}$$

$$X_7 = \begin{cases} 1 & \text{if Low Sodium,} \\ 0 & \text{else.} \end{cases}$$

$$X_4 = \begin{cases} 1 & \text{if contains caffeine,} \\ 0 & \text{else;} \end{cases}$$

previous marketing literature on the market structure of soft drinks (Hoffman & Franke, 1986; Lehmann 1985) where similar brands of soft drinks appeared to be perceptually differentiated on the basis of Diet-NonDiet and Cola-NonCola characteristics, a design matrix of some seven brand features was utilized to reparameterize **B**. Table 6 presents the seven features/attributes most often mentioned by students (in a pretest) that were most relevant to them in the selection of a soft drink. Price was not included in this Table since most 12 oz cans are comparably priced, especially in vending machines. While small price

differences exists for larger volume containers, often times store promotions discount certain brands during various times of the year. Note, Market Share here is a proxy for the students' mention (in the pretest) of "popularity".

B. Other Analyses

A correspondence analysis was performed on the "undoubled" data in Table 5. Three or four dimensions appear to adequately account for the structure in the choice data with proportions of inertia of .143, .159, .100, and .098 respectively. Unfortunately, only the first dimension was clearly interpretable (Diet-NonDiet) as ascertained by a supplemental points analysis with X (in Table 6) on the resulting space and visual interpretation.

Treating the data as metric, an MDPREF (Carroll, 1980) analysis was also conducted on Y in Table 5 given that this is a type of vector model. Based on goodness of fit, two or three dimensions appear to parsimoniously describe the structure in the data. The variance accounted for statistics were .499, .238, and .069 respectively for each dimension. Utilizing correlation as a type of property fitting procedure between X in Table 6 and the resulting brand coordinates, the first dimension was most highly correlated with Market Share ($r = -.731$), while the second was clearly related to Diet-NonDiet ($r = .961$). The third is a bit harder to interpret but appeared to be related to sodium content ($r = -.771$).

C. The Vector Threshold Methodology

The spatial vector model of choice was estimated in $T = 1, 2, 3,$ and 4 dimensions with $\sigma_{ij} = 1$, $\tau_i = 1$, and the reparameterization option: $B = X\gamma$. The columns of X were initially standardized to zero mean and unit variance in order to aid in the interpretation of the resulting γ . As Table 7 shows, the $T = 3$ dimensional solution appears most appropriate as delineated by the asymptotic χ^2 test and AIC statistics. Even if these tests were completely invalid, the accompanying goodness of fit measures also provide substantial evidence of the appropriateness of the $T = 3$ dimensional solution.

TABLE 7

Analyses of soft drink choice data

<u>T</u>	<u>df</u>	<u>$K^* = \ln L^*$</u>	<u>Deviance</u>	<u>SS</u>	<u>r_{pb}</u>	<u>ϕ</u>	<u>Matching Coefficient</u>	<u>Deviance Difference</u>	<u>A.I.C.</u>
1	96	-254.18	508.36	86.77	.55	.49	.75	—	700.36
2	145	-187.69	375.38	59.34	.74	.69	.84	132.98**	665.38
3	192	-56.75	113.50	17.47	.92	.90	.95	261.88**	487.50 ⁺
4	228	-49.37	98.74	14.76	.94	.92	.97	14.76NS	554.74

** p < .01
⁺ min A.I.C.

Table 8 presents the correlation between the rotated solution brand coordinates (**B**) and the design matrix **X** presented in Table 6. As this table depicts, the first brand dimension is dominated by Market Share (and perhaps sodium), the second by Diet-NonDiet, and the third by Cola-NonCola, caffeine (which tends to be associated with cola beverages), and perhaps fruity-nonfruity (which is usually negatively correlated with cola

TABLE 8
Correlations between \underline{B} and \underline{X} and threshold parameters

