

## TSCALE: A NEW MULTIDIMENSIONAL SCALING PROCEDURE BASED ON TVERSKY'S CONTRAST MODEL

WAYNE S. DESARBO

DEPARTMENTS OF MARKETING AND STATISTICS  
SCHOOL OF BUSINESS ADMINISTRATION, UNIVERSITY OF MICHIGAN

MICHAEL D. JOHNSON

MARKETING DEPARTMENT  
SCHOOL OF BUSINESS ADMINISTRATION, UNIVERSITY OF MICHIGAN

AJAY K. MANRAI

DEPARTMENT OF BUSINESS ADMINISTRATION  
COLLEGE OF BUSINESS AND ECONOMICS, UNIVERSITY OF DELAWARE

LALITA A. MANRAI

DEPARTMENT OF BUSINESS ADMINISTRATION  
COLLEGE OF BUSINESS AND ECONOMICS, UNIVERSITY OF DELAWARE

ELIZABETH A. EDWARDS

STATISTICS DEPARTMENT  
SCHOOL OF BUSINESS ADMINISTRATION, UNIVERSITY OF MICHIGAN

Tversky's contrast model of proximity was initially formulated to account for the observed violations of the metric axioms often found in empirical proximity data. This set-theoretic approach models the similarity/dissimilarity between any two stimuli as a linear (or ratio) combination of measures of the common and distinctive features of the two stimuli. This paper proposes a new spatial multidimensional scaling (MDS) procedure called TSCALE based on Tversky's linear contrast model for the analysis of generally asymmetric three-way, two-mode proximity data. We first review the basic structure of Tversky's conceptual contrast model. A brief discussion of alternative MDS procedures to accommodate asymmetric proximity data is also provided. The technical details of the TSCALE procedure are given, as well as the program options that allow for the estimation of a number of different model specifications. The non-linear estimation framework is discussed, as are the results of a modest Monte Carlo analysis. Two consumer psychology applications are provided: one involving perceptions of fast-food restaurants and the other regarding perceptions of various competitive brands of cola soft-drinks. Finally, other applications and directions for future research are mentioned.

**Key words:** multidimensional scaling, asymmetric proximity data, Tversky's contrast model, consumer psychology.

### 1. Introduction

Tversky's (1977) contrast model provides a flexible framework for understanding

The authors wish to acknowledge the reviews of prior versions of this manuscript by three anonymous reviewers and the editor.

Requests for reprints should be sent to Wayne S. DeSarbo, Marketing and Statistics Depts., School of Business Administration, The University of Michigan, Ann Arbor, MI 48109-1234.

similarity across a range of stimulus and judgment contexts. This model describes the similarity between two stimuli as a linear (or ratio) combination or contrast of measures of their common and distinctive features. It captures a family of possible similarity relations depending upon the feature measures of the stimuli (Gati & Tversky, 1982, 1984; Sattath & Tversky, 1977, 1987; Tversky, 1977; Tversky & Gati, 1978, 1982; Tversky & Hutchinson, 1986). Ideally, the representation of these proximity relationships should incorporate this same flexibility. Yet, existing procedures only represent special cases of the general model. Additive tree (Sattath & Tversky, 1977) and extended tree (Corter & Tversky, 1986) procedures, for example, express the distance between stimuli only in terms of distinctive features, while hierarchical clustering (S. C. Johnson, 1967) and additive clustering procedures (Shepard & Arabie, 1979) express distance only in terms of common features. No one procedure captures the full range of possible similarity measures permissible under this contrast model.

This article describes TSCALE, a multidimensional scaling procedure based on Tversky's contrast model for asymmetric three-way, two-mode proximity data. A unique aspect of TSCALE is the conceptualization of a latent dimensional structure to describe the judgmental stimuli. Whether stimuli are dimensional or feature-based in their attribute representations, these attributes can be captured at a more abstract level using continuous dimensions (Johnson & Fornell, 1987). This abstract or latent dimensional structure is linked to a corresponding feature-based representation and the underlying similarity judgment process. Our proposed procedure utilizes the information in the proximity data to estimate both the latent structure and the degree to which various common and distinctive aspects of this latent structure surface in the particular judgment task.

We begin by describing Tversky's contrast model in more detail. A number of alternative MDS procedures for asymmetric proximity data are briefly described. The TSCALE procedure is then presented followed by the results of a preliminary Monte Carlo analysis and two consumer psychology applications. Finally, directions for future research are discussed.

## 2. Tversky's Contrast Model

Traditional spatial models of similarity represent stimuli as points in a derived multidimensional space where the metric distances between these points correspond in some manner to the empirical proximity data (Shepard, 1962). Nonspatial models, such as ultrametric and additive trees, represent stimuli at the terminal nodes in a defined graph structure with a specified distance metric (S. C. Johnson, 1967). As Tversky (1977) observes, the metric axioms implicit with the use of such models (e.g., minimality, symmetry, and the triangle inequality) are often systematically violated in such collected data. Tversky's contrast model arose as an alternative (to the metric-distance approaches) that could account for these violations. This conceptual model was initially based on the presumption that stimuli are cognitively represented using features rather than dimensions. Whereas dimensions are attributes on which stimuli vary as a matter of degree, features tend to be more dichotomous aspects that a stimulus either has or does not have (Garner, 1978). Feature-based representations are common in psychological research. Originally investigated by Restle (1959), features are central to models of semantic judgment (Smith, Shoben, & Rips, 1974), stimulus categorization (Rosch, 1975; Rosch & Mervis, 1975), choice (Tversky, 1972), as well as similarity (Shepard & Arabie, 1979; Tversky, 1977).

Tversky (1977) views similarity judgments as comparisons or contrasts of common and distinctive features. When faced with a similarity task, people extract and compile

a limited list of relevant features from remembered information. Their judgment of similarity is based on a comparison of these features. Formally stated (for the linear version of the model), the dissimilarity between two stimuli  $i$  and  $j$  is modeled as:

$$\delta_{ij} \cong \alpha f(I - J) + \beta f(J - I) - \theta f(I \cap J), \quad (1)$$

where the two stimuli  $i$  and  $j$  are associated with feature sets  $I$  and  $J$ , respectively,  $\alpha$ ,  $\beta$ , and  $\theta$  are nonnegative scalar parameters,  $\delta_{ij}$  is the observed (asymmetric) dissimilarity between stimuli  $i$  and  $j$ , and  $f$  is some specified function. According to this model, the dissimilarity between the two stimuli is a function of their common features, ( $I \cap J$ ), the features distinctive to  $i$ , ( $I - J$ ), and the features distinctive to  $j$ , ( $J - I$ ). This particular model expresses the proximity of  $i$  and  $j$  as a linear combination, or a contrast, of their common and distinctive features. Overall similarity (dissimilarity) increases (decreases) with the measure of common features and decreases (increases) with the measure of distinctive features. The scale values  $f(I)$  and  $f(J)$  capture the overall measure of the feature sets of stimuli  $i$  and  $j$ , respectively, which vary with the "intensity, frequency, familiarity, good form, and informational content" of the stimuli and their features (Tversky, 1977, p. 332). This linear version can accommodate both of Restle's (1961) models of psychological distance where either ( $\alpha = \beta = 0$ ,  $\theta = 1$ ) or ( $\alpha = \beta = 1$ ,  $\theta = 0$ ).

A family of dissimilarity relations are possible under this model depending upon the parameters  $\alpha$ ,  $\beta$ , and  $\theta$ , which describe the importance of the different feature measures with respect to the observed dissimilarity judgments. Note, this model is not limited to situations where stimuli are described only by features. Nominal variables with more than two levels can be expressed as a set of features or qualitative dimensions (Gati & Tversky, 1982, p. 329). Ranges of an inherently continuous dimension can also be treated as features depending on the stimuli (Johnson & Fornell, 1987) or the preferred mode of processing (Garner, 1978).

The contrast model in (1) expresses a simple form of feature matching. Tversky's (1977) framework allows for other matching functions, including a ratio model in which similarity or dissimilarity is normalized. For example, dissimilarity can be modeled as a ratio of distinctive to total features, where

$$\delta_{ij} \cong \frac{\alpha f(I - J) + \beta f(J - I)}{\alpha f(I - J) + \beta f(J - I) + \theta f(I \cap J)}. \quad (2)$$

In addition, similarity can be expressed as a ratio of common to total features. These ratio models assume that the proximities are normalized between 0 and 1. Ratio formulations of the contrast model are somewhat attractive because they generalize several existing set-theoretic models of similarity (see Tversky, 1977). For example, Gregson's (1975) and Sjöberg's (1972) models are special cases of the similarity analog to (2) where  $\alpha = \beta = \theta = 1$ , while Bush and Mosteller's (1951) model presumes  $\alpha = \theta = 1$ ,  $\beta = 0$ , and Eisler and Ekman's (1959) model presumes  $\theta = 1$ ,  $\alpha = \beta = 1/2$ .

However, several considerations support the appropriateness of a linear, as opposed to a ratio, formulation of the contrast model in equation (1). Linear models of judgment are conceptually simple and straightforward. They have proven to be very powerful predictors in a number of judgment tasks (Dawes, 1979). Linear models are also paramorphic to a wide range of possible judgment processes (Einhorn, Kleinmuntz, & Kleinmuntz, 1979). At the same time, Abelson and Levi (1985) suggest that linear models are robust primarily in tasks that are characterized first by conditional monotonicity between the cues or attributes used to make judgments and the corre-

sponding performance or objective being judged, and second by large error components where nonlinear relationships are easily hidden.

Tversky provides considerable support for the contrast model in studies involving people, countries, faces, forms, and figures (Gati & Tversky, 1982; Sattath & Tversky, 1977; Tversky & Gati, 1978, 1982). The flexibility of the contrast model stems from the manner in which  $\alpha$ ,  $\beta$ ,  $\theta$  and the function  $f$  depend on the nature of the similarity task and the construction of the set of stimuli. For example, the model highlights the differential importance of common and distinctive features across task environments. When judging similarity, it is natural to focus on what alternatives have in common (e.g.,  $\theta > \alpha + \beta$ ). When judging dissimilarity, we naturally focus on what is distinctive to the alternatives (e.g.,  $\alpha + \beta > \theta$ ). Thus, similarity and dissimilarity may not be perfectly negatively correlated. If a pair of alternatives has both many common and many distinctive features, it may be both more similar in a similarity task and more different in a dissimilarity task than another pair with fewer common and distinctive features. Consistent with this prediction, Tversky and Gati (1978) report instances in which one group of subjects selected prominent pairs of countries (e.g., East Germany and West Germany) as more similar than nonprominent pairs (Ceylon and Nepal), while a second group selected these same pairs as more different. Intuitively, more prominent countries have both more common and more distinctive features than nonprominent countries. In a slightly different vein, Gati and Tversky (1984) demonstrate the differential importance of adding a common or distinctive feature to verbal and pictorial stimuli (e.g., descriptions of persons versus schematic faces). Their results show that common features are more pronounced for verbal stimuli while distinctive features loom larger for pictorial stimuli.

