A STOCHASTIC MULTIDIMENSIONAL SCALING PROCEDURE FOR THE SPATIAL REPRESENTATION OF THREE-MODE, THREE-WAY PICK ANY/J DATA

KAMEL JEDIDI

MARKETING DEPARTMENT
GRADUATE SCHOOL OF BUSINESS
COLUMBIA UNIVERSITY

WAYNE S. DeSARBO

MARKETING AND STATISTICS DEPARTMENT
GRADUATE SCHOOL OF BUSINESS
UNIVERSITY OF MICHIGAN

This paper presents a new stochastic multidimensional scaling procedure for the analysis of three-mode, three-way pick any/J data. The method provides either a vector or ideal-point model to represent the structure in such data, as well as "floating" model specifications (e.g., different vectors or ideal points for different choice settings), and various reparameterization options that allow the coordinates of ideal points, vectors, or stimuli to be functions of specified background variables. A maximum likelihood procedure is utilized to estimate a joint space of row and column objects, as well as a set of weights depicting the third mode of the data. An algorithm using a conjugate gradient method with automatic restarts is developed to estimate the parameters of the models. A series of Monte Carlo analyses are carried out to investigate the performance of the algorithm under diverse data and model specification conditions, examine the statistical properties of the associated test statistic, and test the robustness of the procedure to departures from the independence assumptions. Finally, a consumer psychology application assessing the impact of situational influences on consumers' choice behavior is discussed.

Key words: multidimensional scaling, binary data analysis, maximum likelihood estimation, stochastic models, consumer psychology.

1. Introduction

The collection of three-mode, three-way binary data is frequently encountered in the social sciences. In consumer psychology, for example, each subject may be presented with a set of consumption situations and a set of J brands and asked to indicate the r (or any) brands in a specified product class that he or she would most likely buy/use under each of the specified situations (see Belk, 1974). If r is specified, then we have what Coombs (1964, p. 33) calls "pick r/J" data. In the case where the subject is not constrained as to how many products to select, the data is classified as "pick any/J" (Coombs, 1964, p. 295). Another area where three-mode, three-way binary data can be collected is in repeated measures experiments, where the subject's choice behavior is measured before and after a given treatment. The objective is to examine the treatment effect on some relevant behavior measured by a binary variable (e.g., presence/absence of some characteristics, yes/no, buy/no buy, etc.). Still another example

occurs when subjects are asked to select the best objects or stimuli (e.g., universities) on each of some specified attributes (e.g., prestige). Coombs labels this as "pick any" data since neither the list of objects to choose from nor how many to choose are provided to the respondent. For analysis purposes, one can convert the pick any data into pick any/J format by simply considering all the objects revealed by the respondents. Note that such types of binary data (i.e., pick any/J and pick r/J) collected across subjects can be represented by a three-mode, three-way array (subject × stimulus × situation). The word "situation" is utilized hereafter to designate the third mode of the data and can represent various aspects such as use occasion, time, place, scenario, treatment, etc..

The development of various methods for the representation of *two-way* binary data has received much attention from psychometricians in a number of different research areas. Item response methods (see Lord, 1980; Takane, 1983), factor analysis of binary data (see Bartholomew, 1980; Bock & Lieberman, 1970; Christoffersson, 1975; Kruskal & Shepard, 1974; Muthén, 1978), correspondence analysis and its variants (see de Leeuw, 1973; Gifi, 1990; Greenacre, 1984; Heiser, 1981; Lebart, Morineau, & Warwick, 1984; Levine, 1979; Nishisato, 1980), and recent methods for stochastic multidimensional scaling (see DeSarbo & Cho, 1989; DeSarbo & Hoffman, 1986, 1987) are among the major procedures for representing the structure in such two-way binary data.

In contrast to this ample and growing methodological development for the analysis of two-way binary data, few researchers have attempted to generalize these two-way binary procedures or to develop new procedures that can accommodate the analysis of three-way, three-mode binary data (see DeSarbo, Lehmann, Gupta, Holbrook, & Havlena, 1987, for one approach). Existing MDS procedures cannot satisfactorily accommodate the spatial analysis of three-way, three-mode binary data. The stochastic multidimensional scaling threshold method proposed in this paper has been purposely designed to accommodate the spatial analysis of three-way, three-mode pick any/J data. It is basically a generalization of the DeSarbo and Hoffman (1986), DeSarbo and Cho (1989), and DeSarbo et al. (1987) approaches to the analysis of pick any/J data. This new procedure fits both vector and ideal-point models, and characterizes the effect of situations by a set of dimension weights reflecting the importance of each of the extracted dimensions in each situation. In cases where such an effect is subject specific (i.e., large subject × situation interactions), the procedure allows the ideal points or the vector directions to be different (floating) from one situation to another (DeSarbo, 1978). With the stochastic formulation of this method, one could (theoretically) utilize asymptotic statistical tests for dimensionality identification and model selection. Finally, the new procedure provides reparameterization options (see Carroll, Pruzansky, & Kruskal, 1980) for the coordinates of individuals and/or stimuli; these options can be utilized for prediction and aid in the interpretation of the derived dimensions.

2. The Models

Let

$i = 1, \ldots, I$ subjects;

$j = 1, \ldots, J$ stimuli;

$k = 1, \ldots, K$ situations;

$n = 1, \ldots, N$ attributes of the subjects (e.g., demographic and/or psychographic background variables);

$m = 1, \ldots, M$ attributes of the stimuli;

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

y_{it}: the t-th coordinate (ideal point or vector terminus) for subject i;

$y_{it}^{(k)}$: the t-th coordinate (ideal point or vector terminus) for subject i in situation k;

x_{jt}: the t-th coordinate for stimulus j;

w_{kt}: the weight for dimension t in situation k;

p_{ijk}: the probability that subject i chooses stimulus j in situation k;

Δ_{ijk}: 1 if subject i chooses stimulus j in situation k, 0 otherwise;

$\overline{a}_{in}$: the value of characteristic n for subject i;

b_{jm}: the value of attribute m for stimulus j;

α_{nt}: the coefficient of the n-th attribute describing subjects on dimension t;

γ_{mt}: the coefficient of the m-th attribute describing stimuli on dimension t.

We assume that the choice process of subject i choosing stimulus j in situation k is Bernoulli (independence is assumed across subjects, stimuli, and situations), with probability of choice given by p_{ijk}. As a consequence of this assumption, $\Sigma_j p_{ijk}$ need not be equal to one. This allows for the consideration of pick any/J choice data where a choice of multiple stimuli in a given situation is possible. However, the independence assumption over stimuli is questionable if the data are pick 1/J or pick r/J since, for a specific situation, knowing which r stimuli are chosen implies that all the other (J-r) stimuli are not selected in that particular situation. The independence assumption over situations is less problematic when situations are independent of each other and stimulus choice is highly affected by situational variations. We examine these independence assumptions later in the paper.

Here, situations are assumed to affect the importance of the dimensions (DeSarbo, 1978; Miller & Ginter, 1979) and/or the ideal points or the vector directions (DeSarbo, 1978; Dickson, 1982). The "fixed" models assume that subjects are homogeneous in terms of their situational perceptions. This is in contrast to the "floating" models where the effect of situation interacts with the subject. These models are described below.

The Fixed-Vector Model

We define a latent, unobservable utility variable F_{ijk} as

$$F_{ijk} = U_{ijk} + e_{ijk},\qquad(1)$$

where

$$U_{ijk} = \sum_{t=1}^{T} w_{kt} x_{jt} y_{it},\qquad(2)$$

and

$$e_{ijk} \text{ is error distributed as iid } N(0, \sigma_{ijk}^2).\qquad(3)$$

Here, U_{ijk} refers to a true, latent utility score for subject i concerning stimulus j in situation k. It is modeled as the weighted scalar product of the stimulus' coordinates (x_{jt}) and a subject's vector (y_{it}). The weights (w_{kt}) reflect the salience or importance of each derived dimension in each situation. This CANDECOMP (Carroll & Chang, 1970), specification in (2) assumes a common (situation-free) stimulus/subject space with differential weighting of dimensions for each situation. We can derive a situation-specific joint space of stimuli and subjects here in a number of ways. One way is to embed the situations' weights in the subject vectors

$$y_{it}^{(k)} = w_{kt}y_{it}. \tag{4}$$

Alternatively, one could apply the situations' weights to both stimulus j's and subject i's coordinates via

$$x_{jt}^{(k)} = \sqrt{w_{kt}}x_{jt}, \tag{5}$$

$$y_{it}^{(k)} = \sqrt{w_{kt}}y_{it}, \tag{6}$$

where $x_{jt}^{(k)}(y_{it}^{(k)})$ represents stimulus j's (subject i's) coordinate on dimension t in the space defined by situation k. This alternative is appropriate if $w_{kt} \geq 0$, for all k, t. Under this vector model specification, the magnitude of utility of the different stimuli to a particular subject in a particular situation is assumed to be given by the projection of these different stimuli onto that particular subject's vector in the space corresponding to that particular situation. The higher the projection of a stimulus, the higher is the utility of that stimulus to that subject in that particular situation. Here, utility is assumed to change monotonically with all dimensions.