<u>Brand Feature:</u>		<u>Correlations</u>		
		<u>I</u>	<u>II</u>	<u>III</u>
X_1	Cola	-.201	-.169	.948
X_2	Fruity Tasting	.394	.060	-.761
X_3	Diet	.392	-.958	.140
X_4	Caffeine	-.072	.015	.868
X_5	Manufacturer	.149	-.019	-.315
X_6	Market Share	-.953	.439	.250
X_7	Sodium	-.691	-.094	.392

<u>Individual</u>		<u>Estimated</u>		<u>Individual</u>		<u>Threshold Parameters</u>	
		<u>Individual</u>	<u>δ_i</u>			<u>Individual</u>	<u>δ_i</u>
1	3.112	16	1.415	31	1.697		
2	1.894	17	.595	32	-1.594		
3	-1.549	18	.384	33	-1.306		
4	1.894	19	1.276	34	2.000		
5	1.894	20	.056	35	.199		
6	1.894	21	.357	36	2.353		
7	.412	22	2.353	37	2.353		
8	.334	23	.340	38	.595		
9	-.170	24	-.192	39	1.371		
10	.056	25	-.585	40	.056		
11	-1.540	26	2.522	41	.334		
12	-1.443	27	2.578	42	1.549		
13	.334	28	-2.293	43	1.506		
14	.216	29	-1.217	44	.506		
15	.334	30	.535	45	1.579		

drinks). Thus, according to these loadings, we can tentatively label Dimension I as a "Popularity" dimension, Dimension II as a "Diet-NonDiet" dimension, and Dimension III as a "Cola-NonCola" dimension. As mentioned previously, these dimensions have also been uncovered in various perceptual studies performed in this market.

Figures 2-3 depict the two dimensional joint spaces for this $T = 3$ dimensional solution. Here, the eleven brands are represented by the letters A-K, while individuals' vectors have been normalized to a constant length for convenience and their termini are represented as "o's". Note, because of several common locations or "overstrikes," we cannot singularly represent all 45 vectors in these plots. The γ coefficients for the seven soft drink feature variables are designated by (normalized) vector termini numbered 1-7. By examining these γ coefficient plots, one can confirm the interpretation of the three dimensions as given by the loadings in Table 8. The figures clearly show the obvious preference of these individuals for those more popular brands with higher market share such as Coke®, Pepsi®, 7UP®, and Sprite®. However, preference for Diet-NonDiet or

