

## A NONSPATIAL METHODOLOGY FOR THE ANALYSIS OF TWO-WAY PROXIMITY DATA INCORPORATING THE DISTANCE-DENSITY HYPOTHESIS

WAYNE S. DESARBO

MARKETING AND STATISTICS DEPARTMENTS  
UNIVERSITY OF MICHIGAN

AJAY K. MANRAI AND RAYMOND R. BURKE

MARKETING DEPARTMENT  
WHARTON SCHOOL  
UNIVERSITY OF PENNSYLVANIA

This paper presents a nonspatial operationalization of the Kruschal (1978, 1982) distance-density model of similarity. This model assumes that the similarity between two objects  $i$  and  $j$  is a function of both the interpoint distance between  $i$  and  $j$  and the density of other stimulus points in the regions surrounding  $i$  and  $j$ . We review this conceptual model and associated empirical evidence for such a specification. A nonspatial, tree-fitting methodology is described which is sufficiently flexible to fit a number of competing hypotheses of similarity formation. A sequential, unconstrained minimization algorithm is technically presented together with various program options. Three applications are provided which demonstrate the flexibility of the methodology. Finally, extensions to spatial models, three-way analyses, and hybrid models are discussed.

Key words: asymmetric similarity, ultrametric trees, hierarchical clustering, Kruschal's distance-density model.

### 1. Introduction

The notion of proximity (i.e., similarity or dissimilarity) has played a central role in several theoretical, methodological, and empirical areas of psychological investigation. Cognitive psychologists have collected proximity judgments from subjects in various experimental contexts in order to measure the underlying structure of designated stimuli. Several different models have been typically employed to summarize the data and display the relationships between stimuli. (We shall restrict our attention to two-way, one-mode proximity data, i.e., a single, square matrix of proximities reflecting the degree of association between all pairs of objects or stimuli.) Unfortunately, the vast majority of these procedures are based on the metric axioms which are often found to conflict with empirically observed proximities. We therefore propose a new class of models which, as will be discussed, can accommodate such axiom violations via the incorporation of Kruschal's (1978, 1982) model of distance-density.

According to Carroll, Clark, and DeSarbo (1984) and Carroll and Arabie (1980), models for the analysis of proximity data can be categorized into one of three major classes.

The authors wish to thank three anonymous reviewers and the editor for their insightful comments on a previous draft of this manuscript.

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

### *Spatial Models*

Continuous spatial models of the type typically associated with multidimensional scaling (MDS) usually embed the objects or stimuli of interest in some coordinate space so that a specified measure of distance (e.g., Euclidean) between the points in the space represents the observed proximities among the respective objects. Thus, a metric space is assumed to underlie the proximity data. Distances in the derived space are specified to be related to the corresponding proximities in some simple way (e.g., linearly or monotonically). Thus,

$$\delta_{ij} \cong f(d_{ij}),$$

$$d_{ij} = \left[ \sum_{t=1}^T (x_{it} - x_{jt})^P \right]^{1/P},$$

summarizes the typical spatial/continuous MDS model, where  $\delta_{ij}$  is the empirical, symmetric dissimilarity between stimuli  $i$  and  $j$ ,  $d_{ij}$  is the distance computed in the derived MDS space between coordinate locations for stimuli  $i$  and  $j$ ,  $x_{it}$  is the  $t$ th coordinate for stimulus  $i$ ,  $f$  is some function (e.g., linear, monotone), and  $P \geq 1$  defines the specific form of the Minkowski metric (usually  $P = 2$ ). The problem here is, given the two-way proximity data  $\delta_{ij}$  and a specification of  $f$  and  $P$ , to construct an estimate of the underlying stimulus coordinates  $x_{it}$ . This estimation is typically done to optimize some goodness (badness) of fit index between the actual and predicted proximities (Kruskal 1964a, 1964b; Shepard 1962a, 1962b; Torgerson 1958).

### *Nonspatial models*

The most common nonspatial or network model is a tree, which represents each stimulus as a terminal node in a connected graph without cycles. Each pair of nodes is joined by a unique path so that the relationships between the terminal nodes in the graph reflect the observed proximity relations among the stimuli. A tree structure lends itself to a natural interpretation as a hierarchical clustering scheme (Johnson 1967). Hartigan (1967) was the first to develop an explicit algorithm to optimize a least-squares criterion for fitting an ultrametric tree structure to proximity data. An ultrametric tree structure is obtained by enforcing the ultrametric inequality which implies that given two disjoint clusters, all intra-cluster distances are smaller than all inter-cluster distances, and that all the inter-cluster distances are equal for any given pair of clusters. In fact, several researchers such as Jardine, Jardine, and Sibson (1967), Johnson, and Hartigan independently demonstrated that the hierarchical tree metric is governed by the ultrametric inequality:

$$d_{ij} \leq \max (d_{in}, d_{nj}),$$

which requires that all "triangles" formed by the distances of all triples of points have their two equal sides longer than the third side. Several psychometricians generalized the ultrametric tree representation to additive or free tree representations (Carroll, 1976; Carroll & Chang 1973; Cunningham 1978; Sattath & Tversky 1977). An additive tree is a tree in which the distance between any two nodes is given by the sum of the lengths of the links in the unique path between those nodes. An additive tree directly implies the four-point additivity condition:

$$d_{in} + d_{jk} \leq \max \{(d_{jn} + d_{ik}), (d_{kn} + d_{ij})\}.$$

An additive tree has  $N$  external nodes for  $N$  stimuli, and the lengths of all connecting links are then estimated to minimize a least-squares measure. Basically these methods are metric in that they treat the data as linearly related to the underlying path-length distance. Note that an ultrametric tree is a special case of an additive tree with a distinguished node (named the root) which is equidistant from all external nodes. Carroll and Pruzansky (1975) have shown that an additive tree can be decomposed into an ultrametric tree plus a "bush" or star graph. They have also extended the concept of estimating a least-squares tree from two-way proximity data to the fitting of multiple tree structures (see also Appendix II of Carroll, Clark, & DeSarbo, 1984).

### *Hybrid Models*

Carroll (1976) and Carroll and Pruzansky (1980) have introduced a newer class of models to fit two-way symmetric proximity data. Hybrid models combine the fitting of both tree structure(s) and continuous spatial components so that for a given two-way proximity data set, say, one might fit two tree structures and a two-dimensional Euclidean space. This hybrid model concept was demonstrated with the Rosenberg and Kim (1975) data in Carroll (1976) and Carroll and Pruzansky (1980).

Tversky (1977) challenged the adequacy of these traditional approaches on both theoretical and empirical grounds. He raised doubts about the metric properties of the distance measure as well as the dimensional structure of many spatial models of proximity. Tversky argued that if similarity or dissimilarity could be represented by metric distance via these models, as is the case in these three approaches, then the three axioms of a metric distance function ought to follow, namely:

1. Minimality:  $\delta_{ab} \geq \delta_{aa} = 0.$
2. Symmetry:  $\delta_{ab} = \delta_{ba}.$
3. Triangle Inequality:  $\delta_{ab} + \delta_{bc} \geq \delta_{ac}.$

However, Tversky (1977), Sjöberg (1972), Parducci (1965), Birnbaum (1974), and others have provided empirical evidence and several intuitive examples demonstrating that some measures of similarity (e.g., identification probabilities in recognition experiments, errors of substitutions, and directional ratings of pairs of objects) violate one or more of these metric axioms. Tversky proposed an alternative set-theoretic approach, called the contrast model. Here the similarity between two objects is expressed as a linear combination of the measures of the common and distinctive features of the two objects. When additional assumptions are made concerning the parameters of the model, the contrast model can be shown to account for asymmetric similarity measures, certain effects of stimulus context on similarity, and discrepancies between similarity and difference judgments.

Krumhansl (1978) later introduced the distance-density model of similarity which, as she demonstrated, can also account for many violations of these axioms of proximity. The goal of this research paper is to operationalize the Krumhansl distance-density hypothesis with a methodology for the nonspatial analysis of proximity data. We first present the Krumhansl conceptual distance-density theory of similarity and cite empirical support for her theory. Next, a model is constructed to accommodate various hypotheses about proximity judgment formation. The algorithm and program options are technically described. Three applications of the methodology are presented dem-

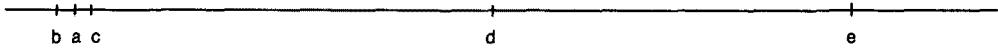


FIGURE 1  
An illustration of Krumhansl's distance-density model in one dimension.

onstrating the flexibility of the approach. Finally, a discussion of model extensions to spatial models, three-way analyses, and hybrid models are provided.