We define F_{ijk} such that if $F_{ijk} \leq V_i$ (some individual threshold level), then we observe $\Delta_{ijk} = 0$ (no choice), and if $F_{ijk} > V_i$, we observe $\Delta_{ijk} = 1$ (a choice). This simply means that stimulus j is selected or chosen by subject i in situation k if its projection onto subject i's vector in situation k lies above his or her threshold level (V_i). This general specification is quite common in the econometrics literature (see McFadden, 1976) where discrete choice models are tied into latent, indirect utility scores and threshold values (also, see DeSarbo & Cho, 1989).

Thus,

$$P(\Delta_{ijk} = 0) = 1 - p_{ijk}$$

$$= P(F_{ijk} \leq V_i)$$

$$= P\left(\sum_{t=1}^{T} w_{kt}x_{jt}y_{it} + e_{ijk} \leq V_i\right)$$

$$= P\left(e_{ijk} \leq -\sum_{t=1}^{T} w_{kt}x_{jt}y_{it} + V_i\right) = \Phi(\cdot), \tag{7}$$

where $\Phi(\cdot)$ represents the standard normal distribution function evaluated at:

$$\frac{V_i - \sum_{t=1}^{T} w_{kt}x_{jt}y_{it}}{\sigma_{ijk}}, \tag{8}$$

and

$$P(\Delta_{ijk} = 1) = P(F_{ijk} > V_i) = p_{ijk}$$

$$= P\left(e_{ijk} > -\sum_{t=1}^{T} w_{kt}x_{jt}y_{it} + V_i\right) = 1 - \Phi(\cdot). \tag{9}$$

Hence, with p_{ijk} expressed in terms of the w_{kt}, x_{jt}, V_i, σ_{ijk}, and y_{it} parameters, we can write the likelihood function as follows:

$$L = \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{k=1}^{K} (p_{ijk})^{\Delta_{ijk}} (1 - p_{ijk})^{1 - \Delta_{ijk}}, \tag{10}$$

and the corresponding log likelihood as

$$\ln L = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} [\Delta_{ijk} \ln (p_{ijk}) + (1 - \Delta_{ijk}) \ln (1 - p_{ijk})]. \tag{11}$$

Finally, by substituting (7) and (9) in (11), we obtain:

$$\ln L = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} [\Delta_{ijk} \ln (1 - \Phi(\cdot)) + (1 - \Delta_{ijk}) \ln (\Phi(\cdot))]. \tag{12}$$

Under this model specification, the procedure estimates $\mathbf{W} = (w_{kt})$, $\mathbf{X} = (x_{jt})$, $\mathbf{Y} = (y_{it})$, $\mathbf{V}$, and σ_{ijk} in maximizing (12), provided that $\Delta = (\Delta_{ijk})$ and T are given. Note, however, that one cannot estimate all the IJK σ_{ijk} parameters in addition to $\mathbf{W}$, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{V}$. There would not be sufficient degrees of freedom for such estimation. Hence, we need to set $\sigma_{ijk} = 1$ for all i, j, k, or for the subsequent ideal-point models, estimate a subject specific $\sigma_{ijk} = \sigma_i$ for all j, k (σ_i is not estimable in the vector model given (8) above). Also, note that the solution produced by this fixed-vector model analysis is determined up to a scale transformation. (One can multiply, say, x_{jt} (across all j) and y_{it} (across all i) by any arbitrary nonzero constants, $c_t^{(1)}$ and $c_t^{(2)}$, respectively, and compensate by dividing w_{kt} (across all k) by $(c_t^{(1)} c_t^{(2)})$ and not alter the right side of (2).) Therefore, for the fixed-vector model, there are T scalar indeterminacies.

The Floating-Vector Model

This model generalizes the fixed-vector model presented above by treating subjects' vectors as freely varying over situations. More specifically, U_{ijk} is now modeled as

$$U_{ijk} = \sum_{t=1}^{T} w_{kt} x_{jt} y_{it}^{(k)}. \tag{13}$$

Unlike the fixed-vector model, the floating-vector model assumes a common space for stimuli and subjects' vectors which vary by situation, both weighted to obtain a situation-specific space. However, in (13), the w_{kt}'s can be embedded in the $y_{it}^{(k)}$'s, so without loss of generality, the w_{kt}'s can all be set equal to 1. The interpretation of the results should be similar to the fixed-vector model case; however, we should note that the vectors are not fixed over the different situations. To derive the log likelihood function corresponding to the floating-vector model, we substitute U_{ijk} in (13) for U_{ijk} in (2) and follow the same steps described by (7) to (12). Thus, we obtain a log likelihood function identical in form to (12), but with

$$\Phi(\cdot) = \Phi \left(\frac{V_i - \sum_{t=1}^{T} w_{kt} x_{jt} y_{it}^{(k)}}{\sigma_{ijk}} \right). \tag{14}$$

The log likelihood function is to be maximized with respect to $\mathbf{X} = (x_{jt})$, $\mathbf{Y}^{(k)} = (y_{it}^{(k)})$, for $k = 1, \ldots, K$, and $\mathbf{V}(\sigma_{ijk} = 1)$, given $\mathbf{\Delta} = (\Delta_{ijk})$ and T. As discussed, this model is characterized by KT indeterminacies in the solution concerning $\mathbf{W}$, which can be set equal to the unit matrix of one's. In addition, there are T^2 indeterminacies with respect to $\mathbf{X}$ and $\mathbf{Y}^{(k)}$.

It is worthwhile to note the behavioral difference between the fixed-vector and floating-vector models. In the fixed-vector model, the change in subjects' utility for various stimuli across different situations is affected only by the situations' weights for the derived dimensions. Here, all subjects' vectors vary uniformly as a result of a multiplication of the dimensions across situations, and these vectors cannot vary freely by situation over the derived MDS space. Each subject's vector is modeled to receive the same w_{kt} for a specific situation and dimension. This is in direct contrast to the floating-vector model where the change in subjects' utility across situations is a result of a subject by situation interaction, where situational factors may not affect all subjects' utility in the same manner; here, the effect varies by subject. Thus, in this latter case, subjects' vectors can vary freely over the derived space. The use of either model depends on how homogeneous subjects' choices are in each situation. If there is such commonality in situational effects among subjects, then the fixed-vector model is more appropriate.

The Fixed Ideal-Point Model

We define a latent unobservable disutility variable F_{ijk}^* as

$$F_{ijk}^* = U_{ijk}^* + e_{ijk}, \tag{15}$$

where e_{ijk} is as defined in (3) and

$$U_{ijk}^* = \sum_{t=1}^{T} w_{kt} (x_{jt} - y_{it})^2. \tag{16}$$

Here, U_{ijk}^* refers to a true, latent disutility score for subject i concerning stimulus j in situation k. Unlike the vector model, U_{ijk}^* is represented by the weighted unfolding model (see DeSarbo & Carroll, 1985), where the weighted distance between the stimulus' point (x_{jt}) and a subject's ideal point (y_{it}) corresponds to the respective utility of the stimulus to the subject in a particular situation. In other words, in a situation-specific space, the closer the stimulus point is to a subject's ideal point, the higher the utility of that stimulus is for that subject in that situation. As in the vector model, the weights (w_{kt}) reflect the importance of each derived dimension for each situation. Also, this fixed ideal-point model assumes a common joint space for subjects and stimuli with differential weighting of dimensions for each situation. The situation-specific spaces are derived by applying the situation specific weights (w_{kt}) to the derived stimulus' points (x_{jt}) and subjects' ideal-point coordinates (y_{it}), as described in (5) and (6), assuming $w_{kt} \geq 0$.

