A Stochastic Three-Way Unfolding Model for Asymmetric Binary Data

Wayne S. DeSarbo
University of Pennsylvania

Donald R. Lehmann and Morris B. Holbrook
Columbia University

William J. Havlena
Southern Methodist University

Sunil Gupta
University of California, Los Angeles

This paper presents a new stochastic three-way unfolding method designed to analyze asymmetric three-way, two-mode binary data. As in the metric three-way unfolding models presented by DeSarbo (1978) and by DeSarbo and Carroll (1980, 1981, 1985), this procedure estimates a joint space of row and column objects, as well as weights reflecting the third way of the array, such as individual differences. Unlike the traditional metric three-way unfolding model, this new methodology is based on stochastic assumptions using an underlying threshold model, generalizing the work of DeSarbo and Hoffman (1986) to three-way and asymmetric binary data. The literature concerning the spatial treatment of such binary data is reviewed. The nonlinear probit-like model is described, as well as the maximum likelihood algorithm used to estimate its parameter values. Results of a monte carlo study applying this new method to synthetic datasets are presented. The new method was also applied to real data from a study concerning word (emotion) associations in consumer behavior. Possibilities for future research and applications are discussed.

Three-way, two-mode asymmetric binary data are often collected in the social sciences. In word association data, perhaps the most common form of such data, individuals are presented a stimulus word and are asked to specify, in some manner, other words which are evoked by or related to the word presented. In a sense, such responses can be represented as "pick any" or "pick any/*n*" binary data (Coombs, 1964) in that an individual's responses can be coded *0* (a particular word is not evoked when a given stimulus/word is presented) or *1* (a particular word is evoked when a given stimulus/word is presented). Such responses may be characterized as "pick any" if there is completely free association and the universe of evoked responses (words) is not restricted or provided a priori. Similarly, these responses can be "pick any/*n*" when the complete or desired set of responses is restricted according to a designated list.

For example, Green and Tull (1978) collected free association data ("pick any" data) for shampoo benefits. They wished to examine the semantic associations elicited by various words or phrases related to the central benefit of "body," specifically whether the stimulus "body" (and other related terms) tended to connote "fullness" as opposed to "manageability." In contrast, using the "pick any/*n*" format,

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 11, No. 4, December 1987, pp. 397–418

0146-6216/87/040397-22$2.35

Havlena (1985) explored the relationship between emotions and consumption experiences. In one part of the study, each person was given the words for 28 emotions, one at a time; for each word, they were asked to select any of the remaining 27 emotions that he/she thought were related.

The focus of this paper is the problem of modeling "pick any/*n*" data for cases where the presented and evoked stimuli/words come from the same list. Note that such binary "pick any/*n*" association data collected over persons can be represented in a three-way, two-mode array (persons × words × words). Each two-way matrix for each person is square, but *asymmetric* in general. That is, the evocation of word B by the presentation of word A (A → B) does not necessarily imply the reverse (B → A). Extending this to a sociometric example, Person A may consider Person B as a friend, but the reverse may not hold. Also, note that the main diagonal of each component two-way matrix is either undefined or set equal to 1 in all cases.

One way to deal with such three-way, two-mode binary data is to "collapse" or aggregate over persons to form a two-way, one-mode matrix of counts of the number of persons in whom the presentation of stimulus i evoked stimulus j. The main diagonal elements are often set equal to the sample size. This type of square matrix of integer counts would be appropriate for the use of procedures such as correspondence analysis (Benzecri, 1969) and related procedures (see de Leeuw, 1973; Gifi, 1981a, 1981b; Nishisato, 1980); MDPREF or PREFMAP-Model 4 vector multidimensional scaling (MDS) preference models (Carroll, 1972); unfolding, using a program such as GENFOLD2 (DeSarbo & Rao, 1984, 1986); ASYMSCAL, a MDS model which accommodates asymmetric proximities (Young, 1975), and other procedures accommodating the inherent asymmetry in such data. Unfortunately, such pooling across persons prevents the detection of individual differences in responses. In addition, most two-way methods render solutions which typically suffer from one or more types of rotational indeterminacy (e.g., nonsingular transformations, orthogonal transformations), causing interpretation difficulties.

Another way of accommodating the analysis of such three-way, two-mode binary data is to split the data into a set of separate two-way, one-mode binary matrices per person to perform a separate analysis by person. Substantial work has been done in the area of spatially representing such two-way, one- or two-mode binary choice data. Guttman's (1944) scalogram analysis is a well-known procedure designed to order both row and column items with respect to some underlying cumulative dimension. Coombs' (1953) parallelogram technique is an early unidimensional scaling approach where the data, coded as *1*s and *0*s, are perfectly scaled; after permutation with respect to that scale, each row (or column) of the data table consists of a solid band of *1*s surrounded by *0*s. Unfortunately, these techniques were developed for unidimensional scaling, and their extension to MDS uses is unclear.

Torgerson (1958) developed deterministic and stochastic methods to analyze such two-way binary data spatially. Torgerson developed deterministic conjunctive and disjunctive models of choice and corresponding spatial methods that operate on such binary choice data. Torgerson also generalized Thurstone's (1929) method of similar reactions in obtaining a matrix of "directed distances," which is converted into scalar products and then factor analyzed.

Nonmetric MDS methods (e.g., Kruskal, 1964) estimate a joint space of row and column objects in which, for the unusual case of binary input data, *1*s would correspond to short distances between row and column object points while *0*s would correspond to long distances. Here, a massive number of "ties" would be encountered in the data, and it is not clear what effect this would have on resulting solutions. Similar concerns are held for nonmetric factor analysis (Kruskal & Shepard, 1974). Binary factor analysis (Christoffersson, 1975; Muthén, 1981) is a method which extends classical factor analysis to the analysis of binary data. Zinnes and Wolff (1977) developed a probabilistic multidimensional Thurstonian model for spatially representing the structure in one-mode, two-way binary data using a threshold model; this concept will be discussed below.

There also exist a number of interrelated MDS procedures which have been purposely devised to accommodate the analysis of two-way binary data. Correspondence analysis (Benzecri, 1969, 1973a, 1973b) typically analyzes aggregate two-way, two-mode choice data in the form of a frequency matrix (it can also handle the raw two-way binary data) and derives a joint space of row and column objects based on an eigenstructure analysis of a normalized frequency matrix. (See Greenacre, 1984, and Lebart, Morineau, & Warwick, 1984, for generalizations to multiple correspondence analysis.) Optimal scaling approaches, such as dual scaling (Nishisato, 1980) and homogeneity analysis (de Leeuw, 1973; Gifi, 1981a, 1981b; Heiser, 1981) attempt to estimate a set of best-fitting continuous weights to discrete categorical data values (many versions of these approaches have been shown to be equivalent to correspondence analyses). Finally, Levine (1979) has devised an eigenstructure-based technique to estimate a joint space to represent "pick-any" data. His scaling method represents persons as points whose coordinates are proportional to the centroids of the points representing their choices. Similarly, choice alternatives are represented as points whose coordinates are proportional to the centroids of the points representing persons who have selected them (see Green & DeSarbo, 1980, and Holbrook, Moore, & Winer, 1980, 1982, for applications of this method in marketing). Levine's (1979) "pick-any" methodology can also be viewed as a special case of correspondence analysis where a different type of normalization is employed.

Takane (1983) presented an item response model which considers "pick any/n" data as a special type of successive categories data in which there are only two response categories. In an analysis of this type, each item (stimulus), rather than each category of an item, may be represented as a point, and persons are assumed to select (or not to select) the item according to its closeness to their respective ideal points.

DeSarbo and Hoffman (1986) recently used a threshold concept to develop simple and weighted unfolding models for the analysis of "pick any/n" binary choice data. In their model, an object is selected by a person if that object falls within a threshold distance from the person's ideal point. These authors also provided the capability of restricting configurations to be functions of specified person and/or stimulus background variables (e.g., demographics, psychographics, features, price, etc.).

The methods described above are designed to analyze two-way, one- or two-mode binary data (e.g., two-mode for binary choice data and one-mode for asymmetric binary proximities). Analyzing each slice of the original three-way array individually by such methods may be computationally infeasible for a large number of persons. It would also result in different configurations for different persons, which creates a problem in comparing results across the sample. Obviously, there is no guarantee that the dimensions will be the same for each person, or that there will even be the same number of dimensions per person. Finally, all of these two-way methods suffer from various types of solution (rotation) indeterminacies which make interpretation difficult.