The contrast model also captures asymmetric proximity relationships where the  $(i, j)$  and  $(j, i)$  elements of the right side of (1) or (2) may not be equal. Asymmetric relationships are particularly evident in directional or subject-referent similarity judgments of the form "how similar is  $i$  to  $j$ ?", where  $i$  is the subject and  $j$  is the referent. For example, Tversky and Gati (1978) found subjects' rating of the similarity of North Korea to Red China to be greater than the similarity of Red China to North Korea. Whenever a stimulus is the focus or subject of the judgment, or serves as the anchor against which another stimulus is compared, it is natural to focus on that stimulus' features. As a result, the distinctive features of the subject often detract more from similarity than the distinctive features of the referent ( $\alpha > \beta$ ). The contrast model predicts asymmetry in this context when stimuli differ in their distinctive feature measures ( $\alpha \neq \beta$  and  $f(I) \neq f(J)$ ). In Tversky's (1977) Red China-North Korea example, the contrast model explains the observed asymmetry given a greater distinctive feature measure for Red China that detracts more from similarity when Red China is the subject rather than the referent in the comparison.

### 3. Alternative MDS Procedures for the Analysis of Asymmetric Proximity Data

There are a number of existing MDS procedures that are available for the analysis of (typically one-mode, two-way) asymmetric proximity data. For example, Gower (1978) and Constantine and Gower (1978) proposed the application of multidimensional unfolding for the analysis and spatial representation of asymmetric proximity data (Bennett & Hayes, 1960; Coombs, 1950). Here, the rows and columns are considered as distinct entities (e.g., stimulus and response, subject and referent, etc.) and are represented as two distinct sets of points. The basic mathematical structure of this model is as follows:

$$\delta_{ij} \cong g(d_{ij}), \quad d_{ij} = \left[ \sum_{t=1}^T (Y_{it} - X_{jt})^2 \right]^{1/2},$$

where  $\delta_{ij}$  is the observed asymmetric dissimilarity between row stimulus  $i$  and column stimulus  $j$ ,  $f$  is some monotone function, and  $d_{ij}$  is the Euclidean distance between row stimulus  $i$  and column stimulus  $j$  in a derived  $T$ -dimensional space. Unfolding procedures estimate the coordinates of these two sets of points ( $Y_{it}$ ,  $X_{jt}$ ) in some specified number of dimensions ( $T$ ) whose distances in the derived space optimally approximate the  $\delta_{ij}$ 's. A number of unfolding procedures exist for such analysis such as PREFMAP (Carroll, 1980), GENFOLD2 (DeSarbo & Rao, 1984, 1986), KYST (Kruskal, Young, & Seely, 1973), and ALSICAL (Takane, Young, & de Leeuw, 1977).

Young (1975) proposed an alternative distance approach called ASYMSCAL for the analysis of asymmetric proximity data. ASYMSCAL estimates one set of stimulus coordinates, as well as differential weights for dimensions for either the row or column stimuli, or both. Let  $W_{it}$  (row) and  $C_{jt}$  (column) designate such weights. Then, the model for the general case may be written as

$$\delta_{ij} \cong d_{ij} = \left[ \sum_{t=1}^T W_{it} C_{jt} (X_{it} - X_{jt})^2 \right]^{1/2}.$$

ASYMSCAL produces a multidimensional map of the stimulus space and separate configurations of stimulus weights.

In considering scalar-product models (i.e., spatial, but non-distance models) for asymmetric proximity data, there is Harshman's (1975) metric procedure that involves a matrix decomposition into directional components (DEDICOM). The strong case of the model assumes a common set of dimensions for the rows and columns, so that the model is in that sense symmetric. Asymmetry is modeled by a set of indices of "directional relationship" that indicate the degree to which each dimension affects each other dimension. Let  $\mathbf{S}$  denote the ( $N \times N$ ) matrix of asymmetric similarity data. The DEDICOM model can be written as:

$$\mathbf{S} \cong \mathbf{VDV}',$$

where  $\mathbf{V}$  is an ( $N \times T$ ) matrix of weights of  $N$  stimuli on  $T$  ( $< N$ ) dimensions and  $\mathbf{D}$  is an asymmetric square ( $T \times T$ ) matrix giving the directional relationships between dimensions.

A different approach involving a geometrically interesting generalization of scalar products (defined initially only for two or three dimensions) has been formulated by Chino (1978) for asymmetric data. Chino (1979, 1990) has extended the Chino (1978) procedure into dimensions higher than three. This extended model can be written as

$$s_{ij} = a\mathbf{X}_i\mathbf{X}_j + b\mathbf{X}_i\mathbf{I}^*\mathbf{X}_j + c,$$

where  $s_{ij}$  denotes the similarity judgement between stimuli  $i$  and  $j$ , and  $\mathbf{X}_i$ ,  $\mathbf{X}_j$  denote the  $T$ -dimensional coordinate vectors of stimuli  $i$  and  $j$ , respectively, while  $a$ ,  $b$ , and  $c$  denote constants. Moreover,  $\mathbf{I}^*$  is a skew-symmetric matrix of the form

$$\mathbf{I}^* = \begin{pmatrix} 0 & 1 & -1 & 1 & \dots \\ -1 & 0 & 1 & -1 & \dots \\ 1 & -1 & 0 & 1 & \dots \\ -1 & 1 & -1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \dots \\ \vdots & \vdots & \vdots & \vdots & \dots \end{pmatrix} = \left\{ \text{sgn} \begin{pmatrix} 1 & 2 & \dots & T \\ \dots & p & q & \dots \end{pmatrix} \right\},$$

where

$$\text{sgn} \begin{pmatrix} 1 & 2 & \dots & T \\ \dots & p & q & \dots \end{pmatrix} = \begin{cases} 0, & \text{if two indices } (p, q) \text{ are the same,} \\ 1, & \text{if permutation } (\dots p q \dots) \text{ is even,} \\ -1, & \text{if permutation } (\dots p q \dots) \text{ is odd.} \end{cases}$$

MDPREF (Carroll, 1980) is another type of scalar-products or projection model (vector model) useful for two-mode, two-way dominance or asymmetric proximities data. It is a form of weighted principal components analysis for determining a multi-dimensional representation for the rows and columns of the input data. In the case of asymmetric proximity data, MDPREF produces a joint space of the row stimuli as vectors and the column stimuli as points in the space. Its mathematical structure is given as

$$s_{ij} \cong \sum_{t=1}^T Y_{it} X_{jt} \quad \text{or} \quad \mathbf{S} = \mathbf{YX}',$$

where  $\mathbf{S} = ((s_{ij}))$ ,  $Y_{it}$  is the  $t$ -th coordinate of the vector terminus for row stimulus  $i$  ( $\mathbf{Y} = ((Y_{it}))$ ) and  $X_{jt}$  is the  $t$ -th coordinate of the point for column stimulus  $j$  ( $\mathbf{X} = ((X_{jt}))$ ). The asymmetry is captured by the patterns of projections of column points onto row vectors in the joint multidimensional space.

The approaches discussed above implicitly assume that the symmetric and asymmetric aspects of the data are inseparable parts of the same fundamental process. Another approach is to model these aspects of the data directly to reflect different underlying processes. The simplest example of this approach is the common tendency to estimate symmetric dissimilarity as the mean of  $\delta_{ij}$  and  $\delta_{ji}$ . This procedure thus attributes any deviations from symmetry to random error. Bishop, Fienberg, and Holland (1975) describe models that predict the asymmetry in  $\delta_{ij}$  from two functions, one on  $i$  and the other on  $j$ , that may or may not be identical. Constantine and Gower (1978) proposed a method in which  $\delta_{ij}$  is partitioned into symmetric and skew-symmetric orthogonal components. While the symmetric part is represented by some established distance-based method, the skew-symmetric part is represented by points whose relationships are interpreted in terms of areas of triangles. Holman (1979) proposed a series of models using a matrix decomposition approach. These models proposed by Holman represent the data as a monotonic combination of a symmetric function on pairs of stimuli, and a "bias function" on individual stimuli. His models make no prior assumption about the symmetric function; however, all the models assume that the bias function is one-dimensional, and they impose additional conditions on the bias function. Many of the models proposed by Holman concerning the matrix decomposition approach are mathematically elegant, but generally do not provide any theoretical basis for the analysis and geometric representation of the skew-symmetric part.

Perhaps a more appropriate manner to account for asymmetry is to model it directly into the derived metric space (e.g., Krumhansl, 1978; Nakatani, 1972). This amounts to "redefining" distances in the space so as to alleviate the symmetry con-

straint. This may be accomplished by superimposing an additional structural property inherent in the space onto the basic distance model. This is exactly what Krumhansl conceptualized in her distance-density model, where the Euclidean distance model is augmented by the spatial density of the points in the surrounding configuration. (DeSarbo, Manrai, & Burke, 1990, have operationalized Krumhansl's distance-density model in a nonspatial, hierarchical clustering context; DeSarbo and Manrai (1992) generalize the methodology to spatial models.) Based on a principle similar to Krumhansl's distance-density hypothesis, Saito (1986) has proposed a metric MDS procedure for *symmetric* dissimilarity data. His model can be written as

$$\delta_{jk} \cong d_{jk} - e_j - e_k,$$

where  $d_{jk}$  is the Euclidean distance between stimuli  $j$  and  $k$ , and the  $e$ 's are stimulus specific (row and column) scalar constants. (Okada & Imaizumi, 1987, have recently developed a nonmetric extension of this type of model.) With certain restrictions on the parameters, Saito's model represents stimuli as regions (rather than points) in a psychological space. He conjectures proximity judgment to correspond to inter-region distance (rather than inter-point distance). While an interesting supposition, it requires further empirical support.

Note that virtually all of the techniques discussed above are data analytic procedures that attempt to fit a particular model structure to proximity data. In most cases, there is no theoretical or empirical evidence to support such specific model fitting. For example, we have no evidence to suggest that subjects cognitively perform unfolding analyses or project points onto vectors when eliciting their (asymmetric) proximity judgments. The popularity of these procedures is mainly due to the parsimonious geometrical/spatial structures generated to summarize the structure in the data. Tversky's (1977) contrast model, however, introduces a theoretical motivation for the potential causes of the asymmetry. This theory discusses asymmetry from the notion of a contrast of the common and distinctive features of stimuli. Our research goal is to devise a MDS-based procedure that is more theoretically justifiable. We wish to incorporate Tversky's (1977) contrast model within a new MDS procedure for the spatial analysis of asymmetric proximity data.

#### 4. The TSCALE Procedure

Tversky's (1977) contrast model is quite general and can potentially explain a variety of empirical findings. Unfortunately, this generality creates difficulties when operationalizing the model, and attempts at estimating the model have been somewhat limited. Two general methods have been advanced to directly estimate the model's parameters. The first, described in Gati and Tversky (1984), estimates the weight of common to distinctive features by manipulating the independent components of separable, controlled stimuli (e.g., schematic faces, landscapes), and comparing the resulting similarity ratings. The second approach relies on memory probes to measure  $f(I)$  and  $f(J)$ . In Tversky (1977), for example, memory probe estimates of common and distinctive features were correlated with subjects' similarity ratings of vehicles. Product-moment correlations revealed that common features increased similarity while distinctive features decreased similarity, supporting the contrast model. Using a similar procedure, Johnson (1986) found support for the contrast model using memory probes and multiple regression to estimate the models' parameters across similarity, dissimilarity, and subject-referent judgment tasks.