F_{ijk}^* is defined such that if $F_{ijk}^* \leq V_i^*$ (some individual threshold level), then one observes $\Delta_{ijk} = 1$ (a choice); if $F_{ijk}^* > V_i^*$, then one observes $\Delta_{ijk} = 0$ (no choice). This simply means that stimulus j is chosen by subject i in situation k if its weighted distance

from subject i's ideal point in situation k is less than or equal to that subject's threshold value V_i^*. Thus,

$$P(\Delta_{ijk} = 1) = P(F_{ijk}^* \leq V_i^*) = p_{ijk}$$

$$= P(U_{ijk}^* + e_{ijk} \leq V_i^*)$$

$$= P\left(e_{ijk} \leq -\sum_{t=1}^{T} w_{kt}(x_{jt} - y_{it})^2 + V_i^*\right) = \Phi(\cdot), \qquad (17)$$

where $\Phi(\cdot)$ represents the standard normal distribution function evaluated at

$$\frac{V_i^* - \sum_{t=1}^{T} w_{kt}(x_{jt} - y_{it})^2}{\sigma_{ijk}}. \qquad (18)$$

Similarly,

$$P(\Delta_{ijk} = 0) = P(F_{ijk}^* > V_i^*) = 1 - p_{ijk}$$

$$= P(U_{ijk}^* + e_{ijk} > V_i^*)$$

$$= P\left(e_{ijk} > -\sum_{t=1}^{T} w_{kt}(x_{jt} - y_{it})^2 + V_i^*\right) = 1 - \Phi(\cdot). \qquad (19)$$

Therefore, one can write the log likelihood function for this model as

$$\ln L^* = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} [\Delta_{ijk} \ln (\Phi(\cdot)) + (1 - \Delta_{ijk}) \ln (1 - \Phi(\cdot))], \qquad (20)$$

where $\Phi(\cdot)$ is as expressed in (18).

Thus, for the fixed ideal-point model, the method will estimate the weights for the situations $W = (w_{kt})$, the coordinates for the stimuli $X = (x_{jt})$, the ideal points $Y = (y_{it})$, V^*, and σ_{ijk} (= 1 or σ_i) by maximizing (20), given $\Delta = (\Delta_{ijk})$ and T. Also, for the fixed ideal-point model, there are T indeterminacies because of the fact that we can multiply $w_{kt} > 0$ by a positive constant c_t, and divide x_{jt} (across all j) and y_{it} (across all i) by $\sqrt{c_t}$ and still obtain the same U_{ijk}^*. In addition, there are T additional indeterminacies because one can center the concatenated matrix

$$L = \begin{pmatrix} X \\ Y \end{pmatrix}$$

by dimension and not alter the distance between the two sets of points (for subjects and stimuli).

The Floating Ideal-Point Model

This model is a generalization of the fixed ideal-point model. Ideal points are no longer fixed, but assumed to vary over situations. Thus, we can re-express U_{ijk}^* in (16) as

$$U^*_{ijk} = \sum_{t=1}^{T} w_{kt}(x_{jt} - y_{it}^{(k)}),$$ (21)

to accommodate the analysis of such ideal points that vary across situations. As a consequence, we need to estimate a common space for stimuli represented by the coordinates $\mathbf{X}$, a set of K situation-specific ideal-point coordinates for subjects represented by $\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(K)}$, and the weights for situations represented by $\mathbf{W}$. The situation-specific spaces are determined similarly, as in the fixed ideal-point model, except for the fact that the ideal points now float or vary by situation. Here, the distance between a stimulus' location and a subject's ideal point reflects the magnitude of the probability of choice of that stimulus by that particular individual in that particular situation. The smaller this distance, the higher the probability of choice. The likelihood function is determined similarly, as in the fixed ideal-point model, but with U^*_{ijk} in (16) replaced by U^*_{ijk} in (21). Thus, we obtain a log likelihood function identical in form to (20), but with

$$\Phi(\bullet) = \Phi\left(\frac{V_i^* - \sum_{t=1}^{T} w_{kt}(x_{jt} - y_{it}^{(k)})^2}{\sigma_{ijk}}\right).$$ (22)

This log likelihood function is to be maximized with respect to $\mathbf{W}$, $\mathbf{X}$, $\mathbf{Y}_k$, for $k = 1, \ldots, K$, $\mathbf{V}^*$, and σ_{ijk} (= 1 or σ_i), given Δ and T. Also, note that the floating ideal-point model is characterized by the same types of indeterminacies as the fixed ideal-point model.

Model Degrees of Freedom

The model degrees of freedom (df) are defined as the number of independent parameters to be estimated. The number of indeterminacies of a particular model is to be subtracted from the total number of parameters to be estimated to obtain that particular model's degrees of freedom. The various models' (those without reparameterizations) degrees of freedom are summarized in Table 1 (assuming $\sigma_{ijk} = 1$, for all i, j, k, and $V_i = V_i^* = V$, for all i).

Program Options

The new procedure can accommodate both external and internal analyses. In an external analysis, the procedure solves for the set of nonfixed parameters. For example, if the user supplies $\mathbf{X}$ (which might be derived from another MDS analysis), then one would solve for $\mathbf{W}$, $\mathbf{V}$, σ, and $\mathbf{Y}$. In an internal analysis, one solves for all of the designated parameters of the particular model specified (i.e., $\mathbf{W}$, $\mathbf{X}$, $\mathbf{Y}$, σ, $\mathbf{V}$).

As indicated previously, this method provides reparameterization options (Carroll, et al., 1980; DeSarbo & Rao, 1984; 1986; DeSarbo, Carroll, Lehmann, & O'Shaughnessy, 1982) whereby users can reparameterize or constrain the subjects' coordinates as linear functions of subjects' background variables. This can also be done for the coordinates of the stimuli. For example, if data are available on attributes for the stimuli, then one can reparameterize the coordinate, x_{jt} (regardless of the utility model specified), as:

$$x_{jt} = \sum_{m=1}^{M} b_{jm} \gamma_{mt},$$ (23)

TABLE 1

Model Degrees of Freedom

Model	Parameters	Number of Indeterminancies	Model Degrees of Freedom
Fixed-Vector Model	$x_{jt}, y_{it}, w_{kt}, V$	T	$[(I+J+K)T+1]-T$
Floating-Vector Model	$x_{jt}, y_{it}^{(k)}, w_{kt}, V$	$T(K+T)$	$[(IK+J+K)T+1]-T(K+T)$
Fixed Ideal-Point Model	$x_{jt}, y_{it}, w_{kt}, V$	$2T$	$[(I+J+K)T+1]-2T$
Floating Ideal-Point Model	$x_{jt}, y_{it}^{(k)}, w_{kt}, V$	$2T$	$[(IK+J+K)T+1]-2T$

Note: This assumes an internal analysis without reparameterizations, $\sigma_{ijk}=1$, and a constant threshold V across subjects.

where b_{jm} is the value of attribute m for stimulus j, and γ_{mt} measures the contribution of attribute m to dimension t. These coefficients (γ_{mt}) are assumed to be constant over situations. As a result of this reparameterization, one obtains an MDS space where locations of the stimuli are expressed as linear functions of their respective attributes. The γ_{mt} coefficients can also aid in interpreting the resulting dimensions as well as delineating the contribution of these attributes on each of the derived dimensions. Similarly, if data on subjects' background variables (e.g., demographics or psychographics) are available, then a subject's coordinates (y_{it}) can be reparameterized as

$$y_{it} = \sum_{n=1}^{N} a_{in} \alpha_{nt}, \tag{24}$$

where a_{in} is the value of characteristic n for subject i, $A = (a_{in})$, and α_{nt} is the contribution of characteristic n on dimension t. Note, however, that this way of reparameterizing subjects' coordinates is appropriate only for the fixed vector and ideal-point models where a common subjects' space with differential weighting for each situation is assumed. In the floating vector and ideal-point models, the subjects' coordinate reparameterizations can be specified in a number of ways. For example,

$$y_{it}^{(k)} = \sum_{n=1}^{N} a_{in} \alpha_{nt}^{(k)}, \tag{25}$$

in those cases where one can assume that the subjects' characteristics are constant, but their effects are varying across situations. Alternatively, they can be specified as

$$y_{it}^{(k)} = \sum_{n=1}^{N} a_{in}^{(k)} \alpha_{nt}. \tag{26}$$

Here, the subjects' characteristics are changing (across situations), but their effects are constant across situations. However, when both subjects' characteristics and their effects change across situations, we can specify

$$y_{it}^{(k)} = \sum_{n=1}^{N} a_{in}^{(k)} \alpha_{nt}^{(k)}. \tag{27}$$

Also, in the vector model(s), to avoid problems that might result from constraining subjects' vectors to be linear function of individuals' background variables (see Carroll et al. 1980), an option of estimating a stretching/shrinking parameter, θ_i, by subject is available. This parameter, θ_i, would appear as a multiplicative term on the right sides of (24) through (27). Finally, one can reparameterize both sets of coordinates if the appropriate background data are available. Also, it should be noted that in all the reparameterized models, the number of stimulus' attributes, M (subject's characteristics, N), can not exceed the number of stimuli, J (subjects, I), since we can identify at most JT (IT or IKT) coordinates.