In this paper, the DeSarbo and Hoffman (1986) approach is generalized to the analysis of three-way, two-mode "pick any/n" binary data. As far as is known, no appropriate procedure currently exists to analyze such three-way, two-mode data. DeSarbo (1978) and DeSarbo and Carroll (1980, 1981, 1985) used their three-way, metric unfolding procedure to analyze asymmetric proximities, but only where these proximities are metric and not binary. Other three-way procedures for three-way, two-mode data, such as INDSCAL (Carroll & Chang, 1970), IDIOSCAL (Carroll & Chang, 1972), Three Mode Scaling (Tucker, 1972), PARAFAC 2 (Harshman, 1972), MULTISCALE (Ramsay, 1977), PINDIS (Lingoes & Borg, 1978), and ALSCAL (Takane, Young, & de Leeuw, 1977), also assume symmetric and metric data (though ALSCAL does provide for a nonmetric fitting of the INDSCAL model). CANDECOMP (Carroll & Chang, 1970), PARAFAC (Harshman, 1970), DEDICOM (Harshman, 1975), and SMACOF (de Leeuw & Heiser, 1980) are procedures that can accommodate the asymmetry, but most are metric procedures. SMACOF accommodates nonmetric analyses, but it is unclear how the *0*s and *1*s and the resulting massive number of ties would affect the solutions obtained.

The Stochastic Three-Way Unfolding Model for Asymmetric Binary Data

Let: $i,j = 1, \ldots, N$ stimuli,

$k = 1, \ldots, M$ persons,

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

$y_{ijk} = \begin{cases} 1 \text{ if presentation of stimulus } i \text{ evokes stimulus } j \text{ in person } k, \\ 0 \text{ otherwise,} \end{cases}$

w_{kt} = the importance or salience of dimension t for person k,

x_{it} = the tth coordinate for the ith presented stimulus,

z_{jt} = the tth coordinate for the jth evoked stimulus, and

α_k = an additive constant for person k.

Now, define a latent unobservable response variable:

$$y^*_{ijk} = f(\mathbb{W},\mathbb{X},\mathbb{Z}) + \mu_{ijk} \quad , \tag{1}$$

where

$$f(\mathbb{W},\mathbb{X},\mathbb{Z}) = \sum_{t=1}^{T} w_{kt}(x_{it} - z_{jt})^2 + \alpha_k \quad , \tag{2}$$

μ_{ijk} = error $\sim N(0,\sigma^2_{ijk})$, $\mathbb{W} = \|w_{kt}\|$, $\mathbb{X} = \|x_{it}\|$, and $\mathbb{Z} = \|z_{jt}\|$, such that $y_{ijk} = 1$ if $y^*_{ijk} \leqslant c_k$, or $y_{ijk} = 0$ if $y^*_{ijk} > c_k$, where c_k is some threshold value indexed by person. This formalization implies that presentation of stimulus i will evoke stimulus j in person k if the coordinate location for stimulus i is close enough to the coordinate location for stimulus j. Thus, each stimulus, assuming the row and column elements of $\mathbb{Y} = \|y_{ijk}\|$ are the same, is represented by two sets of points. One set of points represents the presented stimulus as a row element, and the other set of points represents the evoked stimulus as a column element. This is how the asymmetry in $\mathbb{Y}$ is represented.

Thus, the stochastic three-way unfolding model for asymmetric binary data assumes that some presented stimulus will tend to evoke a given response from a particular person if that response lies within a tolerance distance from it in a multidimensional space. The model assumes a common space for all persons, but allows for different dimension weights and threshold distances that vary across persons.

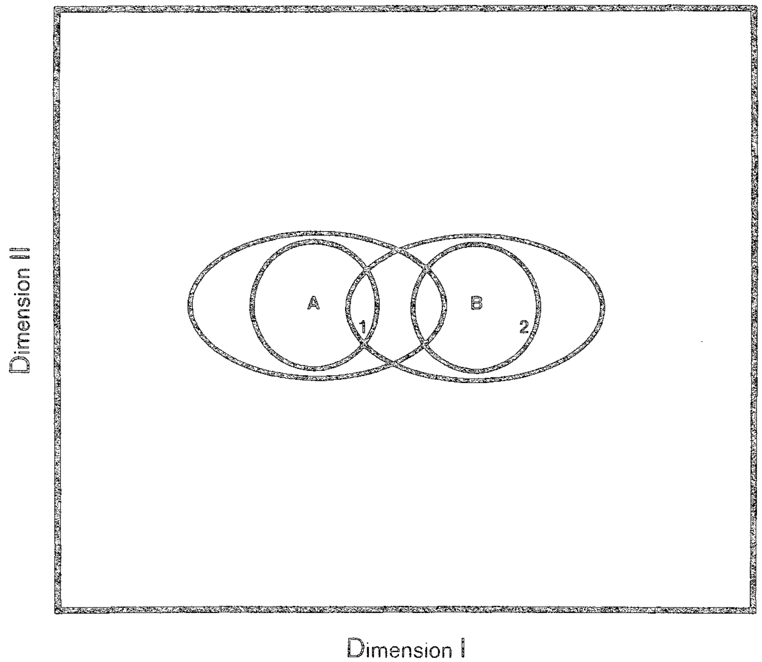
This model may be illustrated graphically by the hypothetical example shown in Figure 1. A and B represent two row stimuli, while 1 and 2 represent two column stimuli. The two circular regions are for Person I, while the two elliptical regions are for Person II. Here, Person I weighs the two dimensions equally with a relatively small tolerance region (i.e., short distance thresholds on both axes) represented by the two circular areas around (row) stimuli A and B. By contrast, Person II adopts a generally larger tolerance region (i.e., larger distance thresholds) represented by the two elliptical areas around (row) stimuli A and B, with greater importance assigned to the vertical dimension. For Person I, stimulus A evokes (column) response 1 but not 2, and stimulus B evokes 2 but not 1. By contrast, for Person II, stimulus A evokes only 1, but stimulus B evokes both 1 and 2. Hence, individual threshold differences are represented by tolerance regions of varying sizes and shapes, while asymmetries are reflected by the presence of two points (a presented stimulus and an evoked response) for each object. In the special case of symmetry, these two points would coincide.

Thus,

$$P(y_{ijk} = 1) = P(y^*_{ijk} \leqslant c_k) = P[f(\mathbb{W},\mathbb{X},\mathbb{Z}) + \mu_{ijk} \leqslant c_k]$$

$$= P[\mu_{ijk} \leqslant -f^*(\mathbb{W},\mathbb{X},\mathbb{Z})] = \Phi\left[\frac{-f^*(\mathbb{W},\mathbb{X},\mathbb{Z})}{\sigma_{ijk}}\right] \quad , \tag{3}$$

Figure 1
A Hypothetical Example of the Threshold Concept



where $\Phi(\cdot)$ is the standard normal distribution function (as in a probit response model) and

$$f^*(\mathbb{W},\mathbb{X},\mathbb{Z}) = \sum_t w_{kt}(x_{it}-z_{jt})^2 + \alpha_k - c_k = f(\mathbb{W},\mathbb{X},\mathbb{Z}) - c_k$$

$$= \sum_t w_{kt}(x_{it}-z_{jt})^2 + \gamma_k \quad , \tag{4}$$

with $\gamma_k = \alpha_k - c_k$. Similarly,

$$P(y_{ijk} = 0) = P(y^*_{ijk} > c_k) = 1 - \Phi\left[\frac{-f^*(\mathbb{W},\mathbb{X},\mathbb{Z})}{\sigma_{ijk}}\right] \quad . \tag{5}$$

It is assumed that the observed values of y_{ijk} are realizations of an independent binomial process with probabilities given by Equations 3 and 5. This assumption is particularly appropriate for "pick any/n" data where selection or nonselection of any item has no structural effect concerning the selection or nonselection of any other(s). Thus, assuming independence over all i, j, and k subscripts, the likelihood function can be formulated:

$$\mathrm{L} = \prod_{y_{ijk}=1} \Phi(\cdot) \prod_{y_{ijk}=0} [1-\Phi(\cdot)] \quad , \tag{6}$$

where

$$\Phi(\cdot) = \Phi\left[\frac{-f^*(\mathbb{W},\mathbb{X},\mathbb{Z})}{\sigma_{ijk}}\right] \quad . \tag{7}$$

Substitution in Equation 6 yields

$$\mathrm{L} = \prod_{k=1}^{M} \prod_{i=1}^{N} \prod_{j \neq i}^{N} [P(y_{ijk}=1)]^{y_{ijk}} [P(y_{ijk}=0)]^{(1-y_{ijk})} \quad . \tag{8}$$

Converting to logs, the expression becomes

$$R = \ln \mathrm{L} = \sum_{k=1}^{M} \sum_{i \neq j}^{N}\sum \{y_{ijk} \ln \Phi(\cdot) + (1-y_{ijk}) \ln [1-\Phi(\cdot)]\} \quad . \tag{9}$$

The procedure attempts to estimate $\mathbb{W}$, $\mathbb{X}$, $\mathbb{Z}$, γ, and $\sigma = \|\sigma_{ijk}\|$, given $\mathbb{Y}$ and T, in order to maximize the log likelihood function in Equation 9.