However, both of these approaches require assumptions regarding feature structures (Tversky, 1977). Specifically, both approaches presume a one-to-one correspon-

dence between either an experimental manipulation (Gati & Tversky, 1984) or a memory probe (M. D. Johnson, 1986; Tversky, 1977), and the feature salience measures in the contrast model:  $f(I)$  and  $f(J)$ . As perceptual variables, these feature measures are not often directly observable. One cannot assume that an experimental manipulation always will result in a different internal feature structure, or that a change in feature structure always will be reflected in a memory probe. At the same time, these theoretical feature measures may be treated as latent variables that can be indirectly observed or manifested at an empirical level. Attaching empirical meaning to these latent variables would require either alternative/redundant indicators or conceptually independent sub-dimensions of the constructs. Another problem is that, while experimental feature manipulations are limited to separable stimuli, using memory probes to measure the salient features of natural stimuli is problematic. The time involved in both administering and coding memory probes makes them prohibitive in most applications, and the coding rules are themselves subjective.

The other alternative is to scale the model directly from the observed perceptions of proximity. Scaling methods have been developed that operationalize special cases of the model. For example, Shepard and Arabie's (1979) additive clustering procedure, ADCLUS, operationalizes a common-features model. Assuming  $f(I) = f(J)$ , similarity is treated as a linear function of the measure of  $i$  and  $j$ 's common features (see Tversky, 1977). Sattath and Tversky (1977) describe an alternative technique, ADDTREE, for estimating rooted additive trees with a path length metric. An additive tree is a special case of the contrast model where similarity is a function of distinctive features, and it is assumed that symmetry and the triangle inequality hold. An extension of ADDTREE, called EXTREE (Corter & Tversky, 1986), accommodates non-nested feature structures and is well-suited for representing stimuli with nominal factorial structures. However, like ADDTREE, EXTREE is also a distinctive-features model. More recently, Manrai and Manrai (1989) present an operationalization of a ratio model of proximity related to a special case of Tversky's ratio contrast model with  $\alpha = \beta = \theta = 1$ .

To summarize, Tversky's contrast model captures a number of alternative similarity relationships. Unfortunately, direct attempts at estimating the model are limited or problematic, and more indirect scaling procedures only represent special cases of the underlying model. In the next section, we propose a new spatial MDS procedure for the analysis of (generally asymmetric) three-way, two-mode proximity data. The procedure, called TSCALE, estimates Tversky's contrast model directly from observed dissimilarity data and allows for a variety of different versions of the model.

### *The TSCALE Model*

Whether cognitively represented using features or dimensions, the properties of stimuli can be captured at a more abstract level using continuous dimensions. This abstraction implies an integration of information into an underlying latent dimensional representation (M. D. Johnson & Fornell, 1987). TSCALE starts with the presumption of an underlying or latent dimensional representation ( $\mathbf{X}$ ). Let  $t$  index derived latent dimensions;  $i, j$  index stimuli;  $r$  index replications (e.g., subjects);  $X_{it}$  = the  $t$ -th coordinate for stimulus  $i$  ( $X_{it} \geq 0$ );  $\delta_{ijr}$  = the observed dissimilarity value on the  $r$ -th replication between the two stimuli  $i$  and  $j$ ; and  $\hat{\delta}_{ijr}$  = the model predicted dissimilarity value between the two stimuli  $i$  and  $j$  for the  $r$ -th replication.

When cognitively representing and processing stimuli, subjects often adopt more feature-based representations. As Tversky (1977) notes, when faced with a task, "we extract and compile from our data base a limited list of relevant features on the basis of which we perform the required task" (p. 329). Garner (1978) similarly argues that inherently dimensional stimuli may be more naturally processed using features as a



special case (also, see Prinz & Scheerer-Neumann, 1974). Conceptually, features are considered a special case of more general dimensional representations; dimensions can refer to any attribute, feature or distinction that can serve as a basis for discriminating between two stimuli (Lopes & Johnson, 1982; also, see Krumhansl, 1978). Therefore, modifying the notion of perceptual "quanta" introduced by Manrai (1986) and Manrai and Sinha (1989), we assume that the latent dimensional structure has a corresponding representation such that

$$f(I \cap J) = \sum_{t=1}^T \theta_r \min(X_{it}, X_{jt}), \quad (3)$$

$$f(I - J) = \sum_{t=1}^T \alpha_r (X_{it} - X_{jt})_+, \quad (4)$$

and

$$f(J - I) = \sum_{t=1}^T \beta_r (X_{jt} - X_{it})_+, \quad (5)$$

where  $(a - b)_+ = \max(a - b, 0)$ . Accordingly, the *common* features of two stimuli are represented as the minimum or intersection along various dimensions (as in (3)), while the *distinctive* features of two stimuli are represented as differences (as in (4) and (5)). Note that this specification is applicable to attribute representations that are either quantitative or qualitative at a given level of abstraction. Both types of attributes are captured at a more abstract or latent scale level by continuous, quantitative dimensions. The transformation of dimensions to features is also congruent with Tversky's notion that subjects extract and compile features from an available base of information.

This allows for an extension of Tversky's model to a latent dimensional structure where, for example,

$$\hat{\delta}_{ijr} = \sum_{t=1}^T \alpha_r (X_{it} - X_{jt})_+ + \sum_{t=1}^T \beta_r (X_{jt} - X_{it})_+ - \sum_{t=1}^T \theta_r \min(X_{it}, X_{jt}), \quad (6)$$

for a linear contrast analog, or,

$$\hat{\delta}_{ijr} = \frac{\sum_{t=1}^T \alpha_r (X_{it} - X_{jt})_+ + \sum_{t=1}^T \beta_r (X_{jt} - X_{it})_+}{\sum_{t=1}^T \alpha_r (X_{it} - X_{jt})_+ + \sum_{t=1}^T \beta_r (X_{jt} - X_{it})_+ + \sum_{t=1}^T \theta_r \min(X_{it}, X_{jt})}, \quad (7)$$

for a ratio, distinctive features model, where

- $\alpha_r$  denotes the impact or salience that is distinctive to the first stimulus,  $i$ , in pair  $ij$ , presented on the  $r$ -th replication (e.g., subject  $r$ );
- $\beta_r$  denotes the impact or salience that is distinctive to the second stimulus,  $j$ , in pair  $ij$ , presented on the  $r$ -th replication; and
- $\theta_r$  denotes the impact or salience that is common to the stimulus pair  $ij$  presented on the  $r$ -th replication.

It is assumed that  $0 \leq \delta_{ijr}, \hat{\delta}_{ijr} \leq 1$  for the ratio version of the contrast model in (7). As in Tversky (1977), we assume  $X_{jt}, \alpha_r, \beta_r, \theta_r \geq 0$  in both (6) and (7) above. Note, there is a multiplicative indeterminacy in the estimation of  $\mathbf{X}$ ,  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\beta}$ , and  $\boldsymbol{\theta}$  in both (6) and (7). More specifically, one can multiply  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\beta}$ , and  $\boldsymbol{\theta}$  by some nonzero constant and then divide  $\mathbf{X}$  by this constant without any effect to  $\hat{\delta}_{ijr}$ .

Notice that the measures defined above in (3) through (5) on the latent structure ( $\mathbf{X} = ((X_{it}))$ ) capture the same information as the  $f$  function in Tversky's model, which reflects the salience or prominence of the various features in any given task. From a researcher's perspective,  $\mathbf{X}$  represents the latent properties underlying a given set of stimuli in a given judgment task. This latent structure will differ for different stimuli in the same task (e.g., conceptual versus perceptual stimuli), and for the same stimuli in different tasks (e.g., similarity versus dissimilarity judgments of conceptual stimuli). Meanwhile, the parameters  $\alpha_r$ ,  $\beta_r$ , and  $\theta_r$  reflect the degree to which the various common and distinctive aspects of the latent dimensional structure actually surface and affect  $\delta_{ijr}$ . Here, these parameters represent (a) the degree to which a latent feature effect can be manipulated asymmetrically within any given task (i.e.,  $\alpha_r$  versus  $\beta_r$ ), (b) the relative impact of common versus distinctive features on the proximity judgments for any given replication (i.e.,  $\alpha_r + \beta_r$  versus  $\theta_r$ ), and (c) differences in the impact of common or distinctive features from replication to replication (e.g., individual differences:  $\theta_1$  versus  $\theta_2$ ).

### Estimation

We wish to estimate  $\mathbf{X} = ((X_{it}))$ ,  $\boldsymbol{\alpha} = ((\alpha_r))$ ,  $\boldsymbol{\beta} = ((\beta_r))$ , and  $\boldsymbol{\theta} = ((\theta_r))$ , given  $\underline{\mathbf{A}} = ((\delta_{ijr}))$  and a value of  $T$ , to minimize the following root-mean-error sum-of-squares:

$$\Phi = \left[ \sum_{i \neq j}^N \sum_{r=1}^R \frac{[\delta_{ijr} - \hat{\delta}_{ijr}]^2}{RN(N-1)} \right]^{1/2}, \quad (8)$$

where  $\hat{\delta}_{ijr}$  is given by (6) or (7). The APL code of TSCALE<sup>1</sup> is currently written for  $\hat{\delta}_{ijr}$  defined by (6), given the popularity of the linear contrast model and certain computational advantages involved in the ability to utilize multiple regression in the estimation of  $\boldsymbol{\alpha}$ ,  $\boldsymbol{\beta}$ , and  $\boldsymbol{\theta}$ . This minimization problem in (8) is decomposed into an alternating (conditional) least-squares procedure involving two major phases:

1. *Phase I: Estimate X.* Given the computational complexity of the analytical derivatives of  $\Phi$  in (8) with respect to  $\mathbf{X}$ , we utilize a conjugate gradient procedure involving forward-finite difference numerical approximations of the derivatives (see Gill, Murray, & Wright, 1981; also, see Rao, 1984 for a more complete discussion of the benefits of finite difference approximations over analytical derivatives in various optimization scenarios). Here, such numerical approximations are obtained via

$$\frac{\partial \Phi}{\partial X_{it}} \approx \frac{\Phi(X_{it}) - \Phi(X_{it} + \varepsilon)}{\varepsilon}, \quad (9)$$

where (8) is substituted into (9) with  $\varepsilon = .001$  based on empirical evidence we derived from the analysis of several synthetic data sets. Note that  $\mathbf{X}$  is normalized to constant length prior to each iteration so that the scale of  $\mathbf{X}$  is stable across different applications. For sake of convenience, assume that the entire set of  $\mathbf{X}$  parameters to be estimated in

<sup>1</sup> An APL listing of TSCALE is available from the senior author.

iteration MIT are contained in the vector  $\mathbf{X}^{(\text{MIT})}$  and that  $\nabla\Phi$  is the vector of partial derivatives for this set of parameters. The conjugate gradient procedure can be briefly summarized (see Rao, 1984) as follows:

1. Start with initial parameter estimates  $\mathbf{X}^{(1)}$  (default option is to generate  $\mathbf{X}$  initially via a singular value decomposition of  $\Delta$ ); set the iteration counter (MIT) = 1.
2. Set the first search direction  $\mathbf{S}^{(1)} = -\nabla\Phi^{(1)}$ .
3. Find  $\mathbf{X}^{(2)}$  according to the relation

$$\mathbf{X}^{(2)} = \mathbf{X}^{(1)} + u^{(1)}\mathbf{S}^{(1)}, \quad (10)$$

where  $u^{(1)}$  is the optimal step length in the direction  $\mathbf{S}^{(1)}$ . The optimal step size is found by quadratic interpolation methods. Set MIT = 2.