## 2. Krumhansl's Distance-Density Model

Krumhansl (1978) suggested that some of the objections to traditional, geometric models raised by Tversky (1977) may be met if the conventional geometric model is augmented by the assumption that the proximity between stimuli is a function of both (a) the interobject distance in a metric space, and (b) the spatial density of stimuli represented as points in some geometric configuration. The specific conceptual model form suggested by Krumhansl is:

$$S_{xy} = f^{-1}(\tilde{d}_{xy}) \quad (1)$$

where

$$\tilde{d}_{xy} = d_{xy} + a\eta(x) + b\eta(y). \quad (2)$$

In the above distance-density model, the similarity between objects  $x$  and  $y$ ,  $S_{xy}$ , is a monotonic decreasing function of modified distance  $\tilde{d}_{xy}$ . Equation (2) defines the modified distance using ordinary distance  $d_{xy}$  and two density terms  $\eta(x)$  and  $\eta(y)$  defined as the densities or concentrations of objects/stimuli surrounding objects  $x$  and  $y$ , respectively. In (2),  $a$  and  $b$  are (non-negative) parameters in this model, and  $d_{xy}$  is the traditional distance defined by spatial or nonspatial methods. To illustrate this model, Figure 1 presents a one-dimensional spatial representation of some five stimuli:  $a$ ,  $b$ ,  $c$ ,  $d$ , and  $e$ . The theory implies that within dense subregions of a stimulus range, finer discriminations are made than within relatively less dense subregions. Two points in a relatively dense region of a stimulus space would thus have a smaller similarity measure than two points of equal interpoint distance located in a less dense region of the space. Thus, distances between the stimulus pairs would have to be augmented with an additional "penalty" (or increment to distance) term reflecting the fact that objects  $a$ ,  $b$ , and  $c$  are close together in a concentrated subregion.

Note that the modified distance function in (2),  $\tilde{d}_{xy}$ , need not satisfy the metric axioms previously discussed. The distance between an object and itself,  $\tilde{d}_{xx} = (a + b)\eta(x)$ , will in general be greater than zero and will depend on the density of points surrounding object  $x$ , so that the minimality axiom will not hold in general unless  $a = b = 0$  or  $\eta(x) = 0$ . The symmetry axiom also need not hold. Here,  $\tilde{d}_{xy}$  and  $\tilde{d}_{yx}$  are equal if and only if  $a = b$  or  $\eta(x) = \eta(y)$ . According to Tversky (1977), in a directional similarity task, greater emphasis may be placed on one member of the object pair than the other. This emphasis can be reflected in the  $a$  and  $b$  weights, with the spatial density surrounding one object affecting the modified distance measure  $\tilde{d}_{xy}$  more than the density surrounding the second object. If  $a > b$ , then  $\tilde{d}_{xy} > \tilde{d}_{yx}$  when  $\eta(x) \geq \eta(y)$ . However, assuming  $a$  and  $b$  are non-negative, the triangle inequality axiom must hold for  $\tilde{d}_{xy}$ . Since the modified distance function  $\tilde{d}_{xy}$  is assumed to be a linear combination of interpoint distance and spatial density,  $\tilde{d}_{xy} + \tilde{d}_{yz} \geq \tilde{d}_{xz}$ . Krumhansl (1978, 1982) presents a list of studies reporting experimental data which violate these axioms.

Krumhansl (1978) and Appelman and Mayzner (1982) define a variety of density

measures (typically in the context of multidimensional scaling). One suggested measure of density is:

$$\eta(x) = \sum_{\substack{i \in R \\ i \neq x}} \frac{1}{d_{ix}^2},$$

where  $R$  stands for the stimulus set and  $d_{ix}$  is the distance between stimulus  $i$  and stimulus  $x$  in the metric space. Another definition of density provided by Krumhansl is based on the number of objects within a fixed radius of a particular object. Here,

$$\eta(x) = \sum_{i \in R} n_{ix},$$

where,

$$n_{ix} = \begin{cases} 1 & \text{if } d_{ix} \leq r \\ 0 & \text{if } d_{ix} > r, \end{cases}$$

$i$  is a member of stimulus set  $R$ , and  $r$  is a radius constant about  $x$ . Density is therefore a monotonic increasing function of the number of stimuli within the fixed radius. Unfortunately, this definition is not sensitive to the particular distribution of stimuli within this radius.

### 3. Methodology

There has been no known methodological development or operationalization of Krumhansl's (1978) conceptual model of distance-density in a proximity modeling context. Saito (1986) has proposed a metric MDS procedure for modeling two-way, one-mode asymmetric proximities by estimating a set of object constants and a configuration of points ( $x_{it}$ ) in  $T$  dimensions. The author claims that by restricting these object constants to be nonpositive, one can interpret them as density constants. Unfortunately, even with nonpositivity restrictions, it is not clear how one can interpret these object constants in a distance-density context, given that they are otherwise free to vary and are not explicitly related to the density of stimulus points in the derived configuration. Our immediate goal, therefore, is to operationalize Krumhansl's conceptual model of distance-density vis-a-vis a new nonspatial methodology. Later, mention will be made of further extending this methodology to a spatial context.

#### *The Model(s)*

Our nonspatial Krumhansl methodology is developed as follows. Let:

- $i, j = 1, \dots, N$  objects in the set  $R$ ;
- $\delta_{ij}$  = the empirical/observed dissimilarity between objects  $i$  and  $j$ ;
- $\mathbf{\Delta} = ((\delta_{ij}))$  where  $\delta_{ij} \neq \delta_{ji}$  and  $\delta_{ii} \neq 0$  in general;
- $d_{ij}$  = the predicted ultrametric tree distance between objects  $i$  and  $j$ , where  $d_{ij} = d_{ji}$  and  $d_{ii} = 0$ .

Then, we can write the "full" Krumhansl model as:

$$\delta_{ij} = d_{ij} + a_i \eta(i) + b_j \eta(j) + e_{ij}, \quad (3)$$

where

$$\eta(h) = \sum_{\substack{p \in R \\ p \neq h}} \Psi_{ph} = \text{a measure of the density of objects around object } h;$$

$$\Psi_{ph} = \max(0, r_h - d_{ph}) = (r_h - d_{ph})_+;$$

$r_h$  = a threshold value for object  $h$ ;  
 $a_i, b_j$  = multiplicative constants;  
 $e_{ij}$  = error.

Note that this is one of many ways of modeling density. We specified  $\Psi_{ph}$  as such to be congruent with Krumhansl's notion of density as the "number of points within a fixed radius." In addition, the  $\Psi_{ph}$  definition above models the particular concentration or distribution of the stimuli within the radius—an aspect missing from one of Krumhansl's initial density notions previously mentioned. The methodology can accommodate other specifications of density as well.

From a substantive perspective, the categorization literature suggests that there may be discontinuities or thresholds in similarity judgments. As context objects move away from (i.e., become less similar to) the pair of objects being judged, at some point they may not be noticed by the respondent or may not be considered relevant to that judgment. This threshold point may be related to the boundaries of basic level concepts for categorization (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Tversky & Hemenway 1984). Objects tend to be spontaneously grouped at this level and subsequent discrimination tends to be easier than at other levels. Categories at the basic level maximize within-category similarity relative to between-category similarity (Mervis & Rosch 1981). Evidence suggests that the basic level categories are defined primarily by concrete perceptual attributes (Murphy & Smith 1982). The threshold values  $r_h$  might be interpreted as the distances between these basic level categories.

### Estimation

We wish to estimate  $a_i, b_j, r_h$ , and  $d_{ij}$ , given  $\Delta$ , where  $d_{ij}$  satisfies the ultrametric inequality, in order to minimize:

$$Z(a_i, b_j, r_h, d_{ij}) = \sum_i \sum_j^N [\delta_{ij} - \hat{\delta}_{ij}]^2 = \sum_i \sum_j^N e_{ij}^2 \quad (4)$$

s.t.:

$$d_{ij} \leq \max(d_{ik}, d_{jk}) \quad \forall i, j, k \quad (5)$$

where

$$\hat{\delta}_{ij} = d_{ij} + a_i \eta(i) + b_j \eta(j). \quad (6)$$

Note that Carroll (1976) and Carroll and Pruzansky (1980) have utilized an exterior penalty function (a constrained optimization procedure that converts a difficult nonlinear constrained problem into a sequence of unconstrained problems initially utilizing parameter estimates that do not satisfy all of the constraints) and mathematical programming methods to minimize (4) subject to (5) for fitting ultrametric trees to ordinary two-way (one mode, symmetric) proximity data. However, in our metric methodology, there are additional parameters to estimate including  $a_i, b_j$ , and  $r_h$ . Note the  $\eta(h)$  must be greater than zero in order for these parameters to be estimable (e.g., the derivatives of  $Z$  with respect to  $r(h)$  are undefined otherwise). We therefore utilize a sequential unconstrained minimization technique (Fiacco & McCormick, 1968) which combines aspects of exterior and interior or barrier penalty functions. In general terms, we now transform the estimation problem:

$$\text{Min } \Phi = Z(a_i, b_j, r_h, d_{ij}) + \lambda P(d_{ij}) - \mu G(r_h, d_{ij}), \quad (7)$$

where the first term on the right-hand side of (7) relates to (4) allowing for the recovery of the  $\delta_{ij}$ , the second term on the right-hand side of (7) is to insure that the  $d_{ij}$  satisfy the ultrametric inequality, and the third term on the right hand side of (7) insures that  $\eta(h) > 0$  so that the respective parameter derivatives exist. This is equivalent to a strict enforcement of the distance-density hypothesis for all stimuli, that is,  $\eta(h) > 0$  for all  $h$ . That is,

$$Z(a_i, b_j, r_h, d_{ij}) = \sum_i \sum_j^N e_{ij}^2 \tag{8}$$

$$P(d_{ij}) = \sum_i^{N-1} \sum_j^N \sum_k^N w_{ik}^j (d_{ij} - d_{jk})^2 \tag{9}$$