At this point, it is important to distinguish this model structure from those proposed by DeSarbo and Hoffman (1986) and Takane (1983), in that it can be thought of as a generalization of each of these two alternative models. DeSarbo and Hoffman proposed a multidimensional unfolding threshold methodology for the spatial representation of two-way binary data. They defined a latent, unobservable "disutility" variable D_{kj} such that

$$D_{kj} = \sum_{t=1}^{T} w_{kt}(x_{kt}-z_{jt})^2 + c_k + \mu_{kj} \quad , \tag{10}$$

where μ_{kj} is a stochastic error component and t, k, and j indicate dimensions, persons, and stimuli, respectively. Equation 10 states that D_{kj}, person k's latent disutility for stimulus j, can be represented by a weighted unfolding model (Carroll, 1972) involving person k's ideal point (x_{kt}), stimulus j's coordinates (z_{jt}), the person's importance weights for each dimension (w_{kt}), and an additive constant (c_k). If $D_{kj} \leq d_k^*$ (some individual threshold value), then a choice is made for stimulus j, that is, $y_{kj} = 1$. If $D_{kj} > d_k^*$, then no choice is made for stimulus j and $y_{kj} = 0$. Then,

$$\begin{aligned} P(y_{kj} = 1) &= P(D_{kj} \leq d_k^*) = P\left[\sum_t w_{kt}(x_{kt}-z_{jt})^2 + c_k + \mu_{kj} \leq d_k^*\right] \\ &= P\left[\mu_{kj} \leq \sum_t w_{kt}(x_{kt}-z_{jt})^2 - c_k^*\right] = P(\mu_{kj} \leq -f_{kj}) \quad , \end{aligned} \tag{11}$$

where $c_k^* = c_k - d_k^*$ and

$$f_{kj} = \sum_t w_{kt}(x_{kt}-z_{jt})^2 + c_k^* \quad . \tag{12}$$

Similarly,

$$P(y_{kj} = 0) = P(D_{kj} > d_k^*) = 1 - P(\mu_{kj} \leq -f_{kj}) \quad . \tag{13}$$

Assuming $\mu_{kj} = 0$, then c_k^* is an estimate of person k's (negative) threshold value. The size of the negative of the estimated additive constant is the *magnitude* of the choice threshold, indicating the sensitivity of a person's choice process.

The general form of the likelihood function here is

$$\mathrm{L} = \prod_{y_{kj}=0} P(y_{kj} = 0) \prod_{y_{kj}=1} P(y_{kj} = 1) \quad . \tag{14}$$

Assuming that $y_{kj} \sim$ binomial $(1, p_{kj})$ with independence across persons and stimuli, the process of person i selecting stimulus j is an independent "coin toss" with probability of choice given by p_{kj}. DeSarbo and Hoffman (1986) assumed that μ_{kj} has a logistic distribution function. Then

$$P(y_{kj} = 1) = \frac{1}{1+\exp(f_{kj})} = p_{kj} \tag{15}$$

and

$$P(y_{kj} = 0) = \frac{\exp(f_{kj})}{1+\exp(f_{kj})} = 1-p_{kj} \quad , \tag{16}$$

so that their spatial model can be expressed as

$$\ln\left[\frac{1-p_{kj}}{p_{kj}}\right] = \sum_t w_{kt}(x_{kt}-z_{jt})^2 + c_k^* \quad . \tag{17}$$

This is a logistic function, where the proximity of a product to a respondent's ideal point indicates some degree of the magnitude of the probability of choice. Thus, the DeSarbo and Hoffman approach is a logit type model which deals with two-way choice data.

Takane (1983) developed an item response model for the multidimensional analysis of unordered categorical data, which typically arise in multiple-choice questionnaire surveys. Item categories and persons are each represented in a multidimensional Euclidean space. The probability of a particular person selecting a particular item category is modeled as a decreasing function of the distance between them. Suppose a group of persons have responded to a set of I items, each having J_i $(i=1, \ldots, I)$ response categories (options). The persons may be classified by distinct response patterns which are indexed by k. Define

$$g^*_{ki(j)} = \begin{cases} 1 \text{ if option } j \text{ of item } i \text{ is chosen in response pattern } k, \\ 0 \text{ otherwise,} \end{cases}$$

and let f_k denote the observed frequency of response pattern k. Takane (1983) assumed that persons (corresponding to each distinct response pattern) and item categories are both represented as points in a multidimensional Euclidean space. The distance between response pattern k and option j of item i is given by

$$d_{ki(j)} = \left[\sum_t (z_{kt} - x_{i(j)t})^2\right]^{1/2} \quad , \tag{18}$$

where z_{kt} is the coordinate of response pattern k on dimension t and $x_{i(j)t}$ is the coordinate of option j of item i on dimension t. Takane's model states that

$$p_{ki(j)} = \frac{\exp(-d^v_{ki(j)})}{\sum_{j'} \exp(-d^v_{ki(j')})} \quad , \tag{19}$$

where $p_{ki(j)}$ is the probability that the person with response pattern k chooses option j of item i. The model postulates that each option has "response strength," $\exp(-d^v_{ki(j)})$, which is a decreasing function of $d_{ki(j)}$. A particular option is chosen with probability proportional to its response strength relative to those of the other options within the same item. Takane (1983) developed a marginal maximum likelihood estimation procedure using an EM algorithm. Like the DeSarbo and Hoffman (1986) model, the Takane model is also restricted to two-way data and uses a logit-type model.

Program Options

The model presented in Equation 2 can be viewed as a special type of DeSarbo's (1978) and DeSarbo and Carroll's (1980, 1981, 1985) three-way unfolding model applied in a stochastic context (i.e., an error term with a posited distribution) to binary data analysis. As such, there is some discussion in the psychometric literature concerning the interpretation of $\mathrm{W} = \|w_{kt}\|$, especially when found to be negative (see Carroll, 1972; DeSarbo & Carroll, 1985). There appears to be substantial controversy in the literature over the desirability of constraining w_{kt}s in the two-way weighted (and general) unfolding model to be positive. Carroll (1972) claimed that in the two-way weighted unfolding model, a negative w_{kt} can have a meaningful interpretation. Specifically, if w_{kt} is negative with respect to dimension t, the ideal point for person k indicates the least preferred rather than the most preferred value, and the farther a stimulus is along that dimension from the ideal point, the more highly preferred is that stimulus. Carroll argued that such a minimally preferred value characterized many persons in the case of the temperature dimension for tea (i.e., many persons liked hot and cold but not lukewarm tea). Other authors, such as Srinivasan and Shocker (1973) and Davison (1976), disputed the value of unconstrained analyses, claiming that the existence of such negative weights may lead to unrealistic predictions for persons' most preferred stimuli. In the present methodology, the user has the option of estimating such weights freely or constraining them to be non-negative.

Another option concerns the estimation of σ^2_{ijk}. It is clear that degrees of freedom are quickly depleted if all σ^2_{ijk} are to be estimated (together with $\mathbb{W}$, $\mathbb{X}$, $\mathbb{Z}$, and γ). Estimation options therefore exist to set $\sigma^2_{ijk} = 1$, for all i, j, k, or to estimate σ^2_k as a variance term which varies by person (for cases where γ is constrained to equal 0). The algorithm used to estimate these sets of parameters is described in the Appendix.

The final option entails constraining γ in Equation 4 to be nonpositive. This additive constant can be interpreted as a "threshold coefficient" for each k. Assuming $\mu_{ijk} = 0$ in Equation 3 allows the following expression:

$$P(y_{ijk} = 1) = P(y^*_{ijk} \leq c_k) = P[f^*(\mathbb{W},\mathbb{X},\mathbb{Z}) \leq 0] = P\left[\sum_t w_{kt}(x_{jt} - z_{it})^2 \leq -\gamma_k\right] \quad . \tag{20}$$

Thus, $y_{ijk} = 1$ if the weighted distance between stimuli i and j for person k is less than $-\gamma_k$. If the w_{kt}s are constrained to be non-negative, then the γ_ks should be ≤ 0 because negative Euclidean distances are impossible. Note that these threshold coefficients can be interpreted in terms of "iso-indicator" contours which can be constructed around the x_{it}s. In the simple case with $w_{kt} = 1$, for all k, t, circular (two-dimensional) iso-indicator contours can be constructed around each x_{it} with radius γ_k. Any adjective j within this iso-indicator contour would be predicted to be evoked by the model when i is presented. A related option is where the user can set $\gamma = 0$.