4. Calculate  $\nabla\Phi^{(\text{MIT})}$  and set

$$\mathbf{S}^{(\text{MIT})} = -\nabla\Phi^{(\text{MIT})} + \frac{[\nabla\Phi^{(\text{MIT})}]'[\nabla\Phi^{(\text{MIT})}]}{[\nabla\Phi^{(\text{MIT}-1)}]'[\nabla\Phi^{(\text{MIT}-1)}]} \mathbf{S}^{(\text{MIT}-1)}. \quad (11)$$

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

$$\mathbf{X}^{(\text{MIT}+1)} = \mathbf{X}^{(\text{MIT})} + u^{(\text{MIT})}\mathbf{S}^{(\text{MIT})}. \quad (12)$$

6. if  $\mathbf{X}^{(\text{MIT}+1)}$  is optimal, stop. Otherwise, set MIT = MIT + 1 and go to Step 4 above (i.e., undertake another iteration).

It has been demonstrated 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 typically will find the global optimum for a quadratic function in  $Q$  steps, where  $Q$  is the number of parameters to be solved. This conjugate gradient method is particularly useful for optimizing functions of several parameters since it does not require the storage of any matrices (as is necessary in quasi-Newton and second derivative methods). However, as noted by Powell (1977), the rate of convergence of the algorithm is linear only if the iterative procedure is "restarted" occasionally (i.e., returning to step b above). Restarts have been implemented in the algorithm automatically, depending on successive improvement in the objective function within this Phase I estimation. However, the maximum MIT value typically is set at 5 (based on the analyses of several synthetic data sets) to reduce computational efforts, and restarting thus is rarely necessary.

Note, the present formulation presents some theoretical difficulties given the potential discontinuities inherent in (3), (4), and (5), when  $X_{it} = X_{jt}$ . As such, a gradient (analytical or numerical) based procedure for such nonlinear estimation is "theoretically" incorrect since the gradients would not be defined at such points of equality. Subgradient optimization (see Shor, 1979) would be more appropriate given such problems, although there are associated difficulties here also. Namely, subgradients defined at such points of equality are not necessarily unique, and deriving them in this context is quite difficult. Also, given the form of (3), (4), and (5), it is difficult to show that (8) is everywhere convex, a necessary step in demonstrating the global properties of subgradients.

*2. Phase II: Estimate  $\alpha$ ,  $\beta$ , and  $\theta$ .* These three sets of multiplicative constants for (6) are estimated by constrained multiple regression using a modification of the Lawson and Hansen (1972) procedure. In the linear specification of TSCALE in (6),

nonnegative estimates of  $\alpha_r$ ,  $\beta_r$ , and  $\theta_r$ , for  $r = 1, \dots, R$ , can be obtained by solving  $R$  linear least-squares problems with linear inequality constraints. We define:

$$\begin{aligned} \mathbf{h}_r &= \text{vec}(\delta_{ijr}) = \text{an } N(N-1) \times 1 \text{ vector composed of the } i \neq j \text{ elements } (i, j = \\ &\quad 1, \dots, N) \text{ in the } r\text{-th replication or slice of } \Delta; \\ \phi_r &= (\alpha_r, \beta_r, \theta_r) \text{ parameter vector;} \\ \mathbf{E}_r &= (\mathbf{E}_r^{(1)}, \mathbf{E}_r^{(2)}, \mathbf{E}_r^{(3)}), \text{ where} \\ \mathbf{E}_r^{(1)} &= \text{vec}(\sum_{i=1}^T (X_{it} - X_{jt})_+), \text{ for } i \neq j, (i, j = 1, \dots, N); \\ \mathbf{E}_r^{(2)} &= \text{vec}(\sum_{i=1}^T (X_{jt} - X_{it})_+), \text{ for } i \neq j, (i, j = 1, \dots, N); \\ \mathbf{E}_r^{(3)} &= \text{vec}(\sum_{i=1}^T \min(X_{it}, X_{jt})), \text{ for } i \neq j, (i, j = 1, \dots, N). \end{aligned}$$

We can then reformulate this estimation problem in terms of  $R$  nonnegative least-squares problems

$$\text{Minimize } \|\mathbf{E}_r \phi_r - \mathbf{h}_r\| \text{ subject to } \phi_r \geq 0, \text{ for } r = 1, \dots, R,$$

which trivially can be shown to conditionally (holding  $\mathbf{X}$  fixed) minimize (8). The algorithm, briefly outlined below, follows directly from the Kuhn-Tucker conditions for constrained minimization. For a given  $r$ , form the  $N(N-1) \times 3$  matrix of "independent variables,"  $\mathbf{E}_r$ , and the  $N(N-1) \times 1$  vector of dissimilarities (acting as the dependent variable)  $\mathbf{h}_r$ . In the description below, the  $3 \times 1$  vectors  $\mathbf{w}_r$  and  $\mathbf{z}_r$  provide working spaces. Index sets  $P_r$  and  $Z_r$  will be defined and modified in the course of execution of the algorithm. Parameters indexed in the set  $Z_r$  will be held at the value zero. Parameters indexed in the set  $P_r$  will be free to take values greater than zero. If a parameter takes a nonpositive value, the algorithm will either move the parameter to a positive value or else set the parameter to zero and move its index from set  $P_r$  to set  $Z_r$ . On termination,  $\phi_r$  will be the solution vector and  $\mathbf{w}_r$  will be the dual vector.

1. Set  $P_r := \text{Null}$ ,  $Z_r := \{1, 2, 3\}$ , and  $\phi_r := \mathbf{0}$ .
2. Compute the vector  $\mathbf{w}_r := \mathbf{E}_r'(\mathbf{h}_r - \mathbf{E}_r \phi_r)$ .
3. If the set  $Z_r$  is empty or if  $w_{rj} \leq 0$  for all  $j \in Z_r$ , go to Step 12.
4. Find an index  $a \in Z_r$  such that  $w_{ra} = \max\{w_{rj} : j \in Z_r\}$ .
5. Move the index  $a$  from set  $Z_r$  to set  $P_r$ .
6. Let  $\mathbf{E}_P^{(r)}$  denote the  $N(N-1) \times 3$  matrix defined by

$$\text{Column } j \text{ of } \mathbf{E}_P^{(r)} := \begin{cases} \text{column } j \text{ of } \mathbf{E}_r & \text{if } j \in P_r \\ 0 & \text{if } j \in Z_r. \end{cases}$$

Compute the vector  $\mathbf{z}_r$  as a solution of the least-squares problem  $\mathbf{E}_P^{(r)} \mathbf{z}_r \cong \mathbf{h}_r$ . Note that only the components  $z_{rj}$ ,  $j \in P_r$ , are determined by this problem. Define  $z_{rj} = 0$  for  $j \in Z_r$ .

7. If  $z_{rj} > 0$  for all  $j \in P_r$ , set  $\phi_r := \mathbf{z}_r$  and go to Step 2.
8. Find an index  $v \in P_r$  such that  $\phi_{rv}/(\phi_{rv} - z_{rv}) = \min\{\phi_{rj}/(\phi_{rj} - z_{rj}) : z_{rj} \leq 0, j \in P_r\}$ .
9. Set  $Q_r := \phi_{rv}/(\phi_{rv} - z_{rv})$ .
10. Set  $\phi_r := \phi_r + Q_r(\mathbf{z}_r - \phi_r)$ .
11. Move from set  $P_r$  to set  $Z_r$  all indices  $j \in P_r$  for which  $\phi_{rj} = 0$ . Go to Step 6.
12. End.

On termination, the solution vector  $\phi_r$  satisfies

$$\phi_{rj} > 0, \quad j \in P_r; \tag{13}$$

and

$$\phi_{rj} = 0, \quad j \in Z_r, \quad (14)$$

and is a solution vector to the least-squares problem

$$\mathbf{E}_P^{(r)} \underline{\phi}_r \cong \mathbf{h}_r. \quad (15)$$

The dual vector  $\mathbf{w}_r$  satisfies

$$w_{rj} = 0, \quad j \in P_r; \quad (16)$$

and

$$w_{rj} \leq 0, \quad j \in Z_r, \quad (17)$$

where

$$\mathbf{w}_r = \mathbf{E}_r'(\mathbf{h}_r - \mathbf{E}_r \underline{\phi}_r). \quad (18)$$

Equations (13), (14), (16), (17), and (18) constitute the Kuhn-Tucker conditions characterizing a solution vector  $\underline{\phi}_r$  for this constrained least-squares problem. Equation (15) is a consequence of (14), (16), and (18). These twelve steps are then repeated for the next value of  $r$ .

#### *Program Options*

TSCALE allows for a variety of linear contrast models to be fit to the empirical  $\delta_{ijr}$ . The  $\alpha$ ,  $\beta$ , and  $\theta$  parameters can be specified as given constants, indexed by replication (e.g.,  $\alpha_r$ ) or indexed by replication and dimension ( $\alpha_{rt}$ ). Constraints can be imposed to estimate symmetric versions of these models (i.e.,  $\alpha = \beta$ ), as well as confirmatory models using equality constraints to allow for competing models akin to those of Restle (1961) discussed earlier. Options exist to hold  $\mathbf{X}$  fixed, estimate  $\mathbf{X}$ , or reparameterize  $\mathbf{X}$  via

$$X_{it} = \sum_{m=1}^M H_{im} \gamma_{mt}, \quad (19)$$

where  $m$  indexes 1, . . . ,  $M$  observed descriptor variables or features,  $H_{im}$  is the value of the  $m$ -th descriptor variable or feature for stimulus  $i$ , and  $\gamma_{mt}$  is the coefficient denoting the contribution of the  $m$ -th descriptor variable or feature on the  $t$ -th dimension. Such a reparameterization can aid in the interpretation of the derived dimensions (see Bentler & Weeks, 1978; Bloxom, 1978; de Leeuw & Heiser, 1980; Noma & Johnson, 1977), and can replace the post-analysis (property-fitting) methods often used in an attempt to interpret subsequent results. As in CANDELINC (Carroll, Pruzansky, & Kruskal, 1980), Three-Way Multivariate Conjoint Analysis (DeSarbo, Carroll, Lehmann, & O'Shaughnessy, 1982), and GENFOLD2 (DeSarbo & Rao, 1984), these reparameterization options can be used to examine what effect such features/attributes have on the derived solution. This option can be utilized to directly relate observed features to the derived dimensions. Note, when such reparameterization options are utilized, the number of background variables in the linear function cannot exceed the number of entities that exist for those variables. For example, if  $N$  stimuli have  $M$  attributes,  $N \geq M$  since, at most, only  $NT$  coordinates can be identified (depending on respective parameter indeterminacies associated with the particular model by being estimated with TSCALE). Thus, in most applications, such a reparameterization actually improves the degrees of freedom of the model by reducing the number of parameters to be estimated.