The new method also allows for both constrained (i.e., nonnegative w_{kt}) and unconstrained W for the fixed and floating ideal-point models. Note that in the vector models, one does not need to constrain the sign of these weights. If a particular dimension has a negative weight in a particular situation, then the direction of the subjects' vectors with respect to that particular dimension in that particular situation would be opposite to the corresponding directions in the situation-specific space.

V_i^* in the fixed and floating ideal-point models plays the role of an individual threshold parameter. If W is constrained to be nonnegative, then we also need to constrain V_i^* to be nonnegative. Concretely, assume that $e_{ijk} = 0$ and rewrite (17) as

$$P(\Delta_{ijk} = 1) = P\left(0 \le - \sum_{t=1}^{T} w_{kt}(x_{jt} - y_{it})^2 + V_i^*\right)$$

$$= P\left(\sum_{t=1}^{T} w_{kt}(x_{jt} - y_{it})^2 \le V_i^*\right).$$

Thus, if $w_{kt} \ge 0$, then the V_i^*'s need to be constrained to be nonnegative to avoid negative Euclidean distances. This special case also produces an additional scale indeterminacy between these weights and thresholds since one can compensate for larger weights by making the thresholds larger, and thus not change $\hat{\Delta}$. The final option entails modeling the threshold parameters as constant or as varying by subjects, situations, or stimuli.

The Algorithm

As noted previously, the estimation of the parameters X (or γ), Y (or α), W, σ, and V of a particular model entails maximizing (minimizing) its log likelihood function (-ln L). Among the many methods available for such a nonlinear optimization problem, the method of conjugate gradient with automatic restarts (Powell, 1977) is utilized. This method has very attractive properties. In particular, the method does not require the storage of any matrices as is necessary in quasi-Newton and second derivative methods, which is important here given the large number of parameters to be estimated. Also, it has been shown that it can avoid the typical cycling often encountered with steepest descent algorithms, and has quadratic termination properties (Himmelblau,

1972)—that is, the method will typically find the global optimum for a quadratic loss function in H steps, where H is the number of parameters to solve for. Also, as shown by Powell (1977), the conjugate gradient method with restarts requires fewer iterations than the conjugate gradient without restarts in finding the optimum solution for a general nonlinear function. Finally, we should note that this method has been successfully used by De Soete, DeSarbo, and Carroll (1985) in their tree-fitting procedures. We briefly summarize the major phases of the algorithm for the fixed-vector model below.

Phase I: Starting Estimates

Starting estimates can either be provided by the user or generated from the procedure. One can obtain "rational" starting estimates by performing a CANDECOMP (Carroll & Chang, 1970) metric analysis (in the case of the fixed-vector model) or a three-way unfolding (DeSarbo & Carroll, 1985) metric analysis (in the case of a fixed or floating ideal-point model) on the three-way, three-mode binary data Δ. Random starting values are generated for W, X (or γ), Y (or $Y^{(k)}$) (or α), and V from different uniform distributions. In the cases of the fixed and floating ideal-point model, the values of W and V must be feasible (e.g., $W \geq 0$ if W is constrained to be nonnegative, and V needs to also be nonnegative as a result). Set $IC = 0$.

Phase II: Estimate X, Y, W, and V

Set $IC = IC + 1$. This phase of the algorithm estimates X (or γ), Y (or α), W and V (or V if the threshold is to be constant over subjects). Estimates of these parameters are sought to minimize the minus of the log of the likelihood function in (12). The partial derivatives are

$$\frac{\partial(\ln L)}{\partial x_{jt}} = -\sum_i \sum_k w_{kt} y_{it} \phi(\bullet) \left[\frac{1 - \Delta_{ijk}}{\Phi(\bullet)} - \frac{\Delta_{ijk}}{1 - \Phi(\bullet)} \right], \tag{28}$$

$$\frac{\partial(\ln L)}{\partial \gamma_{mt}} = -\sum_i \sum_j \sum_k w_{kt} y_{it} b_{jm} \phi(\bullet) \left[\frac{1 - \Delta_{ijk}}{\Phi(\bullet)} - \frac{\Delta_{ijk}}{1 - \Phi(\bullet)} \right], \tag{29}$$

$$\frac{\partial(\ln L)}{\partial y_{it}} = -\sum_j \sum_k w_{kt} x_{jt} \phi(\bullet) \left[\frac{1 - \Delta_{ijk}}{\Phi(\bullet)} - \frac{\Delta_{ijk}}{1 - \Phi(\bullet)} \right], \tag{30}$$

$$\frac{\partial(\ln L)}{\partial \alpha_{nt}} = -\sum_i \sum_j \sum_k w_{kt} y_{it} a_{in} \phi(\bullet) \left[\frac{1 - \Delta_{ijk}}{\Phi(\bullet)} - \frac{\Delta_{ijk}}{1 - \Phi(\bullet)} \right], \tag{31}$$

$$\frac{\partial(\ln L)}{\partial w_{kt}} = -\sum_i \sum_j y_{it} x_{jt} \phi(\bullet) \left[\frac{1 - \Delta_{ijk}}{\Phi(\bullet)} - \frac{\Delta_{ijk}}{1 - \Phi(\bullet)} \right], \tag{32}$$

$$\frac{\partial(\ln L)}{\partial V_i} = \sum_j \sum_k \phi(\bullet) \left[\frac{1 - \Delta_{ijk}}{\Phi(\bullet)} - \frac{\Delta_{ijk}}{1 - \Phi(\bullet)} \right], \tag{33}$$

or, for a constant threshold over subjects,

$$\frac{\partial(\ln L)}{\partial V} = \sum_i \sum_j \sum_k \phi(\bullet) \left[\frac{1 - \Delta_{ijk}}{\Phi(\bullet)} - \frac{\Delta_{ijk}}{1 - \Phi(\bullet)} \right], \tag{34}$$

where $\phi(\cdot)$ represents the standard normal density function evaluated at (8) with $\sigma_{ijk} = 1$. For external analyses, the appropriate set of partial derivatives are set equal to zero. With these partial derivatives specified, we can now briefly describe the conjugate gradient procedure with automatic restarts.

(i) Start with initial parameter estimates $\psi^{(1)}$, and set the iteration counter $IC = 1$. ψ here denotes a vector, stacking all the parameters to be estimated (i.e., $\mathbf{X}$, $\mathbf{Y}$ (or $\mathbf{Y}^{(k)}$), $\mathbf{W}$, $\mathbf{V}$).

(ii) Set the first search direction $\mathbf{S}^{(1)} = -\nabla (\ln L)^{(1)}$, where $\nabla (\ln L)^{(1)}$ denotes the gradient vector of the log-likelihood function evaluated at $\psi^{(1)}$.

(iii) Find $\psi^{(2)}$ using the relationship

$$\psi^{(2)} = \psi^{(1)} + \lambda^{(1)} \mathbf{S}^{(1)}, \tag{35}$$

where $\lambda^{(1)}$ is the optimal step-size in the gradient direction $\mathbf{S}^{(1)}$. A quadratic interpolation method is used for estimating the optimal step-size. Set $IC = 2$.

(iv) Calculate $\nabla (\ln L)^{(IC)}$ and set the (new) search direction

$$\mathbf{S}^{(IC)} = -\nabla(\ln L)^{(IC)} + \beta \mathbf{S}^{(IC-1)}, \tag{36}$$

if $IC = 2$ or if a restart is needed. In this algorithm, restarts are made every H iterations (H is the number of parameters to be estimated) or when the search direction is not "sufficiently downhill". If this is the case, then set $R = IC$ and go to Step (vi). R holds the number of the iteration where a restart is made. Otherwise, set

$$\mathbf{S}^{(IC)} = -\nabla(\ln L)^{(IC)} + \beta \mathbf{S}^{(IC-1)} + \gamma S^{(R)}, \tag{37}$$

where

$$\beta = \frac{\nabla(\ln L)^{(IC)\prime}\nabla(\ln L)^{(IC)}}{\nabla(\ln L)^{(IC-1)\prime}\nabla(\ln L)^{(IC-1)}}, \tag{38}$$

$$\gamma = \frac{(\nabla(\ln L)^{(IC)} - \nabla(\ln L)^{(R)})\prime\nabla(\ln L)^{(IC)}}{(\nabla(\ln L)^{(IC)} - \nabla(\ln L)^{(R)})\prime\mathbf{S}^{(R)}}, \tag{39}$$

$\mathbf{S}^{(R)}$: search direction when a restart is made.

(v) Check if $\mathbf{S}^{(IC)}$ is "sufficiently downhill" using the following condition

$$-\Delta(\ln L)^{(IC)\prime}\mathbf{S}^{(IC)} \geq \rho\|\mathbf{S}^{(IC)}\|\|\nabla(\ln L)^{(IC)}\|, \tag{40}$$

where ρ is a positive constant. Powell (1977) suggests setting $\rho = 0.2$. If this condition is satisfied then go to step (vi), otherwise return to step (iv). This phase is performed only when $\mathbf{S}^{(IC)}$ is computed via (37).