$\begin{matrix} i < j \\ j \neq i, k \end{matrix}$

$$w_{ik}^j = \begin{cases} 1 & \text{if } d_{ik} \leq \min(d_{ij}, d_{jk}) \\ 0 & \text{else} \end{cases} \tag{10}$$

$$G(r_h, d_{ij}) = \sum_{i=1}^N \log \left[ (N-1)r_h - \sum_{\substack{p \in R \\ p \neq h}} d_{ph} + \varepsilon \right] \tag{11}$$

$$\lambda > 0, \quad \lambda \rightarrow \infty \text{ during iterations}$$

$$\mu > 0, \quad \mu \rightarrow 0 \text{ during iterations}$$

$\varepsilon =$  a small positive constant (0.001).

Note how (11) operates to insure  $\eta(h) \geq 0$ . Given the particular operating definition of density in (3):

$$\begin{aligned} \eta(h) &= \sum_{\substack{p \in R \\ p \neq h}} \Psi_{ph} \\ &= \sum_{\substack{p \in R \\ p \neq h}} (r_h - d_{ph})_+ , \end{aligned}$$

for  $\eta(h) > 0$ , then, expanding, one obtains:

$$\sum_{\substack{p \in R \\ p \neq h}} r_h - \sum_{\substack{p \in R \\ p \neq h}} d_{ph} > 0$$

or

$$(N-1)r_h - \sum_{\substack{p \in R \\ p \neq h}} d_{ph} > 0,$$

or, to accommodate the inequality constraint ( $\eta(h) \geq 0$ ) in the SUMT estimation procedure:

$$(N-1)r_h - \sum_{\substack{p \in R \\ p \neq h}} d_{ph} + \varepsilon \geq 0,$$

where  $\varepsilon$  is the small positive constant defined earlier. One can see how the ‘‘barrier function’’ in (11) performs to satisfy these constraints. If the expression above becomes zero or negative, taking the log makes the expression  $\rightarrow -\infty$  which makes the objective

function  $\Phi$  in (7) unusually large. Thus, in the terms of Rao (1984), a "barrier" is constructed in order to avoid a search over the infeasible parameter space.

Some important numerical issues must be noted at this point. First,  $P(d_{ij})$  must be normalized to prevent the solution of  $d_{ij} = 0, \forall i, j$ , which trivially satisfies the ultrametric inequality and minimizes  $Z$ . It is interesting to note that the discontinuities introduced in enforcing (9) (given the fact that  $w_{ik}^j$  is not a continuous argument as seen in (10)) do not appear to disrupt the gradient based procedure (Carroll & Pruzansky, 1980) as does the problem associated with the threshold constants. Second, it is a good idea to normalize all three parts of the right side in (7) so that, in addition to preventing degenerate solutions, the three parts vary in roughly the same range, although this is not nearly as important an issue for the third part of (11). As mentioned, the second part of (7) must be suitably normalized in order to prevent the procedure of setting all  $d_{ij} = 0$ . The third part of (7) requires no normalization since its purpose is to go to infinity if the threshold constants do not satisfy the constraints. Third, one must obtain feasible starting values for  $r_h$  and  $d_{ij}$  with respect to (11), since it is an interior penalty method. Usually  $r_h = \max_p d_{ph} + \varepsilon$  will suffice. Fourth, there are an excessive number of parameters to estimate if we attempt to solve for  $a_i, b_j, r_h$ , and  $d_{ij}$  given  $\delta_{ij}$ . As we will discuss in the next section, constraints are typically placed on either  $a_i, b_j$ , and/or  $r_h$  to reduce the number of parameters to be estimated. For example, we will set  $a_i = a, b_j = b$ , and  $r_h = r$  and reduce the total number of parameters to be estimated given the small data sets discussed. Finally, one could justifiably argue for redundancy in the model in (2) since one could increase the contribution of the density part of the model by either increasing  $a_i$  and  $b_j$  or by increasing  $r_h$ . To simplify this discussion, we assume  $r_h = r > d_{ij}, \forall i, j, a_i = 1$ , and  $b_j = b$  in the asymmetric Krumhansl model. Suppose one enlarges the threshold constant,  $r$ , by a positive constant,  $c$ . How must the  $a$  and  $b$  parameters change to maintain the same density contribution? That is, what are the values of  $\alpha$  and  $\beta$  below that result in the same contribution of density to the modified distance in expression (2)? Thus:

$$\begin{aligned} a\eta(i) &= \alpha\eta^*(i) \\ &= \alpha \sum_{\substack{p \in R \\ p \neq i}} (r + c - d_{pi}) \\ &= \alpha[(N - 1)c + \eta(i)]. \end{aligned}$$

So, for the above equality to hold,

$$\alpha \equiv \frac{a\eta(i)}{c(N - 1) + \eta(i)},$$

and in a similar fashion for  $b\eta(j) \equiv \beta\eta^*(j)$ ,

$$\beta \equiv \frac{b\eta(j)}{c(N - 1) + \eta(j)}.$$

Thus, the contribution by the density parts of the modified distance in (2) can be increased by either increasing  $r$  and/or enlarging  $a$  and  $b$ . Similar arguments can be made concerning the absolute magnitude of the estimated ultrametric distances and  $r, a$ , and  $b$ .

With these issues aside, we can reformulate the objective function:



$$\text{Min } \Phi = \frac{Z(a_i, b_j, r_h, d_{ij})}{\gamma} + \frac{\lambda^{IT} P(d_{ij})}{\theta} - \mu^{IT} G(r_h, d_{ij}), \quad (12)$$

where:

$$\gamma = \sum_i \sum_j^N (\delta_{ij} - \delta_{..})^2$$

$$\delta_{..} = \frac{1}{N^2} \sum_i \sum_j^N \delta_{ij}$$

$$\theta = \sum_{i < j}^N (d_{ij} - d_{..})^2$$

$$d_{..} = \frac{2}{N(N-1)} \sum_{i < j}^N d_{ij}$$

$$\lambda^{IT+1} = c\lambda^{IT}, \quad c = 10$$

$$\mu^{IT+1} = s\mu^{IT}, \quad s = .01 \text{ (Rao, 1984)}$$

$IT$  = iteration counter.

The method of conjugate gradients (Rao, 1984) is utilized as the search procedure. The set of partial derivatives are:

$$\frac{\partial \Phi}{\partial a} = \frac{-2 \sum_i \sum_j^N (\delta_{ij} - \hat{\delta}_{ij}) \cdot \eta(i)}{\gamma} \quad (13)$$

$$\frac{\partial \Phi}{\partial a_i} = \frac{-2 \sum_{j=1}^N (\delta_{ij} - \hat{\delta}_{ij}) \cdot \eta(i)}{\gamma} \quad (14)$$

$$\frac{\partial \Phi}{\partial b} = \frac{-2 \sum_i \sum_j^N (\delta_{ij} - \hat{\delta}_{ij}) \cdot \eta(j)}{\gamma} \quad (15)$$

$$\frac{\partial \Phi}{\partial b_j} = \frac{-2 \sum_{i=1}^N (\delta_{ij} - \hat{\delta}_{ij}) \cdot \eta(j)}{\gamma} \quad (16)$$

$$\frac{\partial \Phi}{\partial r} = \frac{-2 \sum_i \sum_j^N (\delta_{ij} - \hat{\delta}_{ij}) \cdot \left[ a \sum_{k \neq i} t_{ik} + b \sum_{h \neq j} t_{hj} \right]}{\gamma} - \sum_{v=1}^N \frac{\mu(N-1)}{(N-1)r - \sum_{\substack{p \in R \\ p \neq v}} d_{pv} + \epsilon} \tag{17}$$