Monte Carlo Results

The estimation procedure was initially tested on a number of small synthetic datasets. In most cases, the procedure recovered the known configurations with larger likelihood values than were produced with the original parameters. However, in some cases, the algorithm produced locally optimum solutions with worse likelihood values and different parameter estimates; note the problem concerning the recovery of $\mathbb{W}$ and γ when estimating σ_k given the scalar indeterminacy between $\mathbb{W}_k$, σ_k, and γ_k. In order to investigate this aspect in a more rigorous manner, a monte carlo analysis was undertaken in which seven independent data and model factors were experimentally manipulated. Table 1 presents the seven independent factors and their levels and codes.

A fractional factorial design (assuming a main-effects-only model) was used to combine these seven factors into 12 trials, as shown in Table 2. This allows for the estimation of independent main effects (interaction effects are assumed to be 0). The dependent measures used were (1) the simple matching coefficient measuring data recoverability; (2) the average sums of squares between $\mathbb{X}$ and $\hat{\mathbb{X}}$; (3) the average sums of squares between $\mathbb{Z}$ and $\hat{\mathbb{Z}}$; and (4) the average sums of squares between $\mathbb{W}$ and $\hat{\mathbb{W}}$.

Table 1
Seven Independent Factors in the Monte Carlo Analysis

Factor	Levels	Code
A. Number of Dimensions (T)	1	0
	3	1
B. Number of Stimuli (N)	7	0
	12	1
C. Number of Slices (M)	2	0
	4	1
D. Weights	unconstrained	0
	non-negative	1
E. Threshold Coefficients	unconstrained	0
	nonpositive	1
F. Start	random	0
	CANDECOMP	1
G. Sigma	$\sigma = 1$	0
	σ_k	1

Measures 2, 3, and 4 were calculated after the actual and estimated configurations were normalized and permuted to congruence, according to the DeSarbo and Carroll (1985) procedure. The configurations for each trial were generated randomly from a uniform distribution according to the experimental profile. $\mathbb{Y}$ was created after adding $N(0,\sigma_k)$ error.

Table 3 presents the regression results of these monte carlo runs. The table shows that data recovery using the simple matching coefficient is significantly higher for $T = 3$ and a CANDECOMP start. The first result may be due to the well-known instability of one-dimensional unfolding solutions. The second finding underscores the need for a rational starting configuration and potential problems with local optima with random starts. A surprising finding was that somewhat poorer recovery was associated with a larger number of person slices. For the configuration recovery measures 2, 3, and 4, parameter recovery was systematically affected to only a small degree. The $T = 3$ factor appeared to lead to slightly better $\mathbb{Z}$ recovery. None of these regressions is significant. Note, however, the coefficient magnitudes for dependent

Table 2
2^7 Fractional Factorial Design

	Factor						
Trial	A	B	C	D	E	F	G
1	0	0	0	0	0	0	0
2	1	1	0	1	1	1	0
3	0	1	1	0	1	1	1
4	1	0	1	1	0	1	1
5	0	1	0	1	1	0	1
6	0	0	1	0	1	1	0
7	0	0	0	1	0	1	1
8	1	0	0	0	1	0	1
9	1	1	0	0	0	1	0
10	1	1	1	0	0	0	1
11	0	1	1	1	0	0	0
12	1	0	1	1	1	0	0

Table 3
Regression Results for Monte Carlo Analyses

Factor	Dependent Measure 1	2	3	4
T = 3	.201***	-.015	-.058*	-3.103
N = 12	-.008	-.065	-.002	14.245
M = 4	-.065**	-.039	.007	.601
$\tilde{W}$ Constrained	-.024	.027	.010	-2.088
$\tilde{\gamma}$ Constrained	-.025	-.013	.011	-12.433
CANDECOMP start	.075**	.094	.011	1.767
σ_k	-.010	.111	.050	-8.170
S. E.	.031	.099	.042	18.373
R^2	.975	.681	.723	.495
Adjusted R^2	.933	.121	.237	0.0
F	22.74***	1.22	1.49	.561

*$p \leq .10$; **$p \leq .05$; ***$p \leq .01$

measure 4, where the larger coefficients for W recovery indicate that, overall, substantially worse average sums of squares measures were encountered for these weight configurations. In many cases, this may have been caused by the scalar indeterminacy of σ_k, γ_k, and W_k.

Clearly, this is not a definitive monte carlo analysis. More factors and replications for each trial would certainly have been desirable. Similarly, a full factorial design with 128 trials might be more revealing. The intention here was to illustrate the data and parameter recovery of the procedure for a simple experimental design.

An Illustration With Real Data

The Categorization and Description of Emotion

Researchers attempting to develop schemes for emotional classification and description have, in general, followed one of two major approaches. One approach is compositional, in that it views all emotions as stemming from a relatively small number of basic emotional categories (Darwin, 1872; Izard, 1977; Plutchik, 1980). The second approach, as pursued here, is decompositional and views emotions in terms of continuous dimensions that are not themselves emotions, but that can be used to describe them (Mehrabian & Russell, 1974; Osgood, Suci, & Tannenbaum, 1957; Spencer, 1855; Wundt, 1902).

Previous research on both paradigms has used MDS and other multivariate methods to examine emotions. Work in the area of basic emotional categories tends to use empirical data to test these theoretical models of emotion. For example, Conte (1975) used MDS to test Plutchik's hypothesized emotion circumplex. In contrast, researchers examining emotions from the dimensional viewpoint have used factor analytic methods to uncover basic dimensions through exploratory data reduction. For example, Smith and Ellsworth (1985) used principal components analysis and SINDSCAL to recover six dimensions from 18 items measuring emotional responses.

Study Description

The data are from a larger study of emotions in various consumption experiences (Havlena, 1985; Havlena & Holbrook, 1985). Specifically, the responses of (a small subset of) 63 persons to 28 emotion descriptors (drawn from experience descriptions) were reanalyzed. For each descriptor, persons indicated which of the other 27 were likely to occur simultaneously. The resulting aggregate proximity matrix appears in Table 4, whose entries represent the number of times the row item led to a check of the column item. Havlena (1985) and Havlena, Holbrook, and Lehmann (1986) applied ALSCAL to a symmetrized version of this aggregate matrix and found two major dimensions: evaluation and activity. However, this analysis ignored individual differences and assumed symmetry. Different findings might have resulted if these restrictive assumptions had been relaxed.

Analysis

A cluster analysis performed on simple matching coefficients (which, according to Everitt, 1980, is the most commonly used similarity measure for binary data) calculated on $\mathbb{Y}$ for all 63 persons suggested three clusters of respondents. For purposes of illustrating this method, the three persons closest to their respective clusters' centroids were selected for analysis.

The analysis was conducted with $\sigma_{ijk} = 1$, non-negativity constraints on $\mathbb{W}$, and nonpositivity constraints on γ for $T = 1, 2, 3,$ and 4. Table 5 presents the statistical results for the one- to four-dimensional solutions run for this dataset. The asymptotic χ^2 test of the difference of deviance statistics supports the three-dimensional solution as the most parsimonious for this subsample. There appear to be small differences in the corresponding simple matching coefficients (of $\mathbb{Y}$ and $\hat{\mathbb{Y}}$; see Appendix) for these solutions, although the largest jump occurs in going from two to three dimensions. Table 6 presents the simple matching coefficient values by row and by column emotion. Table 6 depicts fairly consistent fitting over all "dimensions" of the three-way array ($\mathbb{Y}$). The associated matching coefficients for Persons 1, 2, and 3 were .904, .862, and .776 respectively. Person 3 and emotion 11 (Excited) do not appear to be fit as well by the model as are the other persons and emotions.

Table 7 presents the three-dimensional solution values for $\mathbb{X}$, $\mathbb{Z}$, $\mathbb{W}$, and γ, as well as the plotting code for each row and column emotion. The pattern of coordinates for the 28 emotions appears similar for the same stimulus ($\mathbb{X}$) and response ($\mathbb{Z}$) emotions. This is reflected in the correlation matrix between the coordinates of $\mathbb{X}$ and those of $\mathbb{Z}$ presented in Table 8. Here, $r(X_1,Z_1) = .914$, $r(X_2,Z_2) = .868$, and $r(X_3,Z_3) = .916$.

Figure 2 presents the one-dimensional joint space representation for each of the three dimensions. The first dimension (Figure 2a) appears to represent "contentedness" or "pleasure" (peacefulness and satisfaction as opposed to sadness or anger), with emotions such as "sad", "in pain", "disappointed", and "angry" loading on the positive end and emotions such as "peaceful", "relaxed", "relieved", and "satisfied" loading on the negative end.

The second dimension (Figure 2b) appears to represent Osgood et al.'s (1957) "potency" or the opposite of Mehrabian and Russell's (1974) "dominance", with the positive end of the scale relatively submissive ("relaxed", "relieved", "sad") and the negative end more assertive ("enthusiastic", "excited", "energetic", "exhilarated"). The third dimension (Figure 2c) separates "peaceful", "relaxed", "relieved", "satisfied" from "anticipatory", "energetic", "fearful", "excited", "horrified" and therefore appears to represent "activation" or "arousal". Taken together, these three dimensions show reasonable consistency with both the Osgood et al. (1957) and Mehrabian and Russell (1974) frameworks.