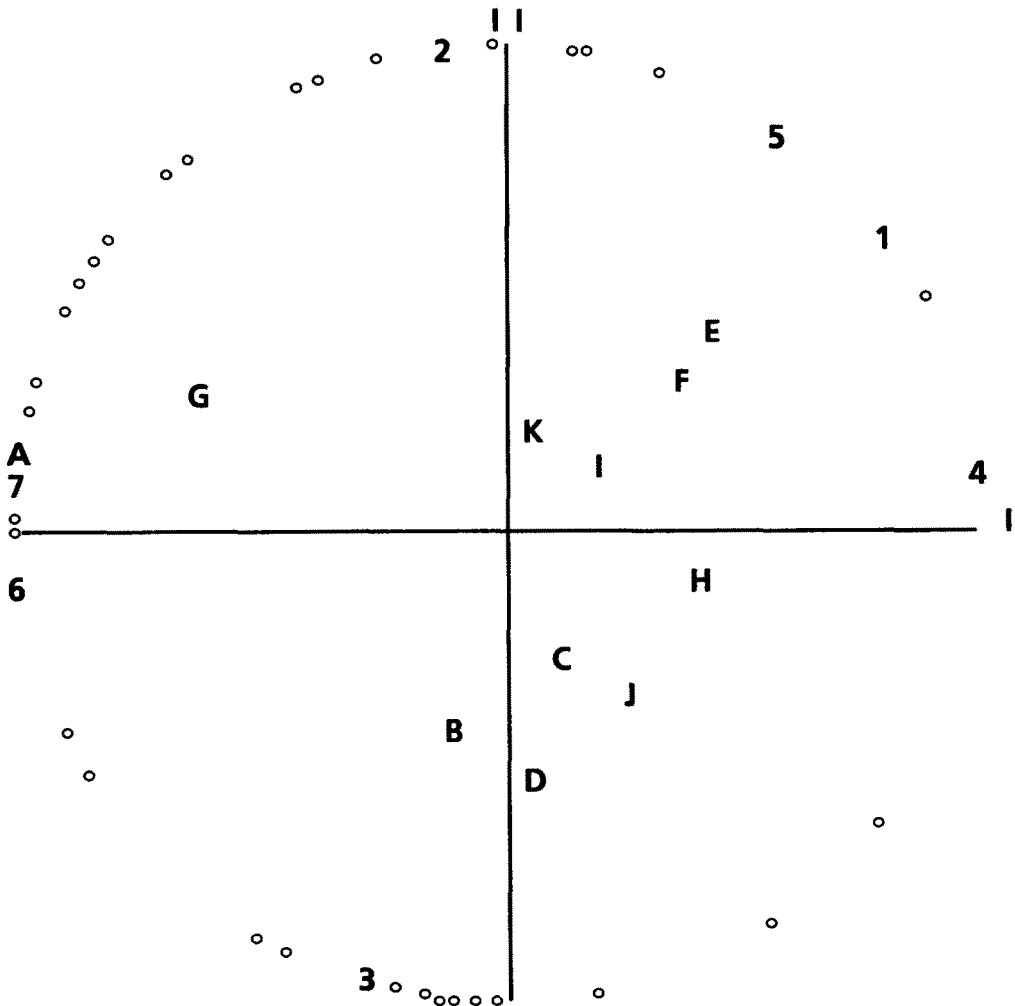


FIGURE 2
Joint Space Plot of Dimensions 1 vs 2 for the Soft Drink Analysis.