Options also exist for fixing any desired parameter set in the analysis to specified values.

### Goodness-of-Fit

A variety of goodness-of-fit measures are computed in TSCALE between the observed data,  $\delta_{ijr}$ , and the model predicted values,  $\hat{\delta}_{ijr}$ , given by (6) or (7). The three primary measures are

1. A root-mean-square (RMS) measure as in (8),
2. A sum-of-squares-accounted-for (SSAF) measure:

$$\text{SSAF} = \frac{\left( \sum_{i \neq j}^{NN} \sum_r^R \delta_{ijr} \hat{\delta}_{ijr} \right)^2}{\sum_{i \neq j}^{NN} \sum_r^R \delta_{ijr}^2 \sum_{i \neq j}^{NN} \sum_r^R \hat{\delta}_{ijr}^2}, \quad (20)$$

3. A variance-accounted-for (VAF) measure:

$$\text{VAF} = 1 - \frac{\left( \sum_{i \neq j}^{NN} \sum_r^R (\delta_{ijr} - \hat{\delta}_{ijr})^2 \right)}{\left( \sum_{i \neq j}^{NN} \sum_r^R (\delta_{ijr} - \delta_{..r})^2 \right)}, \quad (21)$$

where  $\delta_{..r}$  is defined as the average dissimilarity value in the  $r$ -th replication. These measures are also calculated by replication for  $r = 1, \dots, R$ .

### Monte Carlo Results

In preliminary testing of this two-stage algorithm on several synthetic data sets (with no error), it was found that the procedure recovered the true  $\mathbf{X}$ ,  $\alpha$ ,  $\beta$ , and  $\theta$  values in all but two cases where locally optimal solutions occurred. Given this propensity, as well as concerns of model indeterminacy raised by Sattath and Tversky (1987, in a symmetric version of the linear contrast model), and by one helpful reviewer, we decided to conduct a more thorough test of the proposed TSCALE procedure. Some ten independent factors were specified as having potential effects on the performance of TSCALE. These ten factors are shown in Table 1, and reflect various aspects of the input data, the specific TSCALE model, algorithm control parameters, and error. The first seven factors and their respective levels are somewhat self-explanatory; the final three independent factors require some explanation in terms of operationalization. The features model factor specifies magnitude relationships between the  $\alpha$ ,  $\beta$ , and  $\theta$ . Initially, simulated values for these three sets of parameters were all generated from the same uniform distribution. In the mixed-features model, these values were unchanged. In the common-features model,  $\theta$  was multiplied by a large positive constant to enlarge the impact of common features;  $\alpha$  and  $\beta$  were unchanged. In the distinctive-features model,  $\alpha$  and  $\beta$  were multiplied jointly by a large positive constant, and  $\theta$  was unaltered to enlarge the impact of the distinctive features.

The three levels of error are also shown in Table 1, where  $\sigma_2 = 2\sigma_1$ , and  $\sigma_1 = 1$ . Error was added to the synthetically generated proximities to form  $\delta_{ijr}$ , after  $\mathbf{X}$ ,  $\alpha$ ,  $\beta$ ,

**TABLE 1**  
**Independent Factors for the TSCALE Monte Carlo Analysis**

Factor	Level
1. Number of Dimensions	T=2 T=4
2. Number of Stimuli	N=7 N=12
3. Number of Subjects	R=3 R=8
4. Input Proximity Data	Asymmetric Symmetric
5. Stimulus Configuration	Estimate $\underline{X}$ Reparameterize $\underline{X} = H\underline{\gamma}$
6. Start for $\underline{X}$	Random Torgerson method
7. $\epsilon$ -value	$\epsilon = .01$ $\epsilon = .001$
8. Features Model	Common Distinctive Mixed
9. Error to $\underline{\Delta}$	None $N(0, \sigma_1)$ $N(0, \sigma_2)$
10. Variation of $\underline{X}$	Small Medium Large

and  $\theta$  were randomly generated. Finally,  $X$  was initially generated from a uniform distribution:  $U(0, 3)$ . We then investigated the performance of TSCALE for various levels of dispersion of these coordinates. For small dispersion, the coordinates were left unaltered. For medium dispersion, these initial values were squared. For large dispersion, these initial values were cubed.

Some seven dependent measures of algorithm performance were specified measuring computational effort, overall goodness-of-fit, and parameter recovery:

- $Y_1$  = number of major iterations required for convergence;
- $Y_2$  = SSAF ( $\delta_{ijr}, \hat{\delta}_{ijr}$ );
- $Y_3$  = VAF ( $\delta_{ijr}, \hat{\delta}_{ijr}$ );
- $Y_4$  = RMS ( $\alpha, \hat{\alpha}$ );
- $Y_5$  = RMS ( $\beta, \hat{\beta}$ );
- $Y_6$  = RMS ( $\theta, \hat{\theta}$ );
- $Y_7$  = RMS ( $X, \hat{X}$ ) after appropriate permutation and normalization.

The ten independent factors were combined in a  $2^7 3^3$  fractional factorial design (Adelman, 1962) for main effects estimation (an initial full factorial design with factors 4, 8, 9, and 10 was performed at the request of one reviewer and produced only a few significant interaction terms). Such efficient designs have been employed previously in methodology testing by DeSarbo (1982) and DeSarbo and Carroll (1985). Twenty seven trials with a single replication were specified as in a conjoint analysis (see Green and Rao, 1971) where such fractional factorial designs, converted to dummy variables, are used as independent variables in a regression context to explore the effects of designated experimental factors on a selected dependent variables. Here, the seven dependent measures stated above were analyzed via multiple regression to examine the relative effects of the various levels of the ten independent factors. Note, the intercept value would thus represent the predicted value of the specified dependent measure when all independent variables were fixed at their "base" or zero-coded value(s). The remaining regression coefficients denote the additions to or subtraction from this base predicted value.

Table 2 presents the regression results for these seven dependent measures. The only significant regression equation concerns  $Y_1$ , the number of major iterations required for convergence, where  $R = .8$ , symmetric model, reparameterized  $\mathbf{X}$ , and Torgerson start each significantly reduce this dependent variable across all factor levels. While there are scattered and isolated significant effects among the remaining six dependent variables, none of these equations are significant, indicating somewhat consistent fitting over a variety of different model and data types (arc sine transformations were also applied to the second and third dependent variables given their restricted range and no major changes in these regression results occurred). While promising, these results are preliminary given the simple design utilized for Monte Carlo testing.

To examine the *absolute* performance of TSCALE in this Monte Carlo experiment, Table 3 displays the list of means and standard deviations (in parentheses) for each of the seven dependent variables across all factor levels. Overall, there appears to be somewhat consistent fitting over the various factor levels. However, a few cells in this table prove interesting. For example, the choice of  $\epsilon$ -value appears to render rather different SSAF ( $\delta_{ijr}$ ,  $\hat{\delta}_{ijr}$ ) means. The reparameterization option for  $\mathbf{X}$  tends to result in smaller RMS ( $\mathbf{X}$ ,  $\hat{\mathbf{X}}$ ), perhaps due to the fact that additional information/data is provided with such an option (i.e.,  $\mathbf{H}$ ), and there are typically fewer parameters to estimate in such analyses. Oddly, the intermediate variation of  $\mathbf{X}$  level tends to result in better recovery of the model parameters. Finally, the recovery of  $\theta$  appears to be somewhat better in a common features model, as opposed to a distinctive features model.

## 5. TSCALE Applications

### *Method and Procedure*

Proximity data on perceptions of fast-food restaurants and cola soft-drinks were collected and analyzed via TSCALE to illustrate the procedure. Table 4 presents the two sets of stimuli, each consisting of 12 alternatives. Each of the stimuli in the two sets were readily available or accessible in the geographical area of the study. Subjects (graduate students) were screened for a minimal level of awareness of these stimuli. Each set of stimuli contains twelve alternatives, with subjects providing both ( $i, j$ ) and ( $j, i$ ) comparisons. A subject/referent format was adopted.

Subjects were asked to rate the similarity of  $i$  to  $j$  and  $j$  to  $i$  on a scale ranging from 0 (very similar) to 10 (very dissimilar) which appeared to the right of each pair. Indi-



TABLE 2  
TSCALE Monte Carlo Regression Results

	No. of Iterations	SSAF ( $\hat{\delta}_{ijr}, \hat{\delta}_{ijr}$ )	VAF ( $\hat{\delta}_{ijr}, \hat{\delta}_{ijr}$ )	RMS( $\alpha, \hat{\alpha}$ )	RMS( $\beta, \hat{\beta}$ )	RMS( $\theta, \hat{\theta}$ )	RMS( $X, \hat{X}$ )
intercept	27.17	0.95	0.98	0.15	0.17	0.02	0.15
T=4	0.83	0.02	0.01	-0.06	-0.03	-0.03	-0.03
N=12	2.33	-0.01	-0.01	-0.02	-0.00	-0.03	-0.04
R=8	-3.67*	0.02	0.01	-0.04	-0.06	0.02	-0.03
Symmetric	-6.17**	-0.01	-0.01	0.02	0.00	0.02	0.02
$\underline{X} = \underline{H}\gamma$	-7.83**	0.02	0.01	-0.02	-0.05	-0.03	-0.08*
Torgerson start	-4.50*	-0.01	0.00	0.01	0.06	0.03	0.02
$\epsilon = .001$	-1.83	0.01	0.00	-0.04	0.01	-0.05	-0.03
Distinctive	-2.78	0.01	0.01	-0.01	-0.05	0.11*	0.01
Mixed	1.11	0.03*	0.02*	-0.01	-0.03	0.05	-0.02
$N(0, \sigma_1)$	-1.33	-0.02	-0.01	0.00	0.00	-0.00	-0.01
$N(0, \sigma_2)$	-2.67	-0.02	-0.01	0.03	0.03	0.08	0.03
Medium	2.00	0.01	-0.00	-0.06	-0.09*	-0.00	-0.02
Large	2.67	0.01	-0.00	-0.02	-0.04	0.01	0.03
S.E.	4.06	0.03	0.01	0.08	0.08	0.08	0.07
$R^2$	0.82	0.58	0.64	0.43	0.60	0.63	0.57
adj. $R^2$	0.65	0.16	0.29	-0.14	0.19	0.27	0.14
F-Ratio	4.64**	1.38	1.81	0.76	1.47	1.72	1.31

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

vidual level confidence, experience, and preference measures were also collected between the  $i$  to  $j$  and  $j$  to  $i$  judgments to minimize carry-over, and the sex of each respondent was recorded. The subjects rated their confidence in evaluating each alternative on a scale from 0 (not at all confident) to 10 (very confident). Frequency of consumption experience for each stimulus was rated on a five point categorical scale (every day, every week, every month, every year, never). The subjects' preference for each alternative was also measured on an eleven point scale ranging from -5 (very much dislike the alternative) to 0 (neutral) to 5 (very much like the alternative).