$$\frac{\partial \Phi}{\partial r_h} = \frac{-2 \sum_{j>h} (\delta_{hj} - \hat{\delta}_{hj}) \cdot a \sum_{\substack{p \in R \\ p \neq h}} t_{hp}}{\gamma} - \frac{2 \sum_{i<h} (\delta_{ih} - \hat{\delta}_{ih}) \cdot b \sum_{\substack{p \in R \\ p \neq h}} t_{ph}}{\gamma} - \frac{2(\delta_{hh} - \hat{\delta}_{hh}) \cdot [at_{hh} + bt_{hh}]}{\gamma} - \frac{\mu(N-1)}{(N-1)r_h - \sum_{\substack{p \in R \\ p \neq h}} d_{ph} + \epsilon} \tag{18}$$

where:

$$t_{mn} = \begin{cases} 1 & \text{if } r_m - d_{nm} > 0 \\ 0 & \text{else} \end{cases}$$

$$\frac{\partial \Phi}{\partial d_{ij}} = \frac{-2(\delta_{ij} - \hat{\delta}_{ij}) \cdot [1 - at_{ij} - bt_{ji}]}{\gamma} + \frac{2\theta\lambda \sum_k^N [(w_{ik}^j + w_{jk}^i)d_{ij} - w_{ik}^j d_{jk} - w_{jk}^i d_{ik}]}{\theta^2} - \frac{2\lambda \sum_{\substack{i < j < k \\ j \neq i, k}} w_{ik}^j (d_{ij} - d_{jk})^2 \cdot \left[ (d_{jk} - d_{..}) \left( 1 - \frac{2}{N(N-1)} \right) \right]}{\theta^2} + \frac{\mu}{(N-1)r_i - \sum_{\substack{p \in R \\ p \neq i}} d_{pi} + \epsilon} + \frac{\mu}{(N-1)r_j - \sum_{\substack{p \in R \\ p \neq j}} d_{pj} + \epsilon} \tag{19}$$

Figure 2 presents a summary flowchart of the sequential unconstrained minimization technique utilized for this problem.

Various computer analyses with synthetic data were performed to examine the performance of the algorithm. In the first series of analyses, synthetic dissimilarities were constructed from randomly generated **a**, **b**, **r**, and ultrametric  $d_{ij}$  values for each of three major types of models: (1) the ultrametric only model where **a** = **b** = **0**, (2) the symmetric Krumhansl model with **a** = **b**, and (3) the asymmetric Krumhansl model with **a** ≠ **b**. Such analyses were performed for  $N = 8, 10,$  and  $12$  data sets where **a**, **b**, and **r** were limited to single values given degrees of freedom limitations with such small synthetic data sets. In the cases run, perfect recovery of  $\Delta$  was recorded in all but one case. In the second stage of the analysis, error was introduced at two different levels:  $N(0, .1v)$  and  $N(0, .2v)$ , where  $v$  is the standard deviation of the  $\delta_{ij}$  input dissimilarities.

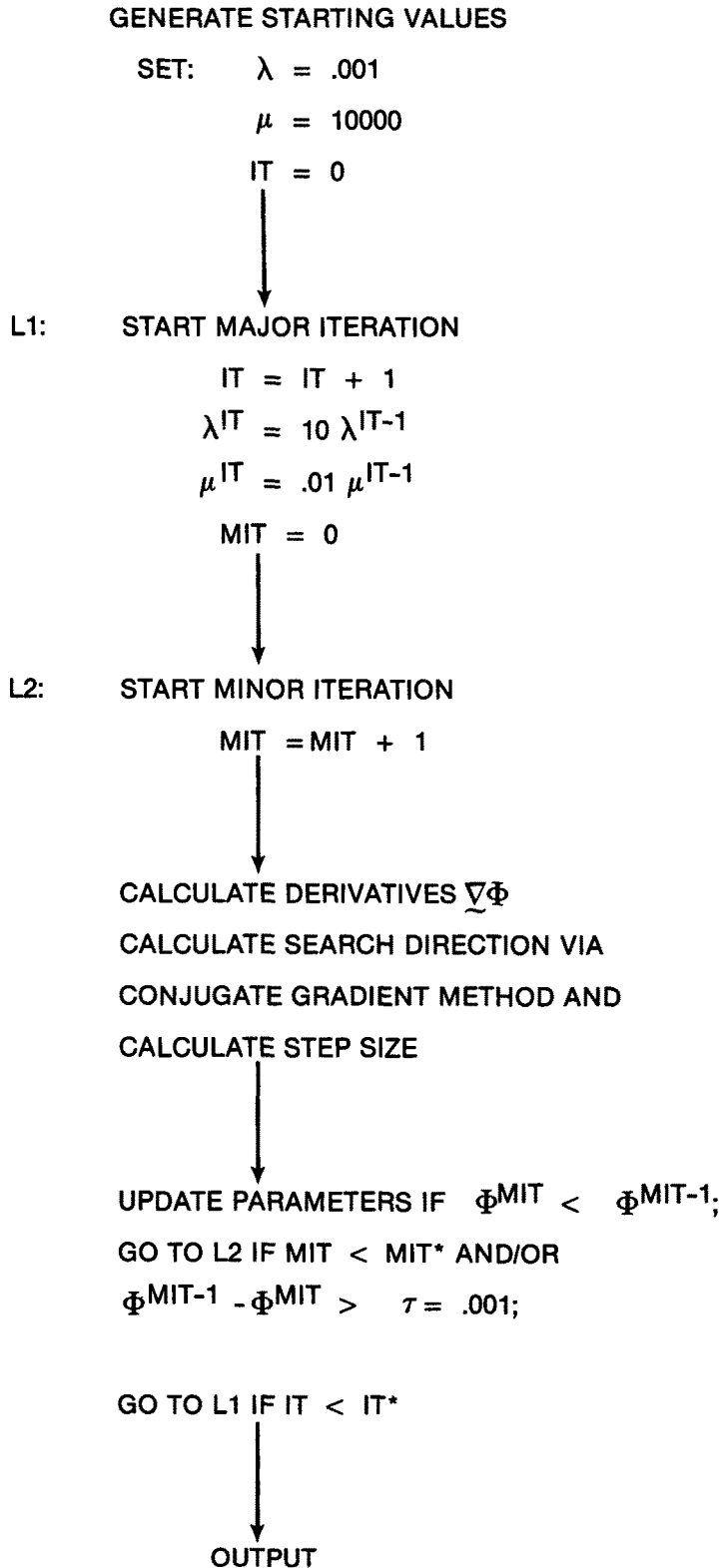


FIGURE 2

Flowchart of the sequential unconstrained minimization algorithm.

Subsequent ANOVA results indicated that these error levels did not significantly detract from recovery parameter.

### *Program Options*

As was previously mentioned, there are a number of options the user can employ to estimate a variety of models and/or reduce the number of estimated parameters. For example, one could set  $a_i = b_j = r_h = 0, \forall i, j, h$ , and solely estimate  $d_{ij}$  for the ultrametric-only model with no density contribution. Or, one could constrain  $a_i = b_j$  and estimate a symmetric version of the Krumhansl model. (We also allow  $a_i$  and  $b_j$  to be unconstrained with respect to sign.) The program allows the user either to estimate  $d_{ij}$  or input them and have them held fixed. This external/internal feature is a useful feature for testing nested models where one might wish to hold the ultrametric part fixed. For the threshold parameters, a variety of options are possible including setting  $r_h = 0, r_h = r$ , have  $r_h$  or  $r$  given and held fixed (e.g.,  $r = \max \delta_{ij} + \varepsilon$ ), or estimate  $r_h$ . Similar options hold for  $a_i$  and  $b_j$ .

A note should be made at this point to contrast the present work with the ultrametric and additive tree fitting methods presented in Carroll (1976) and Carroll and Pruzansky (1980). As mentioned, enforcing the  $a_i = b_j = r_h = 0$  constraints in our methodology is equivalent to merely estimating a "best-fitting" (in a least-squares sense) ultrametric tree as performed by Carroll and Pruzansky. However, the specific density options that are available distinguish these sets of procedures. Given the result that an additive tree can be decomposed into an ultrametric tree and a set of constants, our constrained symmetric model with  $a_i = b_j$  can be viewed as a type of "restricted" additive tree—restricted in the sense that the constants (which typically form a star or bush in the additive tree case) in our Krumhansl methodology are driven by the notions of density as defined in (3) and are not free to vary as in Carroll and Pruzansky's additive tree model (this is a similar argument discussed earlier with respect to the work of Saito, 1986). Carroll and Pruzansky's additive tree conceptualization can be adapted to accommodate asymmetric proximities by estimating two disparate sets of constants—one set for the row stimuli and one set for the column stimuli. The result would be a restricted two-class additive tree (De Soete, DeSarbo, Furnas, & Carroll, 1984) where the  $i$ -th row and  $i$ -th column stimulus were constrained to link to the same internal node of the tree, but where the length of that linkage would vary by row or column stimulus. Our symmetric Krumhansl model could be viewed as a restricted or constrained version of this adaptation where, again, the two sets of constants,  $a_i \neq b_j$ , would reflect the density considerations discussed earlier and not be entirely free to vary.