(vi) Compute the optimal step-size at iteration IC in the direction $\mathbf{S}^{(IC)}$ and then compute

$$\psi^{(IC+1)} = \psi^{(IC)} + \lambda^{(IC)}\mathbf{S}^{(IC)}. \tag{41}$$

(viii) If $\psi^{(IC+1)}$ is optimal, stop. Otherwise set $IC = IC + 1$ and return to step (iv). The optimality of $\psi^{(IC+1)}$ is determined using the following convergence criteria.
     a. The amount of improvement of the log likelihood function between the last and before last iteration (i.e., $|\ln L^{(IC-1)} - \ln L^{(IC)}| \leq TOL$).
     b. The length of the gradient (i.e., $\|\nabla \ln L^{(IC)}\| \leq TOL$).

c. The maximum number of iterations set by the user (i.e., $IC \geq IC^*$).

Phase III: Normalization and Output

In the case of the fixed-vector model, we normalize the dimensions of X and Y to unit sums of squares and then redefine the matrix of situations' weights (W) to compensate for these transformations. Various goodness-of-fit measures (described below) are computed over the entire data, as well as by subject, stimulus, and situation, to examine possible outliers in the data.

These goodness-of-fit measures, computed for a particular solution, are

1. The log likelihood function;
2. A deviance measure (Nelder & Weddenburn, 1972):

$$D = -2 \left[\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} [\Delta_{ijk} \ln (\hat{p}_{ijk}) + (1 - \Delta_{ijk}) \ln (1 - \hat{p}_{ijk})] \right]$$

$$= -2 \ln L(\hat{p}_{ijk}). \tag{42}$$

This is basically (11) with p_{ijk} replaced by its estimated value, $\hat{p}_{ijk}$. The difference between two deviance measures corresponding to two nested models is asymptotically χ^2 distributed with degrees of freedom equal to the difference in the models' degrees of freedom. This difference can be used (theoretically) to test for dimensionality as well as for various model specifications because of the obvious nesting. However, one potential problem with using such a χ^2 test concerns the presence of incidental parameters in the likelihood function. In other words, the number of parameters to be estimated varies with the number of individuals, stimuli, and situations considered. In such a case, the MLEs are not consistent (Anderson, 1980), and thus the χ^2 test is not appropriate. In the next section, a Monte Carlo analysis investigating the use of such a χ^2 test will demonstrate its inappropriateness.

3. The Akaike information criterion (Akaike, 1974):

$$AIC = -2 \ln L(\hat{p}_{ijk}) + 2 \text{ (number of independent parameters in the model)}. \tag{43}$$

This AIC measure can be utilized to test for dimensionality and model selection. According to Akaike (1974), the model with minimum AIC should be selected. However, as noted by Bozdogan (1987), the use of this AIC criterion tends to result in over-fitting certain models.

4. A simple matching coefficient (Match) calculated between Δ and the predicted $\hat{\Delta}$, according to the threshold rule of the model being estimated (Sneath & Sokal, 1973).

5. The point biserial correlation (Pbc) between $\hat{p} = (\hat{p}_{ijk})$ and Δ.

6. The phi coefficient (Phi) calculated between Δ and $\hat{\Delta}$.

All of these goodness-of-fit measures need to be inspected in determining the dimensionality of the space and testing for nested models, given the problems and difficulties in using the χ^2 test and the AIC.

3. Monte Carlo Simulations

Algorithm Performance

A Monte Carlo analysis was performed to examine the performance of the algorithm as a number of model, data, and error factors were experimentally varied. Ten factors were initially hypothesized to affect the performance of the algorithm. These

TABLE 2

Independent Variables for Monte Carlo Analysis

Factor	Designation	Levels
I	Number of Subjects	20 35 50
II	Number of Stimuli	7 14 21
III	Number of Situations	2 4 6
IV	Number of Dimensions	2 3 4
V	Amount of Error (variance)	$.1 \times Var(U_{ijk})$ $.2 \times Var(U_{ijk})$ $.3 \times Var(U_{ijk})$
VI	Error Distribution	Normal Exponential Uniform
VII	Type of Model	Vector Ideal Point
VIII	Fixed/Floating Analysis	Fixed Floating
IX	Type of Analysis	Internal External ($\underset{\sim}{X}$ fixed)
X	Type of Threshold	Varying Constant

factors, along with their various levels, are listed in Table 2. These ten factors were combined via an asymmetric fractional factorial $(3^6 2^4)$ design (Addleman, 1962) for main-effects only estimation as used for such testing in DeSarbo (1982), DeSarbo and Carroll (1985), and DeSoete, DeSarbo, Furnas, and Carroll (1984). Twenty-seven experimental trials were devised according to the constructed design. For each experimental trial, **W**, **X**, **Y**, and **V** were generated randomly from different uniform distributions according to the stipulated factor levels of the particular trial. Then, $U = (U_{ijk})$ was created corresponding to the particular model of the experimental trial. Error was generated randomly from the corresponding distribution of the experimental trial and added to **U**. Finally, **Δ** was created using the threshold rule of the particular model.

With Δ and other input control parameters determined according to the particular trial in the design, the method was then executed.

Some five items were used as dependent (performance) measures for the algorithm: the CPU time (in seconds on an IBM AT (286) clone-12Mhz with a math co-processor) required for convergence, the overall simple matching coefficient between Δ and $\hat{\Delta}$, the root-mean-square error (RMSE) between the actual W and the recovered $\hat{W}$, the RMSE between X and $\hat{X}$, and the RMSE between Y and $\hat{Y}$. The last three dependent measures were calculated after appropriate permutation and normalization (depending on the model estimated and its respective indeterminacies). Thus, the first measure (CPU time) is a measure of computational expense, the second (simple matching coefficient) is a measure of overall goodness-of-fit of the solution, and the last three measures are measures of the goodness-of-recovery of the known configurations.

Given the values of the five dependent variables for each of the 27 trials, multiple regression analyses were conducted using the columns of the experimental design matrix as the independent variables. (Note that since the matching coefficient is restricted in values between 0 and 1, a logit [log $Y/(1 - Y)$] transformation (as recommended in Cox and Snell, 1989, p. 16, as a transformation to normality) was applied to this dependent variable (arc sine transformations were also applied with similar results), and the regression was performed using this transformed variable.) The regression results for the five dependent measures (the experimental design was converted to dummy variables and regressed on each dependent measure as in conjoint analysis) are summarized below.

From the regression analyses, it appears that the CPU time is affected significantly by the number of parameters to be estimated. Here, increasing the number of subjects to 50 significantly increases the amount of CPU time. The significant effect of type of model suggests that the ideal-point model requires more CPU time than the vector model. An unexpected result is that floating analyses require less CPU time than fixed analyses.

The matching coefficient, though having an overall insignificant F statistic, appears to be affected by the amount of error in the data and the type of model. The ideal-point model produces a slightly ($p < 0.05$) better fit to the generated data than the vector model.

Concerning the recovery of W, X, and Y, the regression results suggest that external analysis has a positive effect on the recovery of the three spaces. This is an obvious result given the smaller number of parameters to be estimated in an external analysis as compared to an internal analysis performed on the same data set. Note, however, that such an effect is not significant in the case of W recovery at the $p < 0.05$ level. Also, the recovery of both X and Y is significantly affected by the number of stimuli and subjects. Here, larger I and J levels significantly improve the recovery of these spaces. Generally, the recovery of the three spaces is affected positively by the gain in degrees of freedom as we increase the number of situations, stimuli, and subjects. Finally, note that among these last three regressions, only RMSE $(X, \hat{X})$ has an overall significant F statistic ($p < 0.01$). The RMSE $(W, \hat{W})$ and RMSE $(Y, \hat{Y})$ regressions are insignificant at the $p < 0.05$ level.

The results of this Monte Carlo analysis appear to demonstrate the adequate performance of the procedure. These findings show that the method is somewhat robust to nonnormality. Number of dimensions and type of threshold have no effect on the performance of the algorithm. Better parameter recovery appears to be associated with larger data sets and external (versus internal) analysis. Better overall fit is obtained with smaller amount of error and additional CPU is generally required for larger data sets.

The Validity of the χ^2 Test

As noted earlier, one potential problem with the χ^2 test concerns the presence of incidental parameters in the likelihood function (i.e., the number of parameters to be estimated varies with the order of the data). In such a case, the MLE's are not consistent (Anderson, 1980), and thus the χ^2 test is not appropriate. This is particularly relevant in our case since, typically, no replications are typically collected. In the face of this problem, we performed a Monte Carlo analysis for vector and ideal-point models to test whether the difference of deviance scores is truly asymptotically distributed as χ^2 with the difference in the degrees of freedom as the appropriate degrees of freedom.