TABLE 3  
Monte Carlo Means and Standard Deviations by Factor Levels

Factor	Level	Number of Iterations	SSAF ( $\delta_{ij}^2, \delta_{ij}^2$ )	VAF ( $\delta_{ij}^2, \delta_{ij}^2$ )	RMS( $\alpha, \hat{\alpha}$ )	RMS( $\beta, \hat{\beta}$ )	RMS( $\theta, \hat{\theta}$ )	RMS( $X, \hat{X}$ )
1. Number of Dimensions	T=2	17.5 (6.6)*	0.91 (0.15)	0.96 (0.07)	0.09 (0.08)	0.08 (0.07)	0.11 (0.10)	0.12 (0.09)
	T=4	18.1 (7.1)	0.94 (0.09)	0.98 (0.04)	0.04 (0.03)	0.08 (0.11)	0.06 (0.05)	0.08 (0.04)
2. Number of Stimuli	N=7	16.9 (7.3)	0.93 (0.14)	0.97 (0.07)	0.08 (0.08)	0.09 (0.09)	0.10 (0.10)	0.11 (0.08)
	N=12	19.3 (5.1)	0.90 (0.12)	0.96 (0.05)	0.04 (0.05)	0.07 (0.07)	0.07 (0.06)	0.09 (0.07)
3. Number of Subjects	R=3	18.5 (6.6)	0.91 (0.14)	0.96 (0.07)	0.08 (0.08)	0.10 (0.10)	0.09 (0.10)	0.11 (0.09)
	R=8	16.1 (6.9)	0.94 (0.12)	0.97 (0.05)	0.05 (0.06)	0.05 (0.03)	0.10 (0.08)	0.09 (0.05)
4. Input Proximity Data	Asymmetric	19.3 (5.3)	0.93 (0.10)	0.97 (0.04)	0.07 (0.08)	0.09 (0.09)	0.07 (0.08)	0.10 (0.08)
	Symmetric	14.4 (8.1)	0.90 (0.17)	0.95 (0.09)	0.08 (0.07)	0.08 (0.07)	0.12 (0.09)	0.13 (0.08)
5. Stimulus Configuration	Estimate X	19.2 (6.1)	0.89 (0.15)	0.95 (0.08)	0.08 (0.08)	0.10 (0.09)	0.11 (0.09)	0.14 (0.07)
	Reparameterize $X=Hy$	14.7 (7.0)	0.98 (0.02)	0.99 (0.01)	0.06 (0.05)	0.05 (0.06)	0.06 (0.09)	0.04 (0.04)
6. Start for $X$	Random	19.6 (6.3)	0.91 (0.15)	0.96 (0.07)	0.06 (0.08)	0.06 (0.07)	0.08 (0.08)	0.11 (0.08)
	Torgerson Method	14.0 (6.0)	0.94 (0.08)	0.97 (0.04)	0.08 (0.06)	0.12 (0.11)	0.10 (0.11)	0.10 (0.07)
7. $\epsilon$ -value	$\epsilon=.01$	19.3 (6.7)	0.95 (0.05)	0.98 (0.02)	0.07 (0.08)	0.07 (0.07)	0.10 (0.10)	0.10 (0.08)
	$\epsilon=.001$	14.4 (5.7)	0.85 (0.20)	0.93 (0.10)	0.07 (0.06)	0.12 (0.10)	0.07 (0.06)	0.11 (0.08)
8. Features Model	Common	15.2 (6.3)	0.85 (0.17)	0.93 (0.09)	0.07 (0.07)	0.10 (0.10)	0.05 (0.06)	0.12 (0.08)
	Distinctive	17.7 (7.8)	0.97 (0.02)	0.99 (0.01)	0.07 (0.07)	0.06 (0.05)	0.14 (0.08)	0.11 (0.06)
	Mixed	20.2 (4.8)	0.94 (0.12)	0.98 (0.05)	0.07 (0.08)	0.08 (0.09)	0.08 (0.10)	0.09 (0.10)
9. Error to $\Delta$	None	19.7 (4.7)	0.92 (0.18)	0.96 (0.09)	0.07 (0.07)	0.08 (0.08)	0.07 (0.08)	0.10 (0.08)
	$N(0, \sigma_1)$	17.2 (7.1)	0.91 (0.12)	0.97 (0.05)	0.04 (0.05)	0.05 (0.06)	0.07 (0.06)	0.09 (0.06)
	$N(0, \sigma_2)$	16.2 (7.6)	0.92 (0.07)	0.97 (0.03)	0.10 (0.08)	0.12 (0.11)	0.13 (0.11)	0.12 (0.09)
10. Variation of $X$	Small	15.4 (7.5)	0.92 (0.12)	0.97 (0.05)	0.09 (0.08)	0.12 (0.06)	0.08 (0.10)	0.10 (0.05)
	Medium	19.5 (5.6)	0.95 (0.06)	0.98 (0.02)	0.02 (0.02)	0.02 (0.02)	0.08 (0.08)	0.08 (0.05)
	Large	18.1 (6.4)	0.89 (0.18)	0.95 (0.09)	0.09 (0.09)	0.11 (0.11)	0.11 (0.09)	0.14 (0.11)
Overall		17.7 (6.8)	0.92 (0.13)	0.97 (0.06)	0.07 (0.07)	0.08 (0.09)	0.09 (0.09)	0.11 (0.08)

\*Standard deviations in parentheses.

**TABLE 4**  
**Sets of Stimuli for the Two TSCALE Applications**

<u>Fast Food Restaurants</u>	<u>Plot Codes</u>	<u>Cola Soft Drinks</u>
1. Arby's	A	1. Coke Classic
2. A&W Drive-in	B	2. Pepsi
3. Burger King	C	3. Dr. Pepper
4. Church's Fried Chicken	D	4. Diet Pepsi
5. Hardee's	E	5. Diet Faygo
6. Kentucky Fried Chicken	F	6. Diet Big K
7. Little Caesar's	G	7. New Coke
8. McDonald's	H	8. Cherry Coke
9. Pizza Hut	I	9. Diet Coke
10. Taco Bell	J	10. Faygo Cola
11. Wendy's	K	11. Big K Cola
12. White Castle	L	12. Tab

A pencil and paper format was used to collect data in a classroom setting. For each set of stimuli, half of the subjects gave, in order, the  $i$  to  $j$  proximity ratings, stimulus confidence, experience and preference, and then the  $j$  to  $i$  proximity ratings. The other half gave the  $j$  to  $i$  proximity ratings first and the  $i$  to  $j$  proximity ratings last. The presented pairs of stimuli appeared in random order for the first group of subjects, while this order was reversed for the second group. A total of 71 graduate students participated in the study, 37 for the fast-foods and 34 for the cola soft-drinks. One of the soft-drink subjects was eliminated from the analysis for failing to complete the questionnaire.

### *Hypotheses*

Based on the recent literature in consumer psychology, product experience should play an important role in consumer judgment (see Bettman, 1986), prompting initial focus on this variable. As experience grows, consumers first group products into categories and then distinguish among the brands within those categories (Howard, 1977). Therefore, the weight that consumers place on distinctive features (i.e., the distinguishing features of the brands), should increase with experience. However, any increase in the salience of distinctive features is likely maximized at an intermediate level of experience. Researchers have shown that consumers' recallable knowledge of brand-level differences not only levels off, but may decrease somewhat at higher levels of experience or product familiarity (Bettman & Park, 1980; E. J. Johnson & Russo, 1984). As very experienced consumers rely more on stored overall evaluations and engage in

less evaluative processing, attribute differences are less salient. Therefore, we expect a general increase in distinctive-feature salience with experience that is maximized at an intermediate experience level.

The two sets of stimuli themselves should differ with respect to the importance of experience. The fast-food restaurants represent a more dissimilar and complex set of stimuli than the cola soft-drinks. Any differences across experience levels should be more pronounced for the fast-food stimuli. Quite simply, the cola soft-drinks require little, if any, experience to enable perception of their salient differences (e.g., diet versus nondiet), especially given the massive advertising performed by the major national brands. Finally, given the subject/referent format of the proximity judgments, the distinctive features of the subject should be somewhat greater than the distinctive features of the referent for both sets of stimuli (Tversky, 1977).

### *Fast-Food Restaurant Analysis*

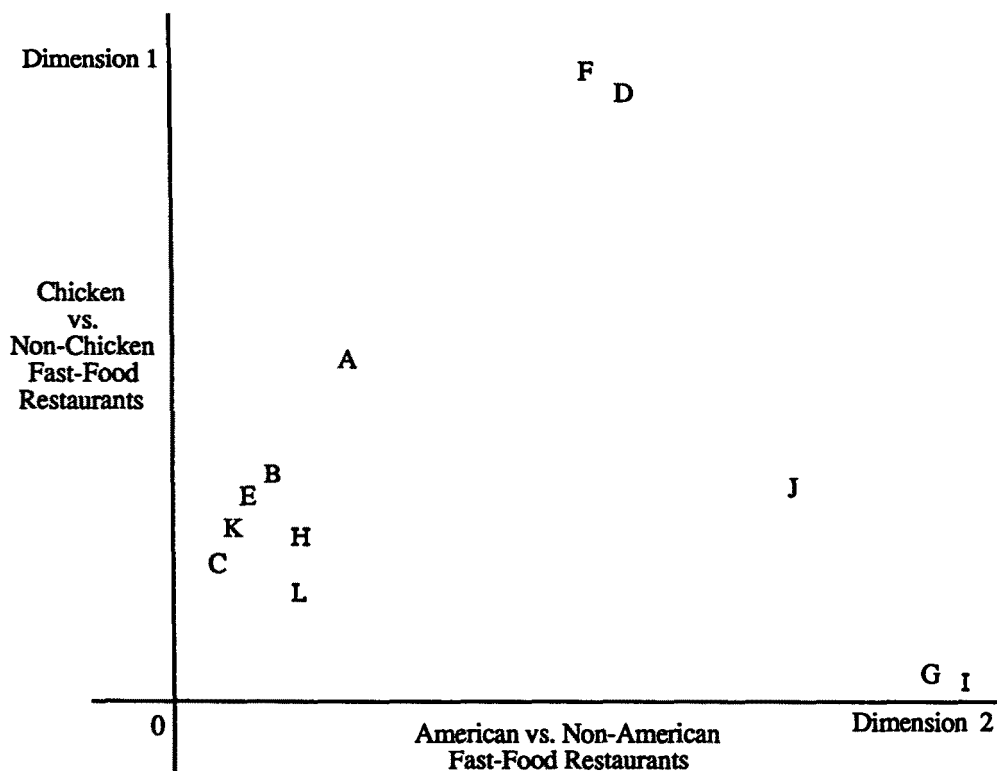
Given the consumer psychology literature mentioned above on the role of experience in consumer perception, three clusters of subjects initially were created prior to TSCALE analysis representing different levels (high, medium, and low) of fast-food restaurant experience (later, the same was done for cola soft-drinks). The five level categorical scales were assigned simple quantitative values from 0 (never eat at the restaurant) to 4 (eat at the restaurant every day), and aggregated across stimuli for analysis. A K-MEANS (Hartigan, 1975) split of the subjects' experience clustered the subjects into low ( $n = 10$ ), intermediate ( $n = 20$ ), and high ( $n = 7$ ) experience groups. We use these clusters as replications to illustrate the TSCALE analyses in detail.