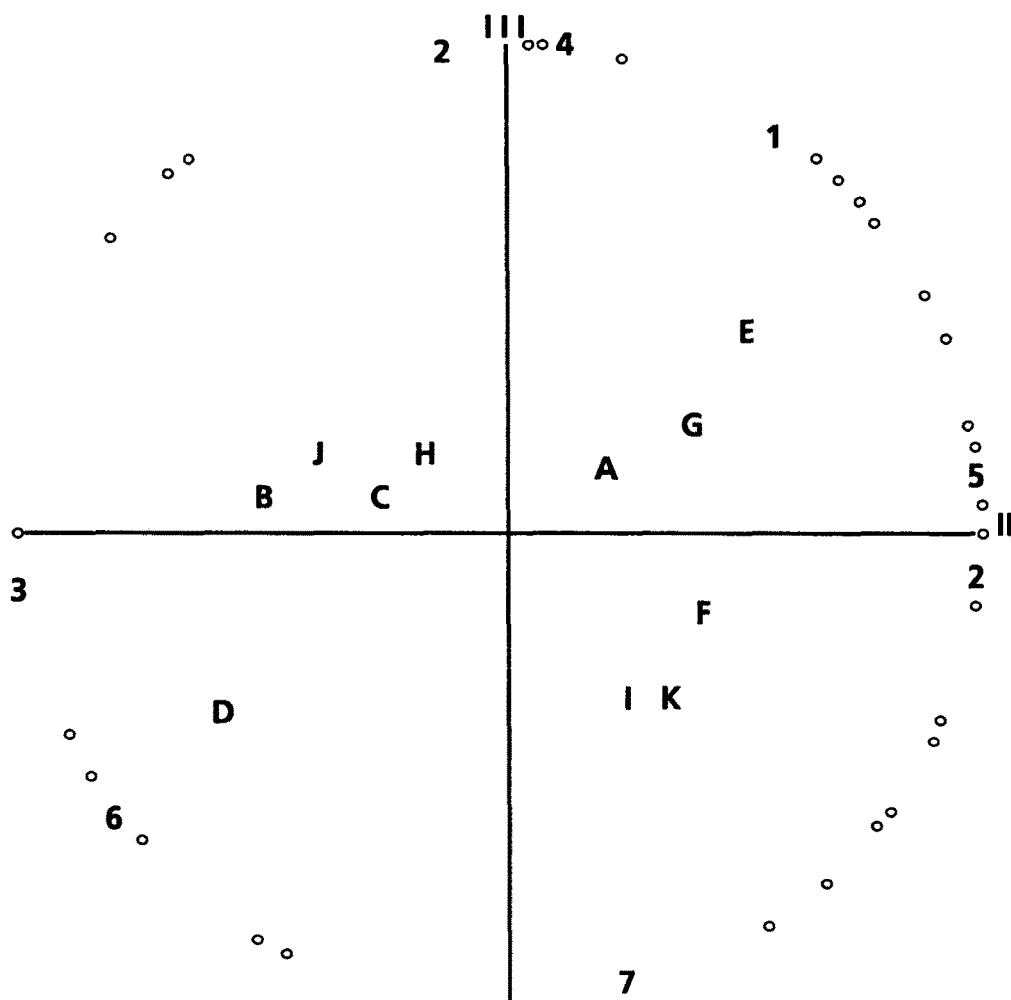


FIGURE 3
Joint Space Plot of Dimensions 2 vs 3 for the Soft Drink Analysis.

Cola-NonCola brands is much more diverse as evidenced by the wider variance in individual choice vector orientations on Dimensions II and III (as shown in Figure 3). While the correlations between dimensions for **A** and **B** will vary according to the particular rotation utilized, for this particular solution, no dimension shared more than 10% common variance with any other dimension in **A** or **B** indicating that these dimensions are providing somewhat "independent" views of the structure in the data.

Another interesting perspective to this analysis concerns the estimated threshold constants δ displayed in Table 8 together with the predicted scalar products (not shown). As mentioned, assuming the model holds with no error, these δ_i can be viewed as estimated threshold values by individual which governs the choice process. According to the random utility model and latent variable approach, whenever the latent variable u_{ij} is greater than some threshold c_i^* , a choice is made. These additive constants δ_i represent an estimate of these threshold values. In comparing them with the computed scalar products, one can make predictions concerning the brands selected by each individual.

D. Prediction of Stimuli Not Included in the Analysis

Two hold-out brands, Pepsi Free® and Sugar Free Dr. Pepper®, were also tested with this same set of students as to their purchase and consumption frequency. In order to illustrate a unique feature of the brand/stimulus reparameterized methodology, we will attempt to predict the individual choices for these two brands given only their attributes or features as specified in Table 6. Since we have the actual purchase/consumption data, we can compare the predictions with the actual data in order to assess some degree of predictive validation. The brand reparameterization allows the user to translate the features for a brand, given an estimated γ , into predicted coordinates. Once these predicted coordinates have been calculated, one can take their scalar products with A and compare them to the estimated threshold constants by individual (see Table 8) to obtain predictions. Table 9 lists the features/attributes for these two hold-out soft drink brands according to the seven variables specified in Table 6. The table also presents the predicted coordinate locations and the resulting proportion of correct choices for the 45 students. Note that each of these proportions is significantly greater than the random or chance proportion of .50. This prediction to stimuli not included in the analysis can be very useful in marketing studies, for example, where one is interested in predicting

TABLE 9

Predictive validation for softdrinks

		<u>Features:</u>						
		\underline{X}_1	\underline{X}_2	\underline{X}_3	\underline{X}_4	\underline{X}_5	\underline{X}_6	\underline{X}_7
1.	Pepsi Free	1	0	0	0	0	1.4	1
2.	Sugar Free Dr. Pepper	1	0	1	1	0	1.0	0