In addition to the results presented paralleling the Mehrabian and Russell (1974) PAD (pleasure-arousal-dominance) scheme and therefore possessing some face validity, the proposed model appears to give a fairly coherent account of the asymmetric relations between presented and evoked emotion descriptors. Figure 2 provides vivid representations of the asymmetries along each dimension. Generally, when the same emotion fails to appear on the same pole of the same stimulus and response dimension, asymmetric relations are present. For example: According to Figure 2a, "annoyed" → "crying", but

Table 4
Adjective Similarity Matrix

Adjective	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1. Absorbed	63	3	2	9	13	0	1	20	17	18	8	10	1	2	2	5	0	54	1	1	4	7	1	10	1	1	13	1
2. Angry	3	63	50	3	4	13	25	3	0	0	19	2	4	48	4	0	1	1	0	12	0	0	0	0	0	11	0	0
3. Annoyed	2	44	63	3	6	1	34	0	0	0	1	0	3	44	1	0	0	0	0	2	0	0	0	0	0	4	0	1
4. Anticipatory	8	0	2	63	32	0	0	20	3	25	48	15	11	4	0	5	1	30	2	1	0	4	1	0	0	0	0	2
5. Aware	16	0	1	15	63	0	2	12	7	9	8	5	3	2	2	11	2	46	2	0	5	6	2	6	5	2	10	1
6. Crying	4	25	9	0	5	63	35	0	1	0	1	1	17	29	45	10	7	0	3	39	1	1	1	1	3	51	0	4
7. Disappointed	3	20	39	1	7	9	63	0	0	0	1	0	3	44	4	0	0	0	0	3	0	0	0	0	2	41	0	4
8. Energetic	11	1	0	15	16	0	0	63	25	57	43	47	0	0	0	26	0	29	12	0	1	17	5	3	1	0	15	0
9. Enjoying	22	0	0	2	16	0	0	22	63	34	26	17	0	0	0	51	0	29	28	0	22	40	6	28	2	0	37	2
10. Enthusiastic	21	0	0	14	11	0	0	52	30	63	50	39	0	0	0	30	0	36	21	0	3	21	5	4	2	0	12	0
11. Excited	11	9	4	22	18	1	3	43	21	45	63	38	5	2	2	29	4	21	19	2	0	13	2	1	2	2	7	7
12. Exhilarated	9	0	0	9	13	0	0	47	26	44	49	63	0	0	0	29	0	15	21	0	2	17	6	3	4	0	14	8
13. Fearful	9	8	11	27	21	5	4	3	0	0	14	1	63	14	5	0	29	2	0	15	0	0	0	0	0	11	0	8
14. Frustrated	9	40	55	9	8	10	40	2	0	0	4	1	11	63	8	0	1	2	0	9	0	0	0	0	0	22	0	1
15. Grief-stricken	13	13	3	2	5	42	19	1	0	0	0	0	17	17	63	0	10	0	0	28	0	0	0	0	0	56	0	0
16. Happy	12	0	0	10	17	4	0	32	49	33	32	30	0	0	0	63	0	20	48	0	24	48	17	26	9	0	42	6
17. Horrified	8	14	10	11	14	8	9	4	0	0	10	2	44	8	8	0	63	1	0	6	0	0	0	0	0	14	0	21
18. Interested	53	0	0	20	41	0	0	22	23	31	16	11	1	1	0	11	1	63	5	0	6	16	3	10	3	0	14	2
19. Joyful	5	0	0	4	11	1	0	24	42	27	29	29	1	0	0	57	0	10	63	0	15	41	7	13	6	0	27	5
20. In pain	14	25	29	7	10	31	14	0	0	0	2	0	25	27	15	1	5	0	1	63	0	0	1	1	0	30	0	1
21. Peaceful	8	0	0	0	12	0	0	1	24	2	1	2	0	0	0	40	0	6	14	0	63	34	4	54	17	0	43	0
22. Pleased	3	0	0	1	11	0	0	3	31	9	11	7	0	0	0	52	0	10	21	0	29	63	18	21	16	0	52	2
23. Proud	2	0	0	2	10	1	0	8	22	20	19	13	0	0	0	48	0	10	26	0	6	47	63	6	4	0	43	2
24. Relaxed	6	0	0	0	4	0	0	1	27	0	0	1	0	0	0	29	0	5	6	0	54	23	1	63	17	0	45	0
25. Relieved	0	0	0	1	7	4	0	2	8	4	3	3	0	0	0	32	0	3	9	0	21	29	2	36	63	0	29	2
26. Sad	6	11	12	1	7	34	44	0	0	0	1	0	10	34	25	0	2	1	0	26	1	0	0	1	1	63	0	0
27. Satisfied	6	0	0	1	10	0	0	4	25	8	3	4	0	0	0	40	0	11	9	0	38	46	14	43	21	0	63	1
28. Surprised	6	7	10	12	16	2	5	10	10	12	38	9	9	1	0	20	4	16	6	2	2	7	2	2	1	0	3	63

Table 5
Analyses for the Emotions Data

T	Model df	Matching Coefficient	-ln L	Deviance	Deviance Difference
1	32	.825	879.54	1759.08	---
2	61	.828	825.12	1650.24	108.84***
3	90	.847	791.15	1582.30	67.94***
4	119	.853	778.69	1556.38	25.94

***$p \leq .01$

"crying" $\nrightarrow$ "annoyed"; according to Figure 2b, "excited" $\rightarrow$ "energetic", but "energetic" $\nrightarrow$ "excited"; and according to Figure 2c, "exhilarated" $\rightarrow$ "absorbed", but "absorbed" $\nrightarrow$ "exhilarated". The three-dimensional joint space (not shown) could also be plotted to examine the proximity of the two sets of points while considering all dimensions simultaneously. Thus, the spatial representation of asym-

Table 6
Matching Coefficients by Row and Column
for Emotions Data

Emotion	Matching Coefficient: Row	Matching Coefficient: Column
Absorbed	.762	.810
Angry	.857	.869
Annoyed	.857	.833
Anticipatory	.833	.869
Aware	.845	.845
Crying	.845	.845
Disappointed	.881	.893
Energetic	.762	.810
Enjoying	.798	.881
Enthusiastic	.750	.833
Excited	.679	.619
Exhilarated	.810	.821
Fearful	.833	.786
Frustrated	.893	.893
Grief-stricken	.869	.869
Happy	.798	.792
Horrified	.857	.905
Interested	.833	.857
Joyful	.809	.833
In pain	.929	.857
Peaceful	.929	.905
Pleased	.881	.833
Proud	.833	.845
Relaxed	.929	.917
Relieved	.952	.905
Sad	.905	.929
Satisfied	.941	.869
Surprised	.857	.833

Table 7
Row and Column Coordinates, Person Weights, Estimates of Sigma, and Threshold Parameter Estimates for the Three-Dimensional Solution for Emotions Data

Emotion	Row Coordinates 1	Row Coordinates 2	Row Coordinates 3	Column Coordinates 1	Column Coordinates 2	Column Coordinates 3
Absorbed	-.121	-.130	-.023	-.075	-.093	-.163
Angry	.321	.164	-.142	.217	.185	-.153
Annoyed	.279	.159	-.144	.094	.190	-.143
Anticipatory	-.136	-.069	-.166	-.130	-.266	-.214
Aware	-.105	-.192	-.032	-.159	-.124	-.132
Crying	.235	.171	-.134	.262	.244	-.138
Disappointed	.339	.182	-.130	.240	.186	-.150
Energetic	-.058	-.326	-.188	-.133	-.311	-.072
Enjoying	-.103	-.106	.182	-.146	-.249	.230
Enthusiastic	-.075	-.316	-.124	-.136	-.274	.076
Excited	.022	-.349	-.168	.134	-.039	-.165
Exhilarated	-.028	-.340	-.158	-.135	-.261	-.048
Fearful	-.060	.117	-.172	-.108	.103	-.157
Frustrated	.273	.160	-.143	.302	.233	-.136
Grief-stricken	.239	.174	-.133	.262	.226	-.135
Happy	-.108	-.168	.228	-.145	-.255	.264
Horrified	-.070	.136	-.155	.148	.121	-.174
Interested	-.107	-.245	-.039	-.160	-.140	-.143
Joyful	-.124	-.061	.217	-.146	-.209	.235
In pain	.307	.183	-.128	.358	.240	-.139
Peaceful	-.171	.168	.336	-.179	.101	.399
Pleased	-.123	-.071	.175	-.163	-.063	.263
Proud	-.150	-.007	.098	-.161	-.079	.103
Relaxed	-.173	.184	.391	-.178	.178	.292
Relieved	-.179	.219	.278	-.176	.106	.207
Sad	.379	.189	-.126	.302	.236	-.132
Satisfied	-.159	.071	.359	-.179	.144	.295
Surprised	-.128	.052	-.147	-.147	-.098	-.065
Person Weights						
Person 1	5.190	4.455	7.622			
Person 2	8.141	10.669	4.495			
Person 3	12.306	2.331	4.856			
Estimates of Sigma						
$\sigma_1 = 1.000$						
$\sigma_2 = 1.000$						
$\sigma_3 = 1.000$						
Threshold Parameter Estimates						
$\gamma_1 = -.021$						
$\gamma_2 = -.940$						
$\gamma_3 = -.610$						

metric binary relations facilitates what would otherwise be a difficult interpretive task of finding and interpreting key asymmetries in the data.