A number of different TSCALE models were estimated in 1-3 dimensions. Based upon goodness-of-fit, interpretation, and parsimony, the two-dimensional, linear TSCALE solution with  $\alpha$ ,  $\beta$ , and  $\theta$  parameters indexed by experience group was selected as best representing the structure in the data, producing an overall sum-of-squares-accounted-for value of 0.990 and a variance-accounted-for value of 0.891. Note, because of a general scale indeterminacy associated with this version of the linear model, we normalized  $\text{vec}(\mathbf{X})$  to constant length and multiplied the corresponding  $\alpha$ ,  $\beta$ , and  $\theta$  by this constant.

The two-dimensional space for the fast food restaurants is presented in Figure 1. The solution is quite interpretable. Dimension one represents a chicken/nonchicken dimension. While hamburgers are the primary product at McDonald's, Burger King, et cetera, chicken is the primary product at Kentucky Fried Chicken and Church's. Arby's has an intermediate location on this dimension given its roast beef image together with the recent addition of a number of chicken items to its menu. The second dimension separates the pizza and Mexican restaurants (serving food of non-American origin) from the chicken and hamburger (traditionally, American fast-food) restaurants.

The common and distinctive-feature parameters, presented in Table 5, vary predictably with experience. The relative weights on distinctive versus common features, presented in the column to the far right, are 10.105, 12.194, and 11.427, from low to high experience levels. As expected, experience has a general positive effect on relative distinctive feature salience and, consistent with previous research, this distinctive feature salience is highest at an intermediate level of experience (Bettman & Park, 1980; E. J. Johnson & Russo, 1984).

However, the results concerning subject-referent asymmetry are mixed. Overall, there is little asymmetry in these proximity data as noted by the similar magnitudes for  $\alpha_r$  and  $\beta_r$  within each experience level. (A Constantine and Gower, 1979, decomposition of the pooled  $\delta_{ijr}$ , averaged over these three experience groups, revealed that the symmetric portion accounts for 99.2% of the variance in this data. The corresponding



Plot Code

A - Arby's	G - Little Caesar's
B - A&W Drive-in	H - McDonald's
C - Burger King	I - Pizza Hut
D - Church's Fried Chicken	J - Taco Bell
E - Hardee's	K - Wendy's
F - Kentucky Fried Chicken	L - White Castle

FIGURE 1.  
The two-dimensional TSCALE solution for fast-food restaurants.

figures by experience group are 96.4%, 98.9%, and 98.1%, respectively, for the low, medium, and high experience groups.) The asymmetry ( $\alpha_r \neq \beta_r$ ) was in the predicted direction only for the high experience group, although the  $\alpha_r - \beta_r$  difference is still quite small. Those subjects with relatively little experience with the fast-food restaurants appear to focus somewhat on the referent of comparison, although these  $\alpha_r - \beta_r$  differences are again small. Recall that the similarity rating scales appeared to the immediate right of each stimulus pair in the data collection task. These subjects' lack of experience, combined with the proximity of the referent to the ratings scale, may produce a focus on the referent rather than the subject. In hindsight, two other factors may have contributed to this lack of symmetry. Unlike the original Tversky and Gati (1978) study, the collected similarity judgments here were not embedded in sentence

TABLE 5  
TSCALE Parameter Values

Stimuli	Experience	$\alpha_r$	$\beta_r$	$\theta_r$	$\alpha_r + \beta_r - \theta_r$
	Low	5.741	5.819	1.455	10.105
Fast-Foods	Intermediate	6.084	6.111	0.001	12.194
	High	5.743	5.685	0.001	11.427
	Low	4.873	4.762	3.739	5.896
Soft Drinks	Intermediate	4.755	4.743	4.174	5.324
	High	4.871	4.713	4.345	5.239
	Gender	$\alpha_r$	$\beta_r$	$\theta_r$	$\alpha_r + \beta_r - \theta_r$
	Male	4.809	4.758	4.642	4.925
Soft Drinks	Female	5.314	5.205	2.964	7.555

form for each pair of stimuli. In addition, the  $(i, j)$  and  $(j, i)$  stimulus pairs were collected within, rather than between, subjects.

#### *Cola Soft-Drinks Analysis*

Following the same procedure, TSCALE was estimated for three clusters of subjects representing three levels of cola soft-drink experience: low ( $n = 5$ ), intermediate ( $n = 18$ ), and high ( $n = 9$ ). (One subject had a significantly higher level of experience than all three groups and was eliminated.) The two-dimensional solution produced a sum-of-squares-accounted-for value of 0.984 (variance accounted-for = 0.884), utilizing a similar model structure as in the fast-food restaurants. The two dimensions extracted separate diet from nondiet brands, and regional from national brands. The estimated parameters, reported in Table 5, reveal that the effects for experience are predictably negligible compared to the fast-food restaurants. There is a very slight decrease in the relative weight of total distinctive versus common features (equal to 5.896, 5.324, and 5.239 for the low, intermediate, and high experience subjects, respectively). Here, distinctive-feature salience is greatest at the lowest experience level, although these differences are again quite small. The less complex nature of the cola soft-drinks may

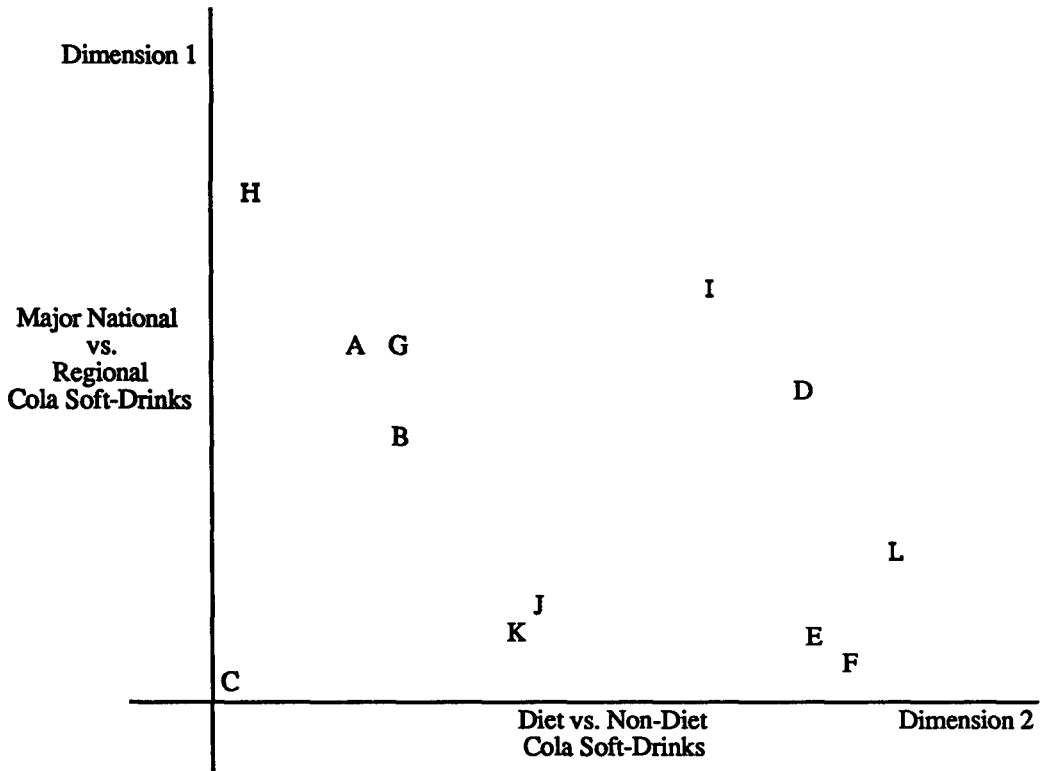
be reflected in these small differences across experience levels. This could also be due to the massive advertising campaigns undertaken by the national brands that would increase overall awareness of the differences between the brands. There are somewhat expected results regarding the subject/referent manipulation. The distinctive features of the subject were slightly more salient than those of the referent for all three experience groups, but these  $\alpha_r - \beta_r$  differences are very small, indicating a lack of asymmetry.

A review of the consumer psychology literature suggests that contrasting male and female perceptions often provides a parsimonious view of product perceptions (e.g., Doyle & Hutchinson, 1973; Gentry, Doering, & O'Brien, 1978; Meyers-Levy, 1988; Turlle & Falconer, 1972). This prompted a second analysis of the cola soft-drink perceptions along these lines. TSCALE was applied to the two gender groups (18 males and 15 females). Again, a variety of models were estimated in 1-3 dimensions. Based on goodness-of-fit and interpretation, the two-dimensional, linear model (sum-of-squares-accounted-for = 0.988, variance-accounted-for = 0.911) is deemed as parsimoniously representing the structure in the data. The resulting two-dimensional solution is presented in Figure 2. Dimension two separates the diet from the nondiet drinks. Dimension one distinguishes the regional from the major national brands, with two exceptions: TAB (L) and Dr. Pepper (C)—two poorly positioned, weak national brands. There are very systematic differences between the males and females with respect to feature salience, as seen at the bottom of Table 5. Looking across the model parameter sets, the females appear to have greater relative distinctive feature salience than males. Finally, consistent with our initial expectations and Tversky's prediction,  $\alpha_r > \beta_r$  for each gender on each dimension, although these differences are again small. (The corresponding Constantine and Gower, 1978, decomposition on the pooled  $\delta_{ijr}$ , averaged over these two gender groups, reveal that the symmetric portion accounts for 98.5% of the variance in this data. The corresponding figures by gender group are 97.7% for males and 97.3% for females.)

## 6. Conclusions

We have presented an alternative spatial MDS procedure for the analysis of three-way, two-mode proximity data based on a latent dimension version of Tversky's (1977) contrast model of similarity. We have demonstrated that such a procedure can be applied to empirical proximity data that do not necessarily satisfy the metric axioms. There are a variety of other potential applications of the TSCALE procedure for the analysis of symmetric and asymmetric proximity data in the domain of psychology. For example, TSCALE could be employed gainfully in the study of confusions data, word association tasks and semantic domains, olfaction, psychoacoustics, and other substantive problems in psychology where MDS has been found useful. Perhaps the most fruitful applications of TSCALE will arise in examinations of stimuli that vary on some underlying or abstract dimensions, yet are likely represented by subjects using dichotomous features (e.g., conceptual stimuli—Pruzansky, Tversky, & Carroll, 1982; Tversky, 1977). In addition, stimulus domains that exhibit substantial individual differences in perception may benefit from TSCALE analysis.