One hundred synthetic data sets corresponding to $I = 30$ subjects, $J = 10$ stimuli, and $K = 3$ situations were generated in $T = 2$ dimensions according to the fixed-vector model in (3) with an internal analysis, a constant threshold V across subjects, and no reparameterization options. Error from $N(0, 1)$ was added to the true utilities (U_{ijk}), and the synthetic choice data Δ were created according to the fixed-vector model threshold rule. Each of these 100 data sets were subjected to a fixed-vector model analysis in $T = 2$ and $T = 3$ dimensions. For each pair of 100 runs (corresponding to $T = 2$ and $T = 3$), the difference in deviance scores was computed. Theoretically, for this particular case, this difference is χ^2 distributed with 42 degrees of freedom. The same procedures were adopted for the fixed ideal-point model case. One hundred synthetic data sets with the same input parameters and options as in the fixed-vector model case were generated according to the fixed ideal-point model in (16). Error from a $N(0, 1)$ distribution was added and the choice data Δ were created. A fixed ideal-point model analysis was performed in $T = 2$ and $T = 3$ dimensions with the difference in corresponding deviance scores (here with 41 degrees of freedom) calculated for each pair of 100 runs.

One hundred sample values from a χ^2 distribution with 42 degrees of freedom were generated for the fixed-vector model. These values and the one hundred values resulting from the fixed-vector model analyses were ordered, and a Q-Q plot (Wilk & Gnanadesikan, 1968) was performed on the resulting quntiles. Similarly, in the case of the fixed ideal-point model, one hundred sample values of a χ^2 distribution with 41 degrees of freedom were generated, ordered, and plotted against the ordered one hundred differences in deviance score values resulting from the ideal-point model Monte Carlo runs. If the difference in deviance scores is χ^2 distributed, then we should expect these plots (assuming equivalent scaling of the axes) to make a 45° line. The results do not indicate this in either case, especially in the tails of the distribution.

In addition to the Q-Q plots, the χ^2 distribution with 42 df was fitted to the fixed vector model difference in deviance score data. The χ^2 goodness-of-fit test is significant at the 0.05 level suggesting a significant lack of fit. Similarly, the χ^2 distribution with 41 df were fitted to the fixed ideal-point model difference in deviance score data. Here also, the χ^2 goodness-of-fit test is significant at the 0.05 level. It, therefore, does appear that there would be difficulties in using the χ^2 test for model selection and dimensionality identification. Similar difficulties in using the χ^2 test were also reported by Takane (1983) in his unfolding-type item response model and by DeSarbo and Hoffman (1986) and DeSarbo and Cho (1989) in their two-way unfolding and vector models, respectively. These results suggest the simultaneous examination of *all* the goodness-of-fit measures for model selection (as well as interpretability of the results).

Testing the Independence Assumptions

As noted, the procedure utilizes assumptions of independence across all modes of the data. In an attempt to check the robustness of this method to violations of inde-

pendence, several synthetic data sets characterized by dependence patterns across rows, columns, slices, and all combinations of these were constructed. For example, for testing independence across columns, structural patterns of relationships were created in the binary data (e.g., stimulus A is never chosen if stimulus B is chosen). In *all* seven series of tests (rows, columns, slices, rows and columns, rows and slices, columns and slices, rows and columns and slices), the procedure recovers the true X, Y, W, and V parameters extremely well in the appropriate dimensionality (as ascertained by the inspection of the various goodness-of-fit indices previously discussed.)

4. Application

Situational Factors in Consumer Psychology

In consumer psychology research, the effect of situational factors on consumers' choices has received growing recognition from many scholars (see Engel, Blackwell, & Kollat, 1969). Besides this recognition, empirical evidence is accumulating to support the explicit consideration of situational factors. Green and Rao (1972) found that consumer perceptions of and preferences for various bread and pastry items change markedly over different meal and menu situations. Sandell (1968) found that nearly 40% of the variance in individual choices for different drinks is explained by brand $\times$ situation interaction effects (as compared to 14.6% and 2.7% for brand and situation main effects, respectively), thereby reflecting the importance of situational influences on brand choice. Similarly, Belk (1974) has demonstrated that situational variables account for nearly half of the explained variance in both meat and snack preferences. Comparable findings have been obtained using other types of products/services such as leisure activities (Bishop & Witt, 1970), fast foods (Belk, 1975; Miller & Ginter, 1979), and time (Hornick, 1982).

Perhaps the most popular approach taken to define the concept of a situation is that proposed by Belk (1974, 1975). In terms of this approach, a situation is defined by a locus in time and space. An objective element of a situation is one that is capable of external verification without the need to construct measures of internal states of the individual such as mood, plan, and purposes (Belk, 1975). Besides the physical features of the situation (i.e., temperature, time of the day, person present, sounds, odors, decor, etc.), objective descriptions may include the existence of external facts and events that bear upon current behavior even though they are not themselves physically a part of the situation. For example, instead of attempting to measure the mood a person brings to the situations, one can ascertain whether the person has had a hectic day with his or her children, has just finished the last of several difficult final examinations, or has been promoted (Belk, 1975). More precisely, Belk (1975) defines a consumption situation as "all those factors particular to a time and place of observation which do not follow from a knowledge of personal (intra-individual) and stimulus' (choice alternative) attributes, and which have a demonstrable and systematic effect on current behavior" (p. 157).

In an attempt to define what he means by "all those factors ... ", Belk (1975) identifies five situational characteristics. *Physical surroundings* include geographical and institutional location, decor, sounds, aromas, lighting, weather, and visible configurations of merchandise or other material surrounding the stimulus object. *Social surroundings* include other persons present and their characteristics, their apparent roles, and interpersonal interactions. *Temporal perspective* is an aspect of situations that may be specified in units ranging from time of the day to season of the year. Time also may be measured related to some past or future event for the situational partici-

pant. *Task definition* features of a situation include an intent or requirement to select, shop for, or obtain information about a general or a specific purchase. In addition, task may reflect different buyer and user roles anticipated by the individual. *Antecedent states* include momentary moods (such as acute anxiety, pleasantness, hostility, and excitation) or momentary conditions (such as cash on hand, fatigue, or illness) rather than chronic individual traits. These momentary conditions are further stipulated as immediately antecedent to the current situation to distinguish states that the individual brings to the situation from individual states that result from the situation.

The strength of this taxonomy lies in its potential for operationalization since we can simply define a situation in terms of some or all the above situational characteristics. For example, the situation "you are at the store to pick up some things for a picnic you are planning with friends and are trying to decide what kind of snack to buy" (Belk 1974, p. 157), is made up of the physical surroundings (picnic), the social surroundings (friends), time (may be a weekend), and task definition (shopping for a picnic with friends).

Study Description

DeSarbo and Cho (1989) report on the collection of intended pick any/*J* choice judgments for various brands of soft drinks in five different consumption situations. The subjects were fifty MBA students from Columbia University in New York. The data for ten of the 50 individuals were discarded since they claimed not to drink any soft drink in one or more of the situations tested. Eleven brands of soft drinks were used: Coke (A), Diet Coke (B), Diet Pepsi (C), Diet 7UP (D), Dr. Pepper (E), Mountain Dew (F), Pepsi (G), Pepsi Light (H), Sprite (I), Tab (J), and 7UP (K).[1] The five situations considered were: (a) to quench your thirst on a very hot day, (b) a social beverage you drink with friends, (c) a beverage you drink with meals, (d) as a mixer, and (e) a beverage you relax with (watching TV or studying). These situations represent various usage occasions for beverages. All but the fourth have been utilized in previous research (see Sandell, 1968). The fourth situation "as a mixer" is also important to consider since many brands of soft drinks are positioned as mixers.

Using Belk's (1975) taxonomy of a situation, one can describe Situation 1 "to quench your thirst on a very hot day" as evoking mood (the feeling of being thirsty) and temporal (summertime) aspects. Situation 2 "A social beverage you drink with friends" captures a social surrounding aspect (friends). Situation 3 "A beverage you drink with meals" encompasses spatial (where you eat your meals), social (with whom you eat meals), and perhaps temporal aspects. Situation 4 "As a mixer" evokes a task or purpose aspect (using the soft drink as a mixer). Finally, Situation 5 "A beverage you relax with (watching TV or studying)" elicits spatial (at home) and mood aspects.