The W matrix of weights presented in Table 7 represents individual differences among the three persons. Person 1 tends to weigh the third ("arousal") dimension most heavily; Person 2 tends to weigh

Table 8
Correlations of Stimulus (X) and Response (Z) Emotion Positions

	Column Dimension		
Row Dimension	Z_1	Z_2	Z_3
X_1	.914	.643	-.614
X_2	.476	.868	-.055
X_3	-.561	-.047	.916

the second ("dominance") dimension most heavily; and Person 3 tends to weigh the first ("pleasure") dimension most heavily. These differences probably reflect the fact that the three persons were chosen for their tendency to represent three contrasting groups derived by means of the cluster analysis. (Note that the individual joint space representations for these three persons could be recovered by multiplying both the row (X) and column (Z) emotion coordinates by the square roots of their person weights.)

Finally, Table 9 illustrates the working of the model in terms of the threshold parameters and weighted Euclidean distances. This table presents the fitted weighted distances, threshold coefficient, and fitted and actual selection values for Person 2 when the emotion "satisfied" is presented. The model correctly classifies the evocation of "peaceful" and "satisfied". Another interesting facet of Table 9 concerns the computed weighted distances of other emotions. For example, "relaxed", "relieved", and "pleased" are emotions relatively close to the presented "satisfied" coordinates, but whose distances are slightly greater than the threshold value of .094. Clearly, these other three emotions are related to the "satisfied" and "peaceful" emotions and should be close to them. Hence, the model captures both their closeness and their failure to reach the elicitation threshold for this person. Emotions such as "in pain", "sad", "grief-stricken", "frustrated", and "crying" have the largest distances from "satisfied". Such an analysis could be performed for any of the 28 emotions for each of the three persons studied here.

Discussion

Other Applications

Clearly, this stochastic three-way unfolding model is not restricted merely to the investigation of emotion associations. As mentioned earlier, such three-way, two-mode binary data occur in virtually all word association contexts. For example, in marketing applications, this method could be used for the investigation of the connotation of particular words in advertising copy (Green & Tull, 1978). It could also facilitate the study of brand associations, such as those rendered by persons in comparing different brands within the same product class (e.g., "Check any of the following brands which are similar to ________"). A sociometric application might involve the study of personal networks based on persons' selections of friends from a designated list over time. Here, the ability to accommodate asymmetry is required because Person A may consider Person B a friend in the absence of a reciprocal choice. From an economic perspective, "competitive" data could be collected over time, in order to learn which other firms are targeted as competitors by decision-makers at a given firm. For example, consider the soft drink market. It is conceivable that one brand, say Sprite, may attempt to compete against another brand, say Seven-Up, but that Seven-Up may attempt to compete against cola drinks (e.g., by positioning itself as "The Uncola"). In short, there are clearly numerous potential applications involving this new methodology.

Figure 2
Single-Dimensional Joint Space Representations

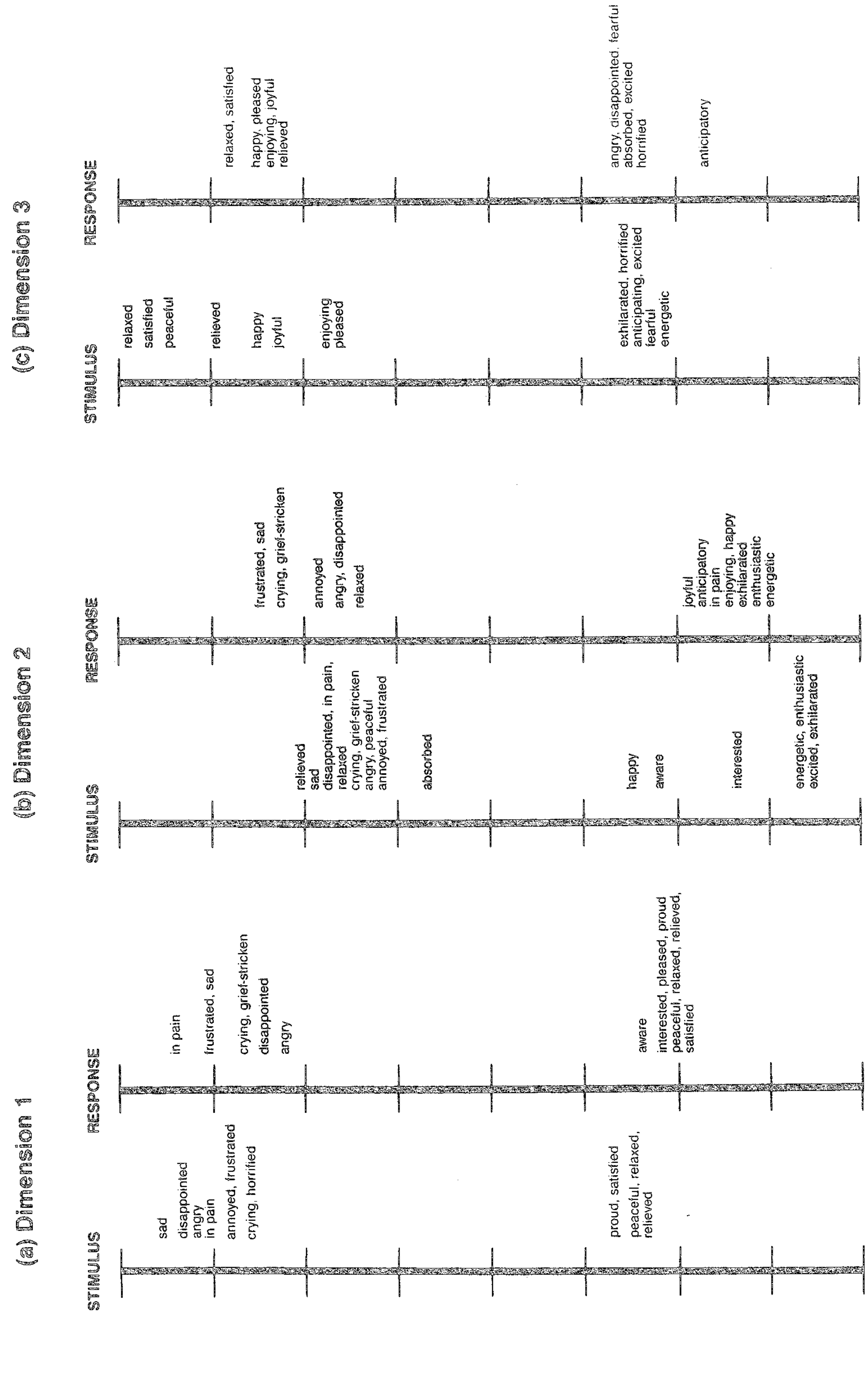


Table 9
Predicted Values for Person 2 for "Satisfied" Emotion
With Threshold Coefficient = .094

Response Emotion	Weighted Squared Distance from Satisfied	Predicted Response	Actual Response
Absorbed	1.57	0	0
Angry	2.47	0	0
Annoyed	1.80	0	0
Anticipatory	2.42	0	0
Aware	1.49	0	0
Crying	2.80	0	0
Disappointed	2.60	0	0
Energetic	2.40	0	0
Enjoying	1.17	0	0
Enthusiastic	1.63	0	0
Excited	2.06	0	0
Exhilarated	1.92	0	0
Fearful	1.23	0	0
Frustrated	3.11	0	0
Grief-stricken	2.80	0	0
Happy	1.18	0	0
Horrified	2.07	0	0
Interested	.68	0	0
Joyful	.91	0	0
In pain	3.59	0	0
Peaceful	.02	1	1
Pleased	.23	0	0
Proud	.53	0	0
Relaxed	.12	0	0
Relieved	.12	0	0
Sad	3.10	0	0
Satisfied	.08	1	-
Surprised	1.11	0	0

Further Research

There are a number of areas that require additional research. Perhaps the most pressing involves a rigorous monte carlo analysis to investigate the appropriateness of the χ^2 test for nested models (in the face of incidental parameters). Here, it is desirable to test whether the difference of deviance scores is truly asymptotically distributed as χ^2 with the difference in degrees of freedom as the appropriate degrees of freedom. This question could be addressed through simulations where, say, a large number of datasets from known two-dimensional solutions are created and the resulting three-way data run in two and three dimensions. The resulting difference in deviance scores can be tested (e.g., using a Q-Q plot or statistical goodness-of-fit test) against a χ^2 distribution.