This distance-density specification is also quite different from the tree-unfolding model proposed by De Soete et al. (1984). In the tree unfolding model, two sets of terminal nodes are estimated—one set for the row stimuli and the other for the column stimuli. Predicted distances are computed, in the ultrametric case, as the height of a common ancestor node for a row ( $i$ ) and a column ( $j$ ) pair of stimuli. While the De Soete et al. approach accommodates asymmetry, there is no explicit modeling of the notion of density. Also, in our operationalization of Krumhansl's (1978, 1982) theory, only one set of terminal nodes is estimated.

## 4. Applications

We present three different applications to illustrate the variety of models that can be fit with this methodology. In particular, we choose to focus upon three types of models: (i) the ultrametric-only model with no density contribution (e.g.,  $\tilde{d}_{ij} = d_{ij}$ ), (ii)

the symmetric model with  $\mathbf{a} = \mathbf{b} = \mathbf{c}$  and  $r_h = r$ , and (iii) the asymmetric model (here  $\mathbf{a} \neq \mathbf{b}$ , but  $a_i = a$ ,  $b_j = b$  and  $r_h = r \forall i, j$ , and  $h$  because of the small data sets and subsequent degrees of freedom). In all three applications, we first estimate the ultrametric-only model. We then perform external analyses holding the initial ultrametric tree fixed for estimation of the symmetric and nonsymmetric Krumhansl models, as well as for other competing discrete models. This sequential estimation (external analysis) allows a direct comparison of different (nested) models. (We also perform internal analyses for the particular solution discussed to examine if the derived tree structure dramatically changes.)

#### *The Rothkopf (1957) Morse Code Data*

Rothkopf (1957) collected a set of confusions among 36 auditory Morse code signals. Each signal consisted of a sequence of dots and dashes. Subjects who were not familiar with Morse code listened to a pair of signals constructed mechanically and separated by a pause of approximately 1.4 seconds. Each subject was required to state whether the two signals presented were the same or different. Table 1 reproduces a portion of this well-known data for only the numerical digits 0–9. Each number in the table is the percentage of roughly 150 respondents who responded “same” to the row signal followed by the column signal. We have also provided the Morse code dash/dot equivalents (signals) in the Table. Note that these numbers were subtracted from 100 to convert them into dissimilarities for the analyses that follow. Following Gower’s (1977) procedure for decomposing an asymmetric matrix into orthogonal symmetric and skew-symmetric parts, for this data set, the symmetric portion accounts for 92.2% of the variance while the skew-symmetric portion accounts for only 7.8%.

The ultrametric-only, symmetric density, and asymmetric density models were fit to these dissimilarities, and Table 2 presents the resulting goodness of fit statistics for the three versions of the model. While we have no formal statistical theory to test nested models and subsequent improvement in fit, for this application it is obvious that the distance-density terms add little explained variance over the ultrametric-only model. The Carroll and Pruzansky (1980) additive tree model produced a variance accounted for statistic (VAF) of .721. The adaptation of Carroll and Pruzansky (1980) to asymmetric proximities with different row and column constants produced a VAF = .752. Each of these was estimated by holding constant the present ultrametric tree and merely estimating the relevant set of constants in order to keep the basis of comparison fixed. As can be seen from the previous decomposition and Table 1, the data do appear to be fairly symmetric with consistently large main diagonal elements that exhibit minor fluctuations. As one reviewer aptly noted, Tversky (1977) found that large asymmetries were found for stimuli differing in the number of components. Here, the subset of digits considered all have the same number of components (five). In addition, Krumhansl (1982) noted that large asymmetries were typically found for pairs where one stimulus is at the boundary (low density) and the other stimulus is in the interior (high density). The digit stimuli considered in this application are all at the boundary of the multidimensional scaling solution (Shepard, 1963) and thus do not appear to differ greatly in terms of density.

Figure 3 presents the ultrametric tree representation estimated for this data set. As can easily be seen, the terminal nodes can be permuted to be nearly perfectly ordered in terms of number of dots/dashes, where, with the exception of 6 and 7, one can, going left to right, encounter increasing (decreasing) numbers of dashes (dots) in the signals. Confusability appears highest with consecutive digit signals, with a few interesting exceptions. For example, 6 is not grouped with 5 and, as Table 1 indicates, is not oftentimes confused with 5 perhaps because the *first* component of the signal for 5 and 6 differ. That is, it may be the case that subjects weigh the first component of the signal

Table 1  
ROTHKOPF (1957) MORSE CODE CONFUSIONS DATA

		Signal presented second									
Digit Stimulus		1	2	3	4	5	6	7	8	9	0
	1	84	63	13	8	10	8	19	32	57	55
	2	62	89	54	20	5	14	20	21	16	11
	3	18	64	86	31	23	41	16	17	8	10
Signal presented first	4	5	26	44	89	42	44	32	10	3	3
	5	14	10	30	69	90	42	24	10	6	5
	6	15	14	26	24	17	86	69	14	5	14
	7	22	29	18	15	12	61	85	70	20	13
	8	42	29	16	16	9	30	60	89	61	26
	9	57	39	9	12	4	11	42	56	91	78
	0	50	26	9	11	5	22	17	52	81	94

\* Percentage of repondents identifying the row signal same as column signal

Stimulus	Morse Code
<u>Digit</u>	<u>Signal</u>
1	.----
2	..---
3	...--
4	....-
5	.....
6	-....
7	--...
8	---..
9	----.
0	-----

more highly in attempting to discriminate between digits. However, the grouping of 1 with 9 and 10 does not support this argument, which may indicate an interesting "interaction" effect. Note that the two major groups formed in Figure 3 separate those digits whose signals start with dots from those starting with dashes, with the sole exception of the 1 digit.

Table 2  
MODEL RESULTS FOR THE MORSE CODE DATA

<u>Model</u>	<u>Error Sums of Squares</u>	<u>V.A.F.</u>
Ultrametric	23489.2	.713
Symmetric	20178.5	.718
Asymmetric	20163.8	.719

Thus the ultrametric tree representation presented in Figure 3 appears to render a parsimonious description of the confusions data in Table 1. The additional explained variance obtained through the application of more complicated models employing different versions of Krumhansl's distance-density notion is small because the data are somewhat symmetric (Kruskal & Wish 1978; Shepard, 1963) with consistently large main diagonal elements.

#### *Soft Drink Brand Switching Data*

Bass, Pessemier, and Lehmann (1972) conducted an experiment with 280 students and secretaries involving the consumption of various brands of soft drinks. The subjects were requested to select a 12-ounce can of soft drink four days a week for three weeks from among the then popular major brands: Coke®, 7-Up®, Tab®, Like®, Pepsi®, Sprite®, Diet Pepsi®, and Fresca®. These brands were selected according to a 2 (cola-noncola) × 2 (diet-nondiet) design as shown in Figure 4.

Note that some brands (e.g., Like®) have altered their position in this design since 1972 via ingredient changes. Table 3 presents the derived brand switching matrix for these eight brands of soft drinks. Here the  $i, j$  element reflects the probability of switching from soft drink brand  $i$  in period  $t$  to soft drink brand  $j$  in period  $t + 1$ . The main diagonal elements reflect differential brand loyalty (i.e., absence of switching). Prior to analysis, the matrix was normalized by dividing each cell by the product of the respective row and column mean—a preprocessing step recommended in DeSarbo (1982) and Rao, Sabavala, and Langfeld (1977) to dampen the effects of differences in market share. In addition, the resulting scale was reversed so as to make the data resemble dissimilarities. Here, the symmetric portion accounts for 68.2% of the variance while the skew-symmetric part accounts for 31.8%.