		<u>Predicted Coordinates:</u>		
		<u>Dimension:</u>		
		<u>I</u>	<u>II</u>	<u>III</u>
1.	Pepsi Free	2.44	3.30	1.61
2.	Sugar Free Dr. Pepper	2.54	-2.40	1.48

		<u>Proportion of Correct Predictions:</u>
1.	Pepsi Free	.711
2.	Sugar Free Dr. Pepper	.733

choice/market share for new brand entries. Naturally, like regression analysis, the predictive aspects of this vector threshold model are valid for stimuli that have feature/background data within the range of such data for the stimuli included in the calibration. (Note that this could also be easily extended for individuals not included in the study if one had selected the reparameterization option for A.)

V. Discussion

Thus, the vector threshold model appears to provide a better description of the structure of the softdrink choice data as compared to the MDPREF and correspondence analysis solutions. In addition, the resulting solution is much easier to interpret and the resulting dimensions correspond to those previously derived in MDS studies presented in the marketing literature. As shown, the reparameterization option allows for the prediction of choice probabilities for stimuli (and/or individuals) not included in the original analysis. One could also generalize the DeSarbo and Hoffman (1987) combinatorial optimization methodology to such a reparameterized vector threshold solution to derive optimal (in terms of profitability) new brands. Two other major areas for further investigation can be defined.

A. Other Applications

The spatial vector model of choice presented can be utilized in a variety of ways to analyze binary data. In addition to analyzing "pick any/ n " or "pick any" data as described in the application, the spatial vector choice model could also provide a spatial representation of binary profile data, for example, brands having or not having certain attributes or benefits. Or, the methodology could be applied to two-way, one or two-mode binary asymmetric proximities to provide a spatial representation of both row and column individuals. For example, one could utilize this methodology to analyze word association data where the general entry y_{ij} would be a 1 if the column response was evoked with the row stimulus.

B. Future Work

A number of interesting research directions have been suggested by the development of this procedure for spatially representing binary data. Further massive Monte Carlo runs should be performed on all model versions under a diverse set of conditions to test the algorithm for local minimum or other potential problems such as other types of violations to the independence assumption. Also, the small sample properties of the χ^2 test must be further examined, as should the consistency of the estimated parameters given the problem of incidental parameters. Finally, Jedidi (1988) has extended this methodology to handle three-way choice arrays where, for example, choices are obtained from subjects for products purchased/chosen over time or across different situations.

References

- Addelman, S. (1962). Orthogonal main-effect plans for asymmetrical factorial experiments. *Technometrics*, 4, 21-46.
- Akaike, H. (1974). A new look at statistical model identification. *IEEE Transactions on Automatic Control*, AC-19, 716-723.
- Andersen, E. B. (1980). *Discrete statistical models with social science applications*. New York: North Holland.
- Bartholomew, D. J. (1980). Factor analysis for categorical data. *Journal of the Royal Statistical Society, Series B*, 42, 293-321.
- Bechtel, G. C., Tucker, L. R., & Chang, W. (1971). A scalar product model for the multidimensional scaling of choice. *Psychometrika*, 36, 369-388.