Other avenues of related research exist in extending the methodology into a number of useful directions. One obvious extension would involve modifications of the algorithm to accommodate three-mode binary data. This could either be in the form of choice data or asymmetric proximity-type data for generally rectangular submatrices. Jedidi (1986) is currently investigating this extension, as well as examining linear restrictions on parameters, "floating" ideal points (see DeSarbo, 1978), and a vector model version (see DeSarbo, Keramidas, & Cho, 1986).

Appendix

The Algorithm

Maximum likelihood methods are used to estimate the desired set of parameters to maximize R (or minimize $-R$) in Equation 9. The method of conjugate gradients (Fletcher & Reeves, 1964) is used to solve this nonlinear, unconstrained optimization problem. The partial derivatives of R in Equation 9 with respect to the various parameters are:

$$\frac{\partial R}{\partial w_{kt}} = \sum_{i}^{N}\sum_{j\neq i}^{N}\frac{\phi(\cdot)(x_{it}-z_{jt})^2}{\sigma_{ijk}}\left[\frac{-y_{ijk}}{\Phi(\cdot)}+\frac{(1-y_{ijk})}{1-\Phi(\cdot)}\right] \quad , \tag{A1}$$

$$\frac{\partial R}{\partial x_{it}} = \sum_{k}^{M}\sum_{j\neq 1}^{N}\frac{2\phi(\cdot)w_{kt}(x_{it}-z_{jt})}{\sigma_{ijk}}\left[\frac{-y_{ijk}}{\Phi(\cdot)}+\frac{(1-y_{ijk})}{1-\Phi(\cdot)}\right] \quad , \tag{A2}$$

$$\frac{\partial R}{\partial z_{kt}} = \sum_{k}^{M}\sum_{i\neq j}^{N}\frac{2\phi(\cdot)w_{kt}(x_{it}-z_{jt})}{\sigma_{ijk}}\left[\frac{-(1-y_{ijk})}{1-\Phi(\cdot)}+\frac{y_{ijk}}{\Phi(\cdot)}\right] \quad , \tag{A3}$$

$$\frac{\partial R}{\partial \gamma_{k}} = \sum_{i}^{N}\sum_{j\neq i}^{N}\frac{\phi(\cdot)}{\sigma_{ijk}}\left[\frac{-y_{ijk}}{\Phi(\cdot)}+\frac{(1-y_{ijk})}{1-\Phi(\cdot)}\right] \quad , \tag{A4}$$

$$\frac{\partial R}{\partial \sigma_{k}} = \sum_{i}^{N}\sum_{j\neq i}^{N}2\sigma_k^{-3}\, f^*(\mathbb{W},\mathbb{X},\mathbb{Z})\ \phi(\cdot)\ \left[\frac{y_{ijk}}{\Phi(\cdot)}-\frac{(1-y_{ijk})}{1-\Phi(\cdot)}\right] \quad , \tag{A5}$$

where $\phi(\cdot)$ represents the evaluation of the standard normal density at $(\cdot)$. In order to enforce optional positivity constraints for w_{kt} and/or negativity constraints for γ_k, a substituted variables approach (Gill, Murray, & Wright, 1981) is used. For example, to enforce such constraints for w_{kt}, $b_{kt}^2 = w_{kt}$ is estimated and derivatives are taken with respect to b_{kt}.

For the sake of convenience, assume that the relevant parameters to be estimated are contained in the vector $\mathbb{\delta}$ and that $\nabla\mathbb{\delta}$ is the vector of partial derivatives for this set of parameters. The complete conjugate gradient procedure for minimizing $-R$ can be summarized as follows:

1. Start with initial parameter estimates $\mathbb{\delta}$; set the iteration counter (MIT) = 1.
2. Set the first search direction $\mathbb{S}^{(1)} = -\nabla\mathbb{\delta}^{(1)}$.
3. Find $\mathbb{\delta}^{(2)}$ according to the relation

   $$\mathbb{\delta}^{(2)} = \mathbb{\delta}^{(1)} + u^{(1)}\mathbb{S}^{(1)} \quad , \tag{A6}$$

   where $u^{(1)}$ is the optimal step length in the direction $\mathbb{S}^{(1)}$. The optimal step size is found by a quadratic interpolation method. Set MIT = 2.
4. Calculate $\nabla\mathbb{\delta}^{(\mathrm{MIT})}$ and set

   $$\mathbb{S}^{(\mathrm{MIT})} = -\nabla\mathbb{\delta}^{(\mathrm{MIT})} + \frac{(\nabla\mathbb{\delta}^{(\mathrm{MIT})})'(\nabla\mathbb{\delta}^{(\mathrm{MIT})})}{(\nabla\mathbb{\delta}^{(\mathrm{MIT}-1)})'(\nabla\mathbb{\delta}^{(\mathrm{MIT}-1)})}\mathbb{S}^{(\mathrm{MIT}-1)} \quad . \tag{A7}$$

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

   $$\mathbb{\delta}^{(\mathrm{MIT}+1)} = \mathbb{\delta}^{(\mathrm{MIT})} + u^{(\mathrm{MIT})}\mathbb{S}^{(\mathrm{MIT})} \quad . \tag{A8}$$

6. If $\mathbb{\delta}^{(\mathrm{MIT}+1)}$ is optimal, stop. Otherwise set MIT = MIT + 1 and return to step 4 above (i.e., undertake another iteration). A number of convergence tests are performed here to test whether additional iterations are required. These involve standard tests for the length of the gradient, the number of iterations, the amount of improvement in subsequent objective function values, and an examination of the amount of change in the parameters from subsequent iterations.

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

This conjugate gradient method is particularly useful for optimizing functions of several parameters (Rao, 1979), as 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 of the algorithms is linear only if the iterative procedure is "restarted" occasionally. Restarts have been implemented in the algorithms automatically, depending upon successive improvement in the objective function.

Several goodness-of-fit measures are computed for this model:

1. Sums of squares between $\mathbb{Y}$ and $\hat{\mathbb{P}} = \|\hat{p}_{ijk}\|$;
2. A deviance measure (Nelder & Wedderburn, 1972):

$$D = -2\left[\sum_{k=1}^{M}\sum_{i \neq j}^{N}\sum^{N} y_{ijk} \ln(\hat{p}_{ijk}) + (1-y_{ijk}) \ln(1-\hat{p}_{ijk})\right] = -2 \ln L \quad , \tag{A9}$$

where $\hat{p}_{ijk}$ is the estimated probability that presentation of stimulus i to person k evokes stimulus j, as expressed in Equation 3. Note that nested models can be theoretically tested as the difference between respective deviance measures. This difference is asymptotically χ^2 distributed with the difference in model degrees of freedom (i.e., the effective number of independent parameters in the estimated model) providing the appropriate χ^2 test degrees of freedom.

This test is appropriate in assessing dimensionality as well as the various optional model specifications because of the nested terms. However, recall that this is an asymptotic test. One obvious problem with this test concerns incidental parameters in the likelihood function (i.e., parameters whose order varies according to the order of $\mathbb{Y}$, such as the x_{it}s). According to Andersen (1980), maximum likelihood estimators in such cases may not be consistent. This is particularly relevant in the present case, where there are no replications. Takane (1983) also demonstrated this problem in his item response model and had to use a marginal likelihood formulation (integrating over person incidental parameters) in order for the asymptotic χ^2 properties to hold. He initially found that his original model and estimation scheme tended to overestimate the true dimensionality. Because of these potential problems, other "goodness-of-fit" indices are introduced.
3. The simple matching coefficient (Sneath & Sokal, 1973) is calculated between the actual $\mathbb{Y}$ and the predicted $\hat{\mathbb{Y}}$. This simple matching coefficient calculates the total number of *0* and *1* matches and divides it by the total number of matches and nonmatches. This is also calculated for each slice of $\mathbb{Y}$ and $\hat{\mathbb{Y}}$, for each row, and for each column.