Table 4 presents the comparative results for the three models. As shown, there is an appreciable drop in the sums of squares in moving from the ultrametric-only tree to the symmetric Krumhansl representation. However, there is little difference between the Krumhansl symmetric and asymmetric models. This small difference is due to the small amounts of asymmetry relative to the high variance across diagonal elements in the proximity matrix. The Carroll and Pruzansky (1980) additive tree procedure yielded a VAF = .751 and the modification for asymmetry produced a VAF = .804. The middle portion of Table 4 presents the estimated  $a = b$  and threshold parameters, as well as the density constants ( $\eta(h)$ ). The lower portion of the table lists the predicted ultrametric

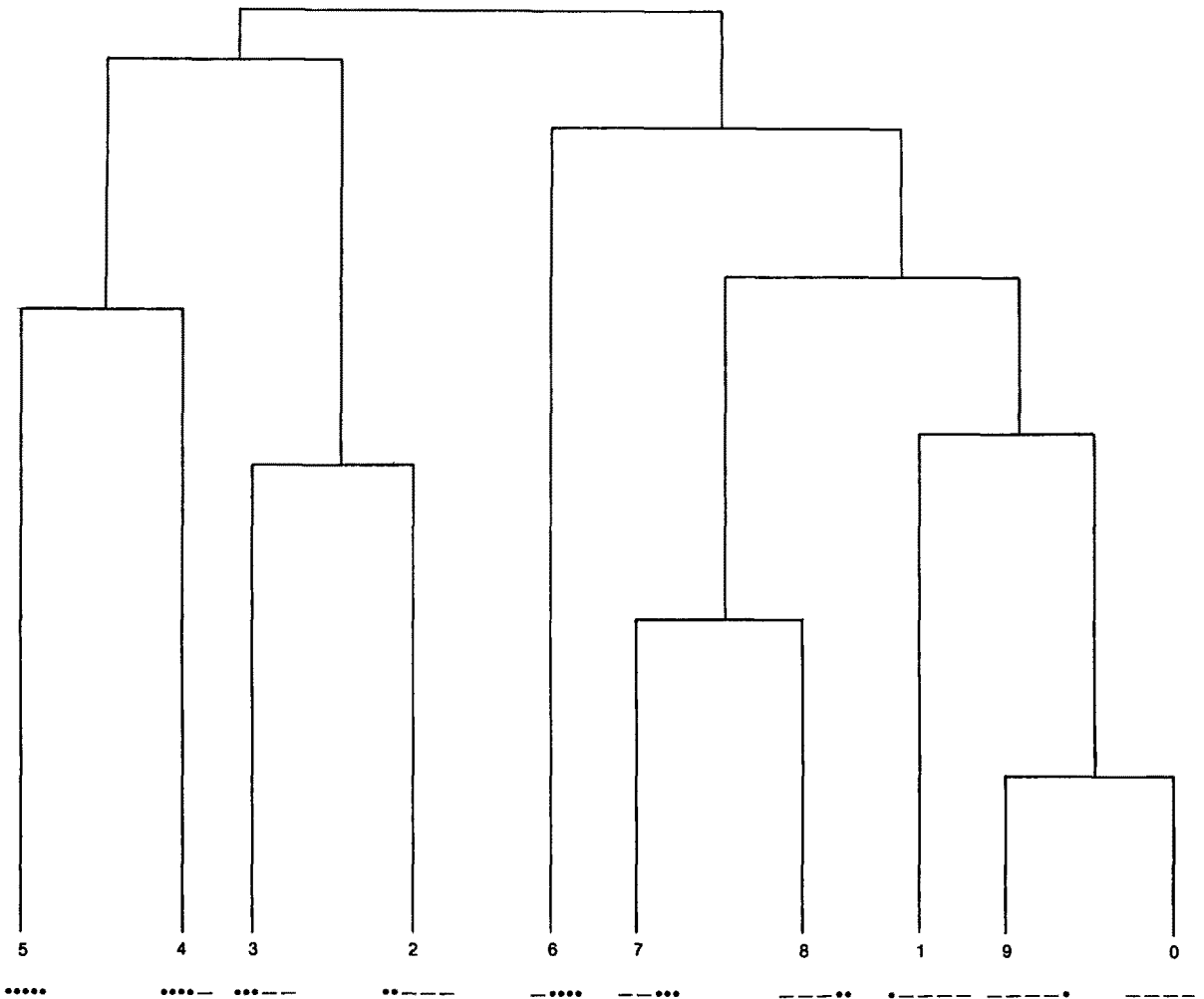


FIGURE 3

Ultrametric tree representation for the Rothkopf (1975) confusion data for the ten digit stimuli.

distances ( $d_{ij}$ ) which can be compared to  $r$  to examine subsequent density contributions (note that  $r > d_{ij}$ ). In particular, note how Tab and Like have a small predicted ultrametric distance producing a high density contribution. Note that an internal analysis for the symmetric Krumhansl model produced congruent results, where the correlation between the two sets of ultrametric distances was 0.989. Figure 5 depicts the estimated ultrametric tree. As can be seen, the diet drinks appear to form a fairly compact group. To the left of this group is the nondiet, lemon-lime group with 7-Up® and Sprite®. And, to its left is the nondiet, cola group with Coke® and Pepsi®. This indicates that switching tends to occur within these three major groups. That is, these consumers are most likely to switch between Coke® and Pepsi®, between 7-Up® and Sprite®, and between the four diet drinks: Fresca®, Like®, Tab®, and Diet Pepsi®. The estimated density constants ( $\eta(h)$ ) presented in Table 4 also reveal some interesting aspects of the brand switching data. The largest values are for the four diet sodas, while the smallest values are for Coke® and Pepsi®. These density constants can, in this context, be interpreted as "penalties" for not achieving a distinctive market position. (They are also strongly negatively correlated with the main diagonals of the data matrix in Table 3.) That is,



	Cola	Non-Cola
Diet	Tab Diet Pepsi	Fresca Like
Non-diet	Coke Pepsi	7-Up Sprite

FIGURE 4  
2 × 2 design for soft drinks.

brands with higher density constants are more likely to have brand share taken away from them by other brands. Note that the two market leaders, Coke® and Pepsi®, suffer the least here due to their dominant market position attained as a result of strong advertising and promotional activity. 7-Up® and Sprite® have intermediate values indicating that their market position is considerably stronger than the diet drinks, but not

Table 3  
BRAND SWITCHING MATRIX (UNNORMALIZED) FOR EIGHT  
BRANDS OF SOFT DRINKS

---



---

		Period t + 1							
		<u>Coke</u>	<u>7-UP</u>	<u>Tab</u>	<u>Like</u>	<u>Pepsi</u>	<u>Sprite</u>	<u>Diet Pepsi</u>	<u>Fresca</u>
Period t	Coke	.612	.107	.010	.033	.134	.055	.013	.036
	7-Up	.186	.448	.005	.064	.140	.099	.012	.046
	Tab	.080	.120	.160	.360	.080	.040	.080	.080
	Like	.087	.152	.087	.152	.239	.043	.131	.109
	Pepsi	.177	.132	.008	.030	.515	.075	.026	.037
	Sprite	.114	.185	.029	.071	.157	.329	.029	.086
	Diet Pepsi	.093	.047	.186	.093	.116	.093	.256	.116
	Fresca	.226	.093	.053	.107	.147	.107	.067	.200

---

Table 4  
MODEL RESULTS FOR THE SOFT DRINK DATA

<u>Model</u>	<u>Error Sum of Squares</u>	<u>V.A.F.</u>
Ultrametric	1186.2	.739
Symmetric	924.8	.764
Asymmetric	874.3	.780

Symmetric Model Results

$a - b = 0.034$   
 $r = 23.604$

Density Constants:	Coke:	1.96
$\eta(h)$	7UP:	17.05
	Tab:	45.40
	Like:	45.43
	Pepsi:	1.99
	Sprite:	17.05
	Diet Pepsi:	33.98
	Fresca:	26.48

Predicted Ultrametric Tree Distances

<u>Soft Drink:</u>	Coke	7-Up	Tab	Like	Pepsi	Sprite	Diet Pepsi	Fresca
Coke	0							
7-UP	23.48	0						
Tab	23.48	20.79	0					
Like	23.48	20.79	1.56	0				
Pepsi	22.37	23.48	23.48	23.48	0			
Sprite	23.48	18.07	20.79	20.79	23.48	0		
Diet Pepsi	23.48	20.79	12.99	12.99	23.48	20.79	0	
Fresca	23.48	20.79	16.74	16.74	23.48	20.79	16.74	0

as dominant as Coke® and Pepsi®. It is interesting to note the close grouping of Tab® and Like® which, at the time of the study, were targeted primarily to a female audience (as were most other diet drinks). These two brands also possess the highest density constants indicating that they (as well as the other two diet drinks to a lesser extent) have not obtained a dominant, unique positioning in the market place. Consumers evidently must have perceived these four brands of diet drinks similarly (i.e., as diet drinks) and switched freely between them. Perhaps consumers were variety seeking within a restricted set of brands that were perceived to have the same dominant characteristic (diet) and/or satisfy the same set of needs.