Analysis

Each of the four choice models (i.e., the fixed and floating-vector models and the fixed and floating ideal-point models) was estimated in $T = 1, 2, 3$, and 4 dimensions with $\sigma_{ijk} = 1$, a varying threshold by subject, and an external analysis option. The external estimates for the brands' coordinates were obtained from the two-way vector model analysis performed on the overall consumption data for these same subjects, as reported in DeSarbo and Cho (1989). In their two-way, binary data, a "1" indicates purchase and consumption at least every other week and "0" indicates purchase and consumption less than every other week. The choice of performing an external analysis was based on the fact that the external results ($T = 1, 2, 3, 4$) were already published.

---

[1] All products are registered trademarks of their respective companies.

TABLE 3

Analyses of Soft Drink Choice Data

T	df	LnL	Deviance	Pbc	Phi	Matching Coefficient	AIC
			Fixed Ideal-Point Model				
1	85	-960.8	1921.6	0.468	0.387	0.805	2091.6
2	130	-744.4	1488.8	0.629	0.536	0.842	1748.8
3	175	-645.3	1290.6	0.703	0.623	0.870	1640.6
4	220	-631.9	1263.8	0.712	0.624	0.872	1703.8
			Fixed-Vector Model				
1	84	-973.2	1946.2	0.460	0.341	0.799	2114.2
2	128	-790.5	1581.0	0.598	0.488	0.832	1837.0
3	172	-605.5	1211.0	0.720	0.635	0.875	1555.0
4	216	-572.3	1144.6	0.743	0.682	0.890	1576.0
3[a]	160	-680.5	1361.0	0.668	0.564	0.852	1681.0
			Floating Ideal-Point Model				
1	245	-844.4	1688.8	0.552	0.438	0.814	2178.8
2	450	-695.4	1390.8	0.654	0.533	0.839	2290.8
3	655	-419.3	838.6	0.816	0.759	0.914	2148.6
4	860	-346.2	692.4	0.854	0.804	0.930	2412.4
			Floating-Vector Model				
1	240	-884.7	1769.4	0.533	0.419	0.816	2249.4
2	440	-596.8	1193.6	0.719	0.620	0.870	2073.6
3	640	-244.1	488.2	0.900	0.873	0.955	1768.2
4	840	-87.6	175.2	0.966	0.951	0.982	1855.2

[a]Estimation with T=3 but with $w_{kt} = 1$ for all k,t.

Also, performing an external analysis has the advantage of estimating fewer parameters and this decreases the chance of locally optimum solutions. More importantly, there is no assurance that the brands' spaces we could obtain from performing internal fixed-vector, floating-vector, fixed ideal-point, and floating ideal-point model analyses would all be similar. Thus, having a common brands' space will facilitate the comparisons between the various solutions. Finally, these dimensions have also been uncovered in various perceptual studies performed in this market (e.g., Lehmann, 1985). DeSarbo and Cho (1989) present a $T = 3$ dimensional solution (reparameterized brands) as best representing the structure in this data. They interpret the first dimension as market
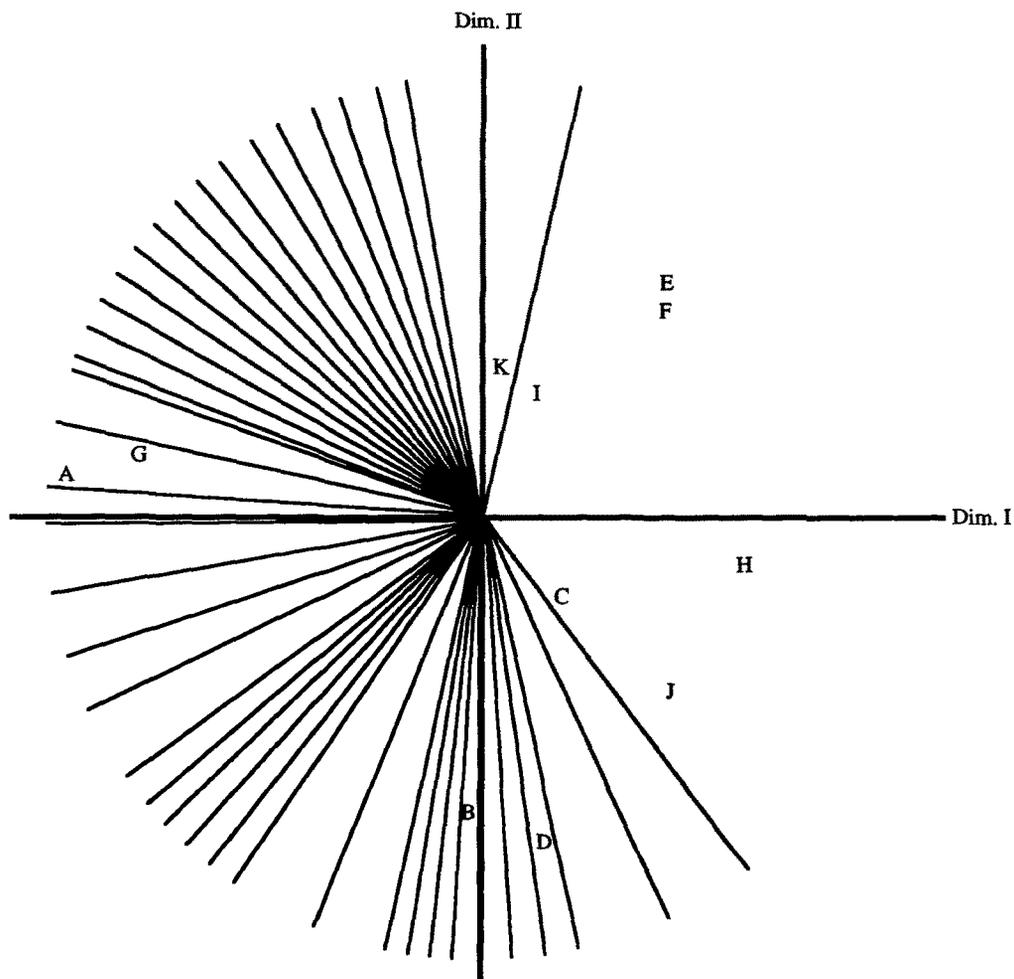
FIGURE 1
Joint space plot of Dimension 1 vs. 2 for the soft drink fixed-vector model.

share (proxy for popularity), the second as a Diet-NonDiet dimensions, and the third as a Cola-NonCola (caffeine) dimension.

Table 3 presents the statistical summaries for the fixed-vector, fixed ideal-point, floating ideal-point, and floating-vector models analyses, respectively. In all of these analyses, the minimum AIC is found with three dimensions. The other goodness-of-fit measures also support this conclusion. This result is congruent with the DeSarbo and Cho (1989) results.

We can also use the minimum AIC rule (Akaike, 1974) as a basis for selecting between fixed versus floating models in three dimensions. The minimum AIC is 1640.6 for the fixed ideal-point model and is 2148.6 for the floating ideal-point model. This suggests that ideal points do not float or vary from situation to situation, but are fixed. Situational factors seem to affect students' choice behavior in the same way. The dimensions' saliencies or importances vary across situations, not the ideal points. Similarly, the minimum AIC for the fixed-vector model is 1555.0 compared to 1768.2 for the floating-vector model. Here also, a fixed representation appears more appropriate than a floating representation. Thus, the two models selected are the fixed ideal-point model and the fixed-vector model. Since the fixed-vector model has a minimum AIC
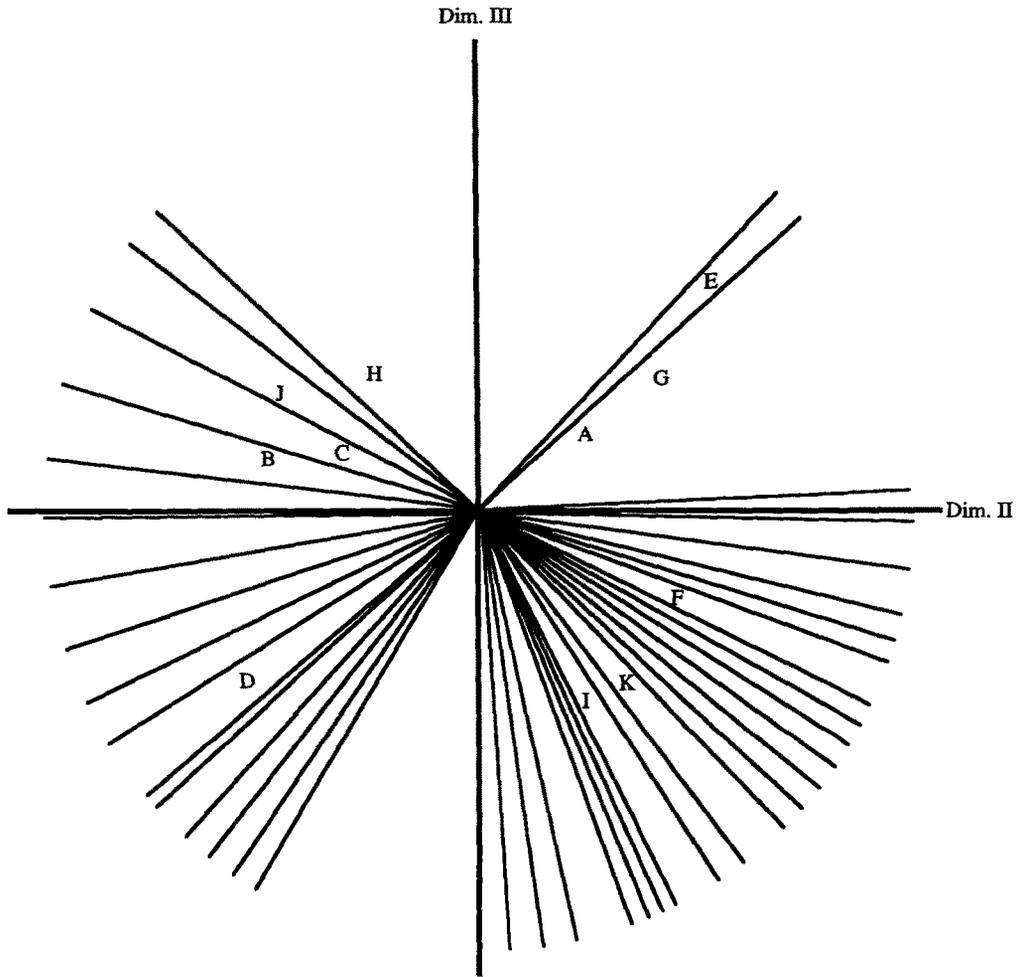
FIGURE 2
Joint space plot of Dimension 2 vs. 3 for the soft drink fixed-vector model.