- Bentler, P. M., & Weeks, David, G. (1978). Restricted multidimensional scaling models. *Journal of Mathematical Psychology*, 17, 138–151.
- Benzécri, J. P., et al. (1973). *L'Analyse des Données. Vol. 1. La Taxinomie, Vol. 2. L'Analyse des Correspondances* [Data analysis. Correspondance analysis]. Paris: Dunod.
- Bloxom, B. (1978). Constrained multidimensional scaling in n spaces. *Psychometrika*, 43, 397–408.
- Bock, R. D., & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, 35, 179–197.
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52, 345–370.
- Carroll, J. D. (1980). Models and methods for multidimensional analysis of preferential choice (or other dominance data). In E. D. Lantermann & H. Feger (Eds.), *Similarity and choice* (pp. 234–289). Vienna: Hans Huber.
- Carroll, J. D., Pruzansky, S., & Kruskal, J. B. (1979). CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika*, 45, 3–24.
- Chow, G. C. (1983). *Econometrics*. New York: McGraw-Hill.
- Christofferson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, 40, 5–32.
- Coombs, C. H. (1964). *A Theory of data*. New York: Wiley.
- Cooper, L. G., & Nakanishi, M. (1983). Two logit models for external analysis of preferences. *Psychometrika*, 48, 607–620.
- Courant, R. (1965). *Differential and integral calculus, Vol. 1* (2nd ed.). New York: Wiley.
- Dawes, R. M. (1964). Social selection based on multidimensional criteria. *Journal of Abnormal and Social Psychology*, 68, 104–109.
- de Leeuw, J. (1973). *Canonical analysis of categorical data*. Unpublished doctoral dissertation, Psychological Institute, University of Leiden, The Netherlands.
- de Leeuw, J., & Heiser, W. (1980). Multidimensional scaling with restrictions on the configuration. In P. R. Krishnaiah (Ed.), *Multivariate analysis, Vol. 5* (pp. 501–522). New York: North Holland.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the E. M. algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1–38.
- DeSarbo, W. S. (1982). GENNCLUS: New models for general nonhierarchical clustering analysis. *Psychometrika*, 47, 426–449.
- DeSarbo, W. S., & Carroll, J. D. (1985). Three-way metric unfolding via alternating weighted least squares. *Psychometrika*, 50, 275–300.
- DeSarbo, W. S., Carroll, J. D., Lehmann, D. R., & O'Shaughnessy, J. (1982). Three-way multivariate conjoint analysis. *Marketing Science*, 1, 323–350.
- DeSarbo, W. S., & Hoffman, D. L. (1986). A new unfolding threshold model for the spatial representation of binary choice data. *Applied Psychological Measurement*, 10, 247–264.
- DeSarbo, W. S., & Hoffman, D. L. (1987). Constructing MDS joint spaces from binary choice data: A new multidimensional unfolding threshold model for marketing research. *Journal of Marketing Research*, 24, 40–54.
- DeSarbo, W. S., & Keramidas, E. M. (1983). A spatial choice model. *Proceedings of the Business and Economic Statistics Section of the American Statistical Association* (pp. 41–46). Washington, DC: American Statistical Association.
- DeSarbo, W. S., Oliver, R. L., and DeSoete, G. (1986). A new probabilistic MDS vector model. *Applied Psychological Measurement*, 10, 79–98.
- 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.
- DeSarbo, W. S., & Rao, V. R. (1986). A constrained unfolding model for product positioning analysis. *Marketing Science*, 5, 1–19.
- DeSoete, G., & Carroll, J. D. (1983). A maximum likelihood method for fitting the wandering vector model. *Psychometrika*, 48, 553–566.
- Einhorn, H. J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychology Bulletin*, 73, 221–230.
- Fletcher, R., & Reeves, C. M. (1964). Function minimization by conjugate gradients. *Computer Journal*, 7, 149–154.
- Gifi, A. (1981a). *Nonlinear multivariate analysis*. Leiden, The Netherlands: University of Leiden, Department of Data Theory.
- Gifi, A. (1981b). *Homogeneity analysis*. Unpublished paper, University of Leiden, Department of Data Theory, Leiden, The Netherlands.
- Green, P. E., & Wind, Y. (1973). *Multiattribute decisions in marketing: A measurement approach*. Hinedale, IL: Dryden Press.