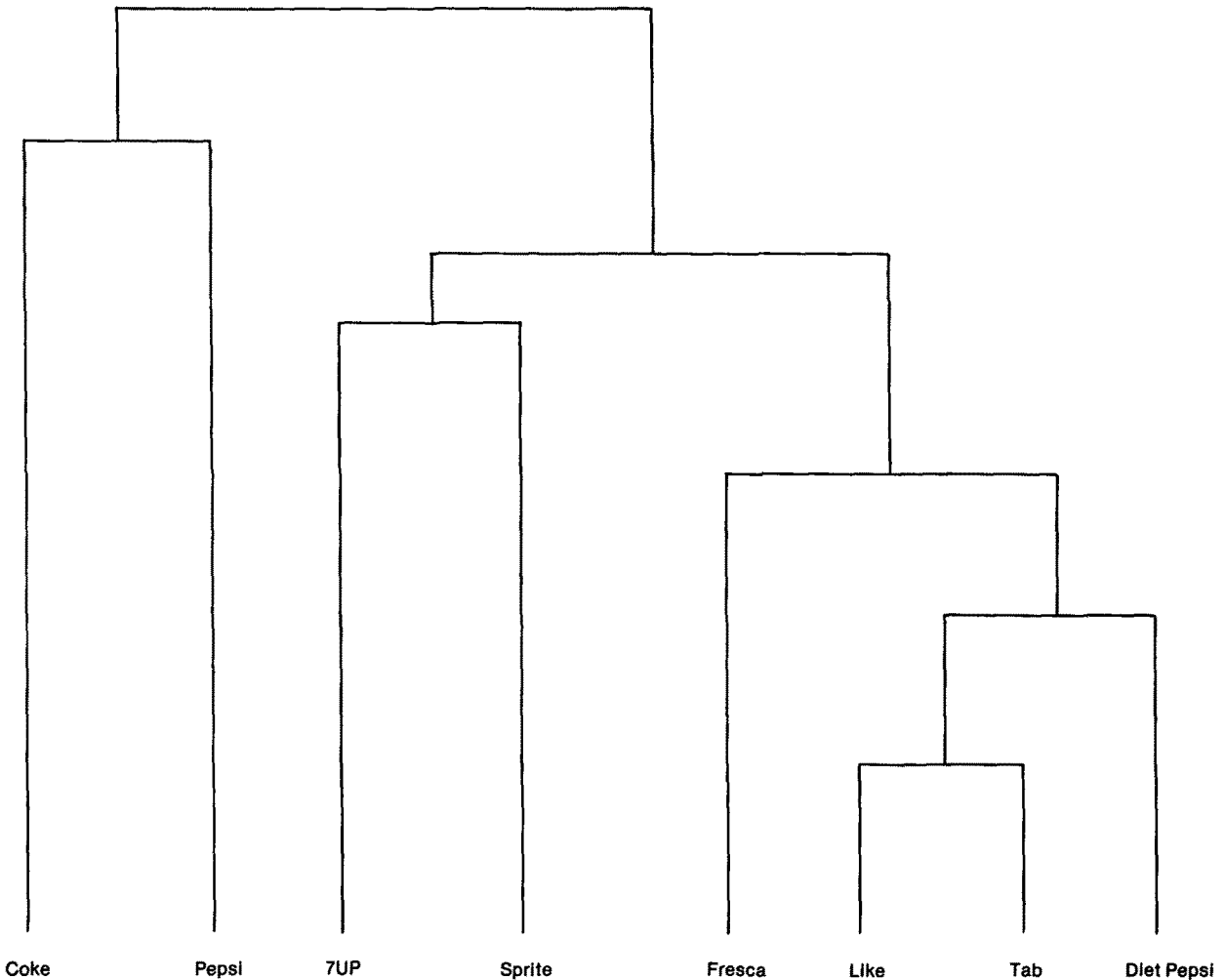


FIGURE 5  
Ultrametric tree representation for the Bass et al. (1972) soft drink brand switching data.

### *Shampoo Word Associations*

Green and Tull (1978, p. 261) present a rectangular array of free-association responses from a marketing study on shampoo benefits. The purpose of the free-association section of the questionnaire was to examine what semantic associations were conjured by various words or phrases that were originally believed to be related to the central benefit of "body". Table 5 presents the frequencies for eight words used as *both* initial stimuli and evoked responses for the  $N = 84$  subjects. (Note that the original data consisted of the  $8 \times 8$  matrix plus another  $8 \times 11$  subsection which included additional evoked response words (e.g., soft, alive, curly, etc.) which were not utilized in this illustration since we are primarily focusing on two-way, one mode asymmetric data.) The main diagonal elements in Table 4 were not collected and are assumed to be equal to the sample size (84). These similarities were subtracted from 84 prior to analysis to convert them into dissimilarities. Here, the symmetric portion accounts for 58.7% of the variance while the skew-symmetric portion accounts for 41.3%.

Table 6 presents the goodness of fit measures for the three models. The Carroll and Pruzansky (1980) additive tree procedure produced a  $VAF = .930$  and the modification

Table 5

## SHAMPOO WORD ASSOCIATION FREQUENCIES

		Evoked Response								Total
		1	2	3	4	5	6	7	8	
Stimulus Word:	1. Body	-	44	5	23	1	19	1	3	96
	2. Fullness	22	-	5	3	1	9	1	2	43
	3. Holds Set	17	21	-	5	0	17	0	5	65
	4. Bouncy	15	12	3	-	1	5	0	14	50
	5. Not limp	28	27	4	18	-	4	1	7	89
	6. Manageable	17	13	11	2	0	-	0	3	46
	7. Zesty	7	9	2	2	0	4	-	13	57
	8. Natural	4	9	1	2	0	7	1	-	24
	Total	110	135	31	75	3	65	4	47	

for asymmetric proximities produced a VAF = .954. The asymmetric model appears to give the most parsimonious explanation of the data's structure based on the much lower sums of squares value. The lower part of the Table presents the predicted ultrametric tree distances ( $d_{ij} < r$ ) which can be compared to  $r = 82.525$  for investigating specific density estimates between pairs of stimuli. Note, in particular, the small  $d_{ij}$  value between body and fullness and the respectively higher contribution to density there. Again, an internal analysis of this asymmetric Krumschl model also produced congruent results, where the correlation between the two sets of ultrametric distances was 0.996. Figure 6 presents the resulting ultrametric tree associated with this asymmetric model. Here, "body" and "fullness" group together at a small height or predicted distance of the tree, as do "holds set" and "manageable", and "bouncy" and "natural". The derived tree indicates that "fullness" is most associated with "body". Next are the control aspects of hair including "holds set" and "manageability". Finally are the liveliness aspects of hair including "zesty", "bouncy", and "natural". Table 6 also presents the  $a$ ,  $b$ ,  $r$  and density constants for these words. Note that the discrepancy between  $a$  and  $b$  values indicates a relatively high degree of asymmetry. In particular,  $b$  is negative while  $a$  is positive indicating opposite row and column contributions to predicted distance. This suggests that higher density increases distance for the stimulus, but reduces distance for the response. Also, note the density constants in the table. There, body and fullness possess the highest  $\eta(h)$  suggesting that the regions around them are quite concentrated with other words. These results suggest that the stimuli

Table 6

MODEL RESULTS FOR THE  
SHAMPOO WORD ASSOCIATIONS DATA

<u>Model</u>	<u>Error</u> <u>Sum of Squares</u>	<u>V.A.F.</u>
Ultrametric - Only	4432.73	0.914
Symmetric	3979.73	0.916
Asymmetric	3132.38	0.935

Asymmetric Model Results

a = .137  
b = -.225  
r = 82.525

Density Constants: ( $\eta(h)$ )	Body	48.39
	Fullness	48.36
	Holds Set	19.10
	Bouncy	26.01
	Not limp	4.05
	Manageable	19.06
	Zesty	20.24
	Natural	26.05

<u>Adjective</u>	Body	Full- ness	Holds Set	Bouncy	Not Limp	Manage- able	Zesty	Natural
Body	0							
Fullness	40.60	0						
Holds Set	80.95	80.95	0					
Bouncy	81.07	81.07	81.07	0				
Not Limp	81.66	81.66	81.66	81.66	0			
Manageable	80.95	80.95	69.87	81.07	81.66	0		
Zesty	81.07	81.07	81.07	74.80	81.66	81.07	0	
Natural	81.07	81.07	81.07	68.99	81.66	81.07	74.75	0

used in this study vary both in terms of their likelihood of being recalled and in their ability to cue the recall of other stimuli. Words that have high spatial densities are closely associated with a large number of other words, presumably as a result of past advertising and promotion. These associations provide many retrieval paths, facilitating recall. For example, the commonly used stimuli—fullness, body, and bouncy—are most often recalled. However, the more distinctive (i.e., low density) words (e.g., not limp, holds set, and zesty) are very difficult to recall, but are the most effective cues for