Note that the model degrees of freedom are defined as the effective number of free model parameters. This can be specified as

$$T(M+2N)+M+\alpha-2T \quad , \tag{A10}$$

where

$$\alpha = \begin{cases} 0 & \text{if } \sigma_{ijk} = 1 \text{ or some constant,} \\ M-1 & \text{if } \sigma_{ijk} = \sigma_k. \end{cases}$$

The term $2T$ is subtracted because of the indeterminacy of the origin for the x_{it}s and z_{jt}s (origin shifts of $\mathbb{L} = \begin{bmatrix}\mathbb{X}\\ \mathbb{Z}\end{bmatrix}$ do not affect Euclidean distances) and the scalar indeterminacy associated with the weights (if w_{kt} is multiplied by a constant n_t and if x_{it}, z_{jt} are divided by $(n_t)^{1/2}$ for all i, j, the constant will cancel). Finally, if σ_k is estimated, then $M - 1$ free parameters will result, as long as one of the σ_ks can be set equal to 1 to set the overall scale. Equation A10 must be altered when all individual parameters are estimated, given the obvious scalar indeterminacy between W_k, γ_k, and σ_k (one of these must be constrained to some constant because all three are not simultaneously estimable).

References

Andersen, E. B. (1980). *Discrete statistical models with social science applications*. New York: North Holland.

Benzecri, J. P. (1969). Statistical analysis as a tool to make patterns emerge from data. In S. Watanabe (Ed.), *Methodologies of pattern recognition* (pp. 35–74). New York: Academic Press.

Benzecri, J. P. (1973a). *L'analyse des donnees I: La taxinomie*. Paris: Dunod.

Benzecri, J. P. (1973b). *L'analyse des donnees II: L'analyse des correspondences*. Paris: Dunod.

Carroll, J. D. (1972). Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences, Vol. 1: Theory* (pp. 105–155). New York: Seminar Press.

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., & Chang, J. J. (1972). *IDIOSCAL (Individual Differences in Orientation SCALing): A generalization of INDSCAL allowing idiosyncratic reference systems as well as an analytic reference system and an analytic approximation to INDSCAL*. Paper presented at the meetings of the Psychometric Society, Princeton NJ.

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

Conte, H. T. (1975). *A circumplex model for personality traits*. Unpublished doctoral dissertation, New York University.

Coombs, C. H. (1953). The theory and method of social measurement. In L. Festinger and D. Katz (Eds.), *Research methods in the behavioral sciences* (pp. 471–535). New York: Wiley.

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

Darwin, C. (1872). *The expression of the emotions in man and animals*. London: J. Murray.

Davison, M. L. (1976). Fitting and testing Carroll's weighted unfolding model for preferences. *Psychometrika*, *41*, 233–247.

de Leeuw, J. (1973). *Canonical analysis of categorical data*. Leiden, The Netherlands: University of Leiden Press.

de Leeuw, J., & Heiser, W. J. (1980). Theory of multidimensional scaling. In P. R. Krishnaiah & L. Kanal (Eds.), *Handbook of statistics* (pp. 285–316). New York: North Holland.

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

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

DeSarbo, W. S., & Carroll, J. D. (1981). Three-way metric unfolding. In J. W. Keon (Ed.), *Marketing: Measurement and Analysis 1981* (pp. 157–183). Providence RI: Institute of Management Science.

DeSarbo, W. S., & Carroll, J. D. (1985). Three-way metric unfolding via alternating weighted least-squares. *Psychometrika*, *50*, 275–300.

DeSarbo, W. S., & Hoffman, D. L. (1986). Simple and weighted unfolding threshold models for the spatial representation of binary choice data. *Applied Psychological Measurement*, *10*, 247–264.

DeSarbo, W. S., Keramidas, E. M., & Cho, J. (1986). *A new stochastic multidimensional scaling vector threshold model for the spatial representation of binary choice data*. Unpublished manuscript, University of Pennsylvania.

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.

Everitt, B. (1980). *Cluster analysis* (2nd ed.). London: Halsted Press.

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

Gifi, A. (1981a). *Nonlinear multivariate analysis*. Unpublished manuscript, University of Leiden, The Netherlands.

Gifi, A. (1981b). *Homogeneity analysis*. Unpublished manuscript, University of Leiden, The Netherlands.

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

Green, P. E., & Tull, D. S. (1978). *Research for mar-*

keting decisions (4th ed.). Englewood Cliffs NJ: Prentice Hall.

Green, P. E., & DeSarbo, W. (1980). Two models for representing unrestricted choice data. In K. B. Monroe (Ed.), *Advances in consumer research, Vol. 8* (pp. 309–312). Ann Arbor MI: Association for Consumer Research.

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

Guttman, L. L. (1944). A basis for scaling qualitative data. *American Sociological Review*, *9*, 139–150.

Harshman, R. A. (1970). *Foundations of the PARAFAC procedure: Models and conditions for an explanatory multimodal factor analysis*. Unpublished manuscript, University of California, Los Angeles.

Harshman, R. A. (1972). *PARAFAC 2: Mathematical and technical notes*. Unpublished manuscript, University of California, Los Angeles.

Harshman, R. A. (1975). *Models for the analysis of asymmetrical relationships among* N *objects or stimuli*. Presented at The US-Japan Seminar on Multidimensional Scaling, University of California, San Diego.

Havlena, W. J. (1985). *The varieties of consumption experience*. Unpublished doctoral dissertation, Columbia University.

Havlena, W. J., & Holbrook, M. B. (1985). *The varieties of consumption experience: Comparing two typologies of emotion in consumer behavior*. Unpublished manuscript, Southern Methodist University, Dallas TX.

Havlena, W. J., Holbrook, M. B., & Lehmann, D. R. (1986). *Contrasting methods for assessing the validity of an emotional typology: Words versus deeds*. Unpublished manuscript, Southern Methodist University, Dallas TX.

Heiser, W. J. (1981). *Unfolding analysis of proximity data*. Leiden, The Netherlands: Leiden Press.

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

Holbrook, M. B., Moore, W. L., & Winer, R. S. (1980). Using "pick-any" data to represent competitive positions. In R. P. Leon (Ed.), *Proceedings: Market measurement and analysis* (pp. 129–134). Providence RI: Institute of Management Science.

Holbrook, M. B., Moore, W. L., & Winer, R. S. (1982). Constructing joint spaces from pick-any data: A new tool for consumer analysis. *Journal of Consumer Research*, *9*, 99–105.

Izard, C. E. (1977). *Human emotions*. New York: Plenum Press.

Jedidi, K. (1986). *A stochastic three-way unfolding threshold methodology for the spatial representation of binary choice data*. Doctoral dissertation proposal, University of Pennsylvania.

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*, 115–129.

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

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

Levine, J. H. (1979). Joint-space analysis of "pick-any" data: Analysis of choices from an unconstrained set of alternatives. *Psychometrika*, *44*, 85–92.

Lingoes, J. C., & Borg, I. (1978). A direct approach to individual differences scaling using increasingly complex transformations. *Psychometrika*, *43*, 491–519.

Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. Cambridge MA: MIT Press.

Muthén, B. (1981). Factor analysis of dichotomous variables: American attitudes toward abortion. In D. N. Jackson & E. F. Borgatta (Eds.), *Factor analysis and measurement in sociological research* (pp. 114–136). Beverly Hills CA: Sage Press.

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society, A*, *135*, 370–384.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications*. Toronto: University of Toronto Press.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana IL: University of Illinois Press.

Plutchik, R. (1980). *Emotion: A psychoevolutionary synthesis*. New York: Harper and Row.

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

Ramsay, J. O. (1977). Maximum likelihood estimation in multidimensional scaling. *Psychometrika*, *42*, 241–266.

Rao, S. S. (1979). *Optimization: Theory and applications*. New York: Wiley.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*, 813–838.

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

Spencer, H. (1855). *Principles of psychology*. New York: Appleton.

Srinivasan, V., & Shocker, A. D. (1973). Linear programming techniques for multidimensional analysis of preferences. *Psychometrika*, *38*, 337–369.

Takane, Y. (1983). Choice model analysis for the "pick-any/*n*" type of binary data. *Handout for talk at the European Psychometric and Classification Meetings*. July, Jouy-en-Josas, France.

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

Thurstone, L. L. (1929). Theory of attitude measurement. *Psychological Review*, *36*, 222–241.

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

Tucker, L. R. (1972). Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika*, *37*, 3–27.

Wundt, W. (1902). *Outlines of psychology*. Leipzig: W. Engelman.

Young, F. W. (1975). *An asymmetric Euclidean model for multiprocess asymmetric data*. Paper presented at the U.S.-Japan Seminar on Multidimensional Scaling, University of California, San Diego.

Zinnes, J. L., & Wolff, R. P. (1977). Single and multidimensional same-different judgments. *Journal of Mathematical Psychology*, *16*, 30–50.

Author's Address

Send requests for reprints or further information to Wayne S. DeSarbo, E. L. Cox School of Business, Southern Methodist University, Dallas TX 75275, U.S.A.