There are a number of avenues for future research. One, more rigorous testing of the procedure is required with more complex Monte Carlo analyses employing larger designs with full factorial specifications, replications, additional factors, etc. Second, an extension of TSCALE to tree structures is needed. Similarly, the development of hybrid models combining the fitting of both tree structures and Euclidean spaces would also be a useful direction for research. Finally, as mentioned above, the use of the



Plot Code

A - Coke Classic	G - New Coke
B - Pepsi	H - Cherry Coke
C - Dr. Pepper	I - Diet Coke
D - Diet Pepsi	J - Faygo Cola
E - Diet Faygo	K - Big K Cola
F - Diet Big K	L - Tab

FIGURE 2.

The two-dimensional TSCALE solution for cola soft-drinks. The product names shown are trademarks of their respective companies.

procedure in different and more varied applications would also prove beneficial in accurately evaluating the value of this new procedure.

References

- Abelson, R. P., & Levi, A. (1985). Decision making and decision theory. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology, Vol. 3* (pp. 231-309). New York: Academic Press.
- Addelman, S. (1962). Orthogonal main-effect plans for asymmetrical factorial experiments. *Technometrics*, 4, 21-46.
- Bennett, J. F., & Hays, W. L. (1960). Multidimensional unfolding: Determining the dimensionality of ranked preference data. *Psychometrika*, 25, 27-43.



- Bentler, P. M., & Weeks, D. G. (1978). Restricted multidimensional scaling models. *Journal of Mathematical Psychology*, 17, 138–151.
- Bettman, J. R. (1986). Consumer psychology. *Annual Review of Psychology*, 37, 257–289.
- Bettman, J. R., & Park, C. W. (1980). Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: A protocol analysis. *Journal of Consumer Research*, 7, 234–248.
- Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: Theory and practice*. Cambridge: MIT Press.
- Bloxom, B. (1978). Constrained multidimensional scaling in  $N$  spaces. *Psychometrika*, 43, 397–408.
- Bush, R. R., & Mosteller, F. (1951). A model for stimulus generalization and discrimination. *Psychological Review*, 58, 413–423.
- Carroll, J. D. (1980). Models and methods for multidimensional analysis of preferential choice (or other dominance) data. In E. D. Lantermann & H. Feger (Eds.), *Similarity and Choice* (pp. 234–289). Bern: Huber.
- 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.
- Chino, N. (1978). A graphical technique for representing the asymmetric relationship between  $N$  objects. *Behaviormetrika*, 5, 23–40.
- Chino, N. (1979). Extension of Chino's ASYMSCAL into higher dimensions. *Paper presented at the annual meeting of the 7th Behaviormetric Society of Japan*, Japan.
- Chino, N. (1990). A generalized inner product model for the analysis of asymmetry. *Behaviormetrika*, 27, 25–46.
- Constantine, A. G., & Gower, J. C. (1978). Graphical representation of asymmetric matrices. *Applied Statistics*, 27, 297–304.
- Coombs, C. H. (1950). Psychological scaling without a unit of measurement. *Psychological Review*, 57, 148–158.
- Corter, J. E., & Tversky, A. (1986). Extended similarity trees. *Psychometrika*, 51, 429–451.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34, 571–582.
- de Leeuw, J., & Heiser, W. (1980). Multidimensional scaling with restrictions on the configuration. In P. R. Krishnaiah (Ed.), *Multivariate analysis-V* (pp. 501–522). New York: North-Holland.
- DeSarbo, W. S. (1982). GENCLUS: New models for general nonhierarchical cluster analysis. *Psychometrika*, 47, 446–469.
- 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., & O'Shaughnessy, J. (1982). Three-way multivariate conjoint analysis. *Marketing Science*, 1, 323–350.
- DeSarbo, W. S., & Manrai, A. K. (in press). A new multidimensional scaling methodology for the analysis of asymmetric proximity data in marketing research. *Marketing Science*.
- DeSarbo, W. S., Manrai, A. K., & Burke, R. R. (1990). A nonspatial methodology for the analysis of two-way proximity data incorporating the distance-density hypothesis. *Psychometrika*, 55, 229–253.
- 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.
- Doyle, P., & Hutchinson, P. (1973). Individual differences in family decision making. *Journal of the Marketing Research Society*, 15, 193–206.
- Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review*, 86, 465–485.
- Eister, H., & Ekman, G. (1959). A mechanism of subjective similarity. *Acta Psychologica*, 16, 1–10.
- Garner, W. R. (1978). Aspects of a stimulus: Features, dimensions, and configurations. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 99–133). New Jersey: Erlbaum.
- Gati, I., & Tversky, A. (1982). Representations of qualitative and quantitative dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 325–340.
- Gati, I., & Tversky, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgments. *Cognitive Psychology*, 16, 341–370.
- Gentry, J. W., Doering, M., & O'Brien, T. V. (1978). Masculinity and femininity factors in product perception and self image. In H. Keith Hunt (Ed.), *Advances in consumer research* (pp. 326–332). Ann Arbor, MI: Association for Consumer Research.
- Gill, P. E., Murray, W., & Wright, M. H. (1981). *Practical optimization*. Orlando, FL: Academic Press.

- Gower J. C. (1978). Unfolding: Some technical problems and novel uses. *Presented at European meeting of Psychometric and Mathematical Psychology*, Uppsala.
- Green, P. E., & Rao, V. R. (1971). Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research*, 8, 355-363.
- Gregson, R. A. M. (1975). *Psychometrics of similarity*. New York: Academic Press.
- Harshman, R. A. (1975). Models for the analysis of asymmetrical relationships among  $N$  objects or stimuli. *Presented at US-Japan seminar on multidimensional scaling*, University of California, San Diego, La Jolla.
- Hartigan, J. A. (1975). *Clustering algorithms*. New York: John Wiley & Sons.
- Himmelfarb, D. M. (1972). *Applied nonlinear programming*. New York, McGraw Hill.
- Holman, E. W. (1979). Monotonic models for asymmetric proximities. *Journal of Mathematical Psychology*, 20, 1-15.
- Howard, J. (1977). *Consumer behavior: Application of theory*. New York: McGraw-Hill.
- Johnson, E. J., & Russo, J. E. (1984). Product familiarity and learning new information. *Journal of Consumer Research*, 11, 542-550.
- Johnson, M. D. (1986). Consumer similarity judgments: A test of the contrast model. *Psychology and Marketing*, 3, 47-60.
- Johnson, M. D., & Fornell, C. (1987). The nature and methodological implications of the cognitive representation of products. *Journal of Consumer Research*, 14, 214-228.
- Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32, 241-254.
- Krumhansl, C. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, 85, 445-463.
- Kruskal, J. B., Young, F. W., & Seery, J. B. (1973). *How to use KYST, a very flexible program to do multidimensional scaling and unfolding*. Bell Laboratories, Murray Hill, NJ.
- Lawson, C. L., & Hanson, R. J. (1972). *Solving least squares problems*. Englewood Cliffs, NJ: Prentice-Hall.
- Lopes, L. L., & Johnson, M. D. (1982). Judging similarity among strings described by hierarchical trees. *Acta Psychologica*, 51, 13-26.
- Manrai, A. K. (1986). *Similarity, perceptions, and choice*. Unpublished doctoral dissertation, Northwestern University, Evanston, IL.
- Manrai, A. K. & Manrai, L. A. (1989). Mathematical models for relating proximity to multidimensional scaling. In N. Avlonitis (Ed.), *Marketing thought and practice in the 1990's* (pp. 853-868). Athens, Greece: European Marketing Academy, The Athens School of Economics and Business Science.
- Manrai, A. K., & Sinha, P. K. (1989). Elimination by cutoffs. *Marketing Science*, 8, 133-152.
- Meyers-Levy, J. (1988). The influence of sex roles on judgment. *Journal of Consumer Research*, 14, 522-530.
- Nakatani, L. H. (1972). Confusion-choice model for multidimensional psychophysics. *Journal of Mathematical Psychology*, 9, 104-127.
- Noma, E., & Johnson, J. (1977). *Constraining nonmetric multidimensional scaling configurations* (Tech. Rep. 60). The University of Michigan, Human Performance Center, Ann Arbor.
- Okada, A., & Imaizumi, T. (1987). Nonmetric multidimensional scaling of asymmetric proximities. *Behaviormetrika*, 21, 81-96.
- Prinz, W., & Scheerer-Neumann, G. (1974). Component processes in multiattribute stimulus classification. *Psychological Research*, 37, 25-50.
- Powell, M. J. D. (1977). Restart procedures for the conjugate gradient method. *Mathematical Programming*, 12, 241-254.
- Pruzansky, S., Tversky, A., & Carroll, J. D. (1982). Spatial versus tree representations of proximity data. *Psychometrika*, 47, 3-24.
- Rao, S. S. (1984). *Optimization: Theory and applications* (2nd ed.). New York: Wiley.
- Restle, F. A. (1959). A metric and an ordering on sets. *Psychometrika*, 24, 207-220.
- Restle, F. A. (1961). *Psychology of judgment and choice*. New York: Wiley.
- Rosch, E. (1975). Cognitive representation of semantic categories. *Journal of Experimental Psychology: General*, 104, 192-233.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-603.
- Saito, T. (1986). Multidimensional scaling to explore complex aspects in dissimilarity judgment. *Behaviormetrika*, 20, 35-62.
- Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika*, 42, 319-345.
- Sattath, S., & Tversky, A. (1987). On the relation between common and distinctive feature models. *Psychological Review*, 94, 16-22.
- Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 27, 125-140.

- Shepard, R. N., & Arbie, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review*, *86*, 87–123.
- Shor, N. Z. (1979). *Minimization methods for non-differentiable functions*. New York: Springer-Verlag.
- Sjöberg, L. (1972). *A cognitive theory of similarity*. Goteborg Psychological Reports (No. 10).
- Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*, 214–241.
- Takane, Y., Young, F. W., & de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least square method with optimal scaling features. *Psychometrika*, *42*, 7–67.
- Turle, J. E., & Falconer, R. (1972). Men and women are different. *Journal of the Marketing Research Society*, *14*, 111–123.
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, *79*, 281–299.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.
- Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 79–98). New Jersey: Erlbaum.
- Tversky, A., & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review*, *89*, 123–154.
- Tversky, A., & Hutchinson, J. W. (1986). Nearest neighbor analysis of psychological spaces. *Psychological Review*, *93*, 3–22.
- Young, F. W. (1975). An asymmetric Euclidean model for multiprocess asymmetric data. *Presented at US-Japan seminar on multidimensional scaling*, University of California, San Diego, La Jolla.

*Manuscript received 2/16/89*

*Final version received 2/21/91*