(1555.0) that is smaller than that of the fixed ideal-point model (1640.6), it is the model selected here for spatially representing situational choice.

Results

Figures 1 and 2 present the two, two-dimensional joint spaces for the $T = 3$ dimensional solution of the fixed-vector model analysis. The lines in the plot represent the normalized vectors for subjects 1–40. These vectors are normalized to a constant length for convenience. The letters A–K represent the eleven brands of soft drinks. Note how the subjects' vectors are mostly oriented towards the left side of dimension one in Figure 1. This suggests that students have higher utility for the more popular (high market share) brands like Coke, Pepsi, 7-up, and Sprite. This homogeneity of utility does not hold for the Diet-NonDiet and Cola-NonCola dimensions where there is substantially more dispersion among the subjects' vector directions as shown in Figure 2.

The significance of the fixed analysis as compared to the floating analysis suggests that situations homogeneously affect student's choice behavior by affecting only the importance or the salience of the various dimensions in the various situations. As

TABLE 4

Importance Weights for Dimensions by Situation

Situation	Dimension: I	II	III
1	1.99	3.91	2.65
2	1.33	2.82	1.31
3	1.73	3.04	1.73
4	1.41	0.97	1.17
5	2.06	3.45	2.08

previously noted, these weights are common to all the subjects. The finding that a fixed-vector representation is the most appropriate for the soft drink choice data might result from the homogeneity of the sample (all are MBA students). Table 4 presents the weights of the three dimensions across the five usage occasions after normalizing X and Y to constant length by dimension. In Situation 1 (to quench your thirst on a very hot day), Dimension 2 (Diet-NonDiet) is the most important, suggesting a calorie concern for a situation where heavy consumption/usage typically occurs. The third dimension also carries a substantial weight, suggesting that students give moderate importance to taste in choosing a soft drink to quench their thirst. In Situation 2 (a social beverage you drink with friends) and in Situation 3 (a beverage you drink with meals), Dimension 2 (Diet-NonDiet) is also the most important. In Situation 4 (as a mixer), Dimension 1 (Popularity) is most important; Dimension 2 (Diet-NonDiet) is the least important. These results are quite intuitive given the set of brands we are considering.

To informally test the hypothesis that the importance of each dimension does not vary by situation, a fixed-vector model analysis with the same options as the one discussed above (i.e., $T = 3$, $\sigma_{ijk} = 1$, external brands' coordinates, and varying threshold by subjects), but with all the weights set equal to one was performed. The AIC of this analysis is 1681 which is *not* smaller than the AIC value for the fixed-vector model, and as such one can conclude that situational factors do appear to significantly affect the importance of the dimensions.

References

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

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, AC-19*, 716–723.

Anderson, 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.

Belk, R. W. (1974). An exploratory assessment of situational effects in buyer behavior. *Journal of Marketing Research, 11*, 156–63.

Belk, R. W. (1975). Situational variables and consumer behavior. *Journal of Consumer Research, 2*, 157–64.

Bishop, D. W., & Witt, P. A. (1970). Sources of behavioral variance during leisure time. *Journal of Personality and Social Psychology, 16*, 352–360.

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 criteria (AIC): The general theory and its analytical extensions. *Psychometrika, 52*, 345–370.

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., Pruzansky, S., & Kruskal, J. B. (1980). CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika, 45*, 3–24.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika, 40*, 5–32.

Coombs, C. H. (1964). *A theory of data*. New York: Wiley.

Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data* (2nd ed.). London: Chapman and Hall.

de Leeuw, J. (1973). *Canonical analysis of categorical data*. Unpublished doctoral dissertation, University of Leiden, The Netherlands.

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

DeSarbo, W. S. (1982). GENNCLUS: New models of general nonhierarchical clustering analysis. *Psychometrika, 47*, 449–475.

DeSarbo, W. S., & Carroll, J. D. (1985). Three-way metric unfolding via weighted alternating 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., & Cho, J. (1989). A stochastic multidimensional scaling vector model for the spatial representation of pick any/*N* data. *Psychometrika, 54*, 105–129.

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., Lehman, D., Gupta, S., Holbrook, M., & Havlena, W. (1987). A three-way unfolding methodology for asymmetric binary proximity data. *Applied Psychological Measurement, 11*, 397–418.

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 methodology for product positioning. *Marketing Science, 5*, 1–19.

De Soete, G., DeSarbo, W. S., & Carroll, J. D. (1985). Optimal variable weighting for hierarchical clustering: An alternating least-squares algorithm. *Journal of Classification, 2*, 173–192.

De Soete, G., DeSarbo, W. S., Furnas, G. W., & Carroll, J. D. (1984). The estimation of ultrametric and path length trees from rectangular proximity data. *Psychometrika, 49*, 289–310.

Dickson, P. R. (1982). Person-situation segmentation: Segmentation's missing link. *Journal of Marketing, 46*, 56–64.

Engel, J. F., Blackwell, R. D., & Kollat, D. T. (1969). *Consumer behavior*. New York: Holt, Rinehart and Winston.

Gifi, A. (1990). *Nonlinear multivariate analysis*. New York: Wiley.

Gill, P. E., Murray, W., & Wright, M. H. (1981). *Practical optimization*. Orlando, FL: Academic Press.

Green, P. E., & Rao, V. R. (1972). *Applied multidimensional scaling*. Hinsdale, IL: Dryden Press.

Greenacre, M. J. (1984). *Theory and applications of correspondence analysis*. London: Academic Press.

Heiser, W. J. (1981). *Unfolding analysis of proximity data*. Unpublished doctoral dissertation, University of Leiden, The Netherlands.

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

Hornik, J. (1982). Situational effects on the consumption of time. *Journal of Marketing, 46*, 44–55.

Kruskal, J. B., & Shepard, R. N. (1974). A nonmetric variety of linear factor analysis. *Psychometrika, 39*, 123–157.

Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive statistical analysis*. New York: 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.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

McFadden, D. (1976). Quantal choice analysis: A survey. *Annals of Economic and Social Measurement, 5*, 363–90.

Miller, K. E., & Ginter, J. L. (1979). An investigation of situational variation in brand choice behavior and attitude. *Journal of Marketing Research, 16*, 111–23.

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika, 43*, 551–560.

Nelder, J. 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.

Powell, M. J. D. (1977). Restart procedures for the conjugate gradient method. *Mathematical Programming, 12*, 241–254.

Sandell, R. G. (1968). Effects of attitudinal and situational factors on reported choice behavior. *Journal of Marketing Research, 4*, 405–408.

Sneath, P. H., & Sokal, R. R. (1973). *Numerical taxonomy*. San Francisco: W. H. Freeman.

Takane, Y. (1983, July). *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-èn-Josas, France.

Wilk, M. B., & Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika, 55*, 1–17.