- Greenacre, M. J. (1984). *Theory and application of correspondence analysis*. London: Academic Press.
- Hausman, J. A., & Wise, D. A. (1978). A conditional probit model for qualitative choice: Discrete decisions recognizing interdependencies and heterogeneous preferences. *Econometrica*, *46*, 403–426.
- Heiser, W. J. (1981). *Unfolding analysis of proximity data*. Doctoral dissertation, University of Leiden, Leiden, The Netherlands.
- Himmelblau, D. M. (1972). *Applied nonlinear programming*. New York: McGraw Hill.
- Hoffman, D. L., & Franke, G. (1986). Correspondence analysis: A multidimensional analysis of categorical data in marketing research. *Journal of Marketing Research*, *23*, 213–227.
- Kruskal, J. B. (1981). Multilinear models for data analysis. *Behaviormetrika*, *10*, 1–20.
- Kruskal, J. B., & Shepard, R. N. (1974). A nonmetric variety of linear factor analyses. *Psychometrika*, *39*, 123–157.
- Jedidi, K. (1988). *A stochastic three-way MDS choice methodology*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia, PA.
- Lancaster K. (1966). A new approach to consumer theory. *Journal of Political Economy*, *74*, 132–157.
- Lancaster, K. (1979). *Variety, equity, and efficiency*. New York: Columbia University Press.
- Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive statistical analysis*. New York: John Wiley.
- Lehmann, D. R. (1985). *Market research and analysis* (2nd ed.). Homewood, IL: R. D. Irwin.
- Levine, J. H. (1979). Joint-space analysis of 'pick any' data: Analysis of choices from an unconstrained set of alternatives. *Psychometrika*, *44*, 85–92.
- Lewin, K., Tamara, D., Festinger, L., & Sears, P. (1944). Level of aspiration. In J. M. Hunt (Ed.), *Personality and the behavior disorders* (Vol. 1, pp. 333–378). New York: Ronald Press.
- Maddala, G. S. (1986). *Limited dependent and qualitative variables in econometrics*. New York: Cambridge University Press.
- McCullagh, P., & Nelder, J. A. (1983). *Generalized linear models*. New York: Chapman and Hall.
- McFadden, D. (1976). Quantal choice analysis: A survey. *Annals of Economic and Social Measurement*, *5*, 363–390.
- Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, *43*, 551–560.
- Muthén, B. (1981). Factor analysis of dichotomous variables: American attitudes toward abortion. In Jackson & Borgatta (Eds.), *Factor analysis and measurement in sociological research* (pp. 114–136). Beverly Hills, CA: Sage Press.
- Nelder, John A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A*, *135*, 370–384.
- Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications*. Toronto: University of Toronto Press.
- Noma, E., & Johnson, J. (1977). Constraining nonmetric multidimensional scaling configurations (Tech. Rep. No. 60). Ann Arbor: The University of Michigan, Human Performance Center.
- Powell, M. J. D. (1977). Restart procedures for the conjugate gradient method. *Mathematical Programming*, *12*, 241–254.
- Siegel, S. (1957). Level of aspiration and decision making. *Psychological Review*, *64*, 253–262.
- Simon, H. A. (1959). Theories of decision-making in economics and behavioral science. *American Economic Review*, *49*, 253–283.
- Simon, H. A. (1978). Rationality as process as product of thought. *American Economic Association*, *68*, 1–16.
- Slater, P. (1960). The analysis of personal preferences. *British Journal of Statistical Psychology*, *13*, 119–135.
- Takane, Y. (1983). *Choice model analysis of the "pick any/n" type of binary data*. Handout for Presentation at the European Meetings of the Psychometric and Classification Societies, Jouy-en-Josas, France.
- Tenenhaus, M., & Young, F. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis, and other methods for quantifying categorical, multivariate data. *Psychometrika*, *50*, 91–119.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
- Tucker, L. R. (1960). Intra-individual and inter-individual multidimensionality. In H. Gulliksen & S. Messick (eds.), *Psychological scaling: Theory and applications* (pp. 155–167). New York: Wiley.

Manuscript received 11/13/84

Final version received 11/27/87