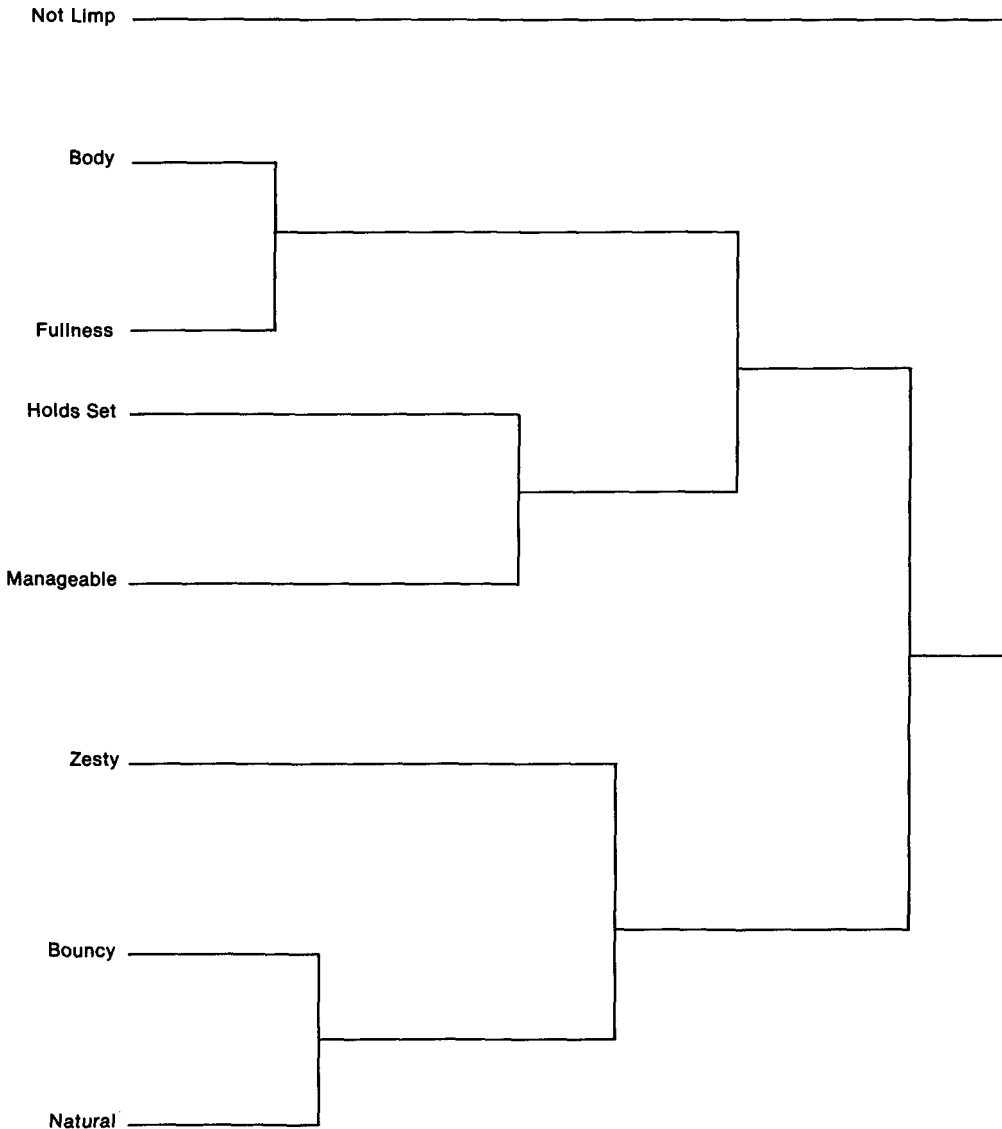


FIGURE 6

Ultrametric tree representation for the Green and Tull (1978) shampoo word association data.

triggering the recall of related stimuli. These cues allow memory access to a number of distant associates.

## 5. Discussion

### *Other Applications*

The distance-density model is most applicable in cases where a directional measure of the association between a set of objects or attributes is available, and this measure is a function of both the distance (dissimilarity) between pairs of stimuli and the distance between these items and other stimuli in the judgment set. The model allows the researcher to estimate the psychological distances, removing this contextual distortion.

Context effects have been empirically demonstrated for a number of measures of

item association. Measures involving memory for objects (e.g., cued recall, item identification, free recall) are especially susceptible. Non-distinctive items (i.e., high spatial density) are often found to be easier to recall than distinctive items since there are a number of closely-associated stimuli which can trigger their remembrance (Gregg 1976). On the other hand, distinctive items (i.e., low spatial density) have been shown to be more effective than non-distinctive items as memory cues (Tulving 1974; Watkins 1979).

Choice judgments are also susceptible to context effects. The likelihood of choosing one alternative over the other has been found to depend on the proximity of these stimuli to the remaining items in the stimulus set (e.g., Debreu 1960; Tversky & Russo 1969). This has implications for the analysis of brand-switching data. For example, consumers may be less likely to switch between two brands if the spatial density of one brand increases. Instead, consumers may split their purchases between this brand and the local (substitutable) context brands (Currim 1982). Increased spatial density might also *increase* switching between two brands in cases where the target brand dominates the local context brands (Huber, Payne, & Puto, 1982). Here, the context serves as a "decoy", increasing the target brand's apparent attractiveness.

Of course, the distance-density model may provide an improved fit over the ultrametric-only model in cases where spatial density is unrelated to the dependent measure, but where, for other reasons, the data matrix has unequal diagonal elements and/or asymmetry. This is more likely to occur when the stimulus densities and distances are estimated simultaneously rather than when independent estimates of stimulus density are used. We therefore recommend that researchers first conduct experimental research to establish an empirical relationship between density and the measure of association. For example, Corter (1987) reports that spatial density has a significant effect on item identification judgments, but not item discrimination or direct ratings of similarity (see also Corter, 1988; Krumhansl, 1988). The distance-density model could then be applied in cases where this relationship is confirmed.

### *Future Research*

We have presented an alternative methodology for the analysis of proximities that may not obey the traditional metric axioms. There are clearly other avenues for methodological research that are available for investigation. One general area for research concerns the specific methodology employed. Because of the noted discontinuities implied by the enforcement of the ultrametric inequality and the specification of the radius-threshold constants, the authors are currently investigating subgradient search routines to replace the conjugate gradient scheme. Other estimation algorithms such as STEPIT (Chandler, 1969) might also be explored for potential computational efficiencies. Also, other specifications of density are to be investigated. One reviewer suggested consideration of  $\Psi_{Ph} = (r_h - d_{Ph})^q$ , with  $q \geq 0$  a parameter to be estimated. The authors are currently in the process of generalizing this methodology to spatial models where a Euclidean metric replaces the ultrametric in (12) and a dimensional space replaces the ultrametric tree. Here, other algebraic expressions for density are accommodated. Another related area for future research concerns the development of corresponding hybrid models which may contain both a tree, a  $T$  dimensional Euclidean MDS space, and the density constants with  $\mathbf{a}$ ,  $\mathbf{b}$ , and  $\mathbf{r}$ . Similarly, fitting multiple trees is a logical extension of this as mentioned in Carroll and Pruzansky (1980). Finally, one could extend this approach to the analysis of three-way proximities as Carroll, Clark and DeSarbo (1984) extend Carroll (1976).

Another direction for future research concerns the empirical investigation of the relationship between stimulus context and similarity judgments. We are not prosely-

tizing for the Krumhansl distance-density hypothesis. We only operationalize it in terms of providing a non-spatial methodology. Experimental studies and analyses such as those performed by Corter (1987, 1988) and Krumhansl (1988) should be conducted to examine the extent to which, or the situations where, Krumhansl's conceptual model is indeed appropriate.

#### References

- Appelman, I. B., & Mayzner, M. S. (1982). Application of geometric models to letter recognition: Distance and density, *Journal of Experimental Psychology: General*, *111*, 60–100.
- Bass, F. M., Pessemier, E. A., & Lehman, D. R. (1972). An experimental study of the relationships between attitudes, brand preference, and choice. *Behavioral Science*, *17*, 532–541.
- Birnbaum, M. H. (1974). Using contextural effects to derive psychological scales. *Perception and Psychophysics*, *15*, 89–96.
- Carroll, J. D. (1976). Spatial, nonspatial, and hybrid models for scaling. *Psychometrika*, *41*, 439–463.
- Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. *Annual Review of Psychology*, *31*, 607–649.
- Carroll, J. D., & Chang, J. J. (1973). A method for fitting a class of hierarchical tree structure models to dissimilarities data, and its application to some body parts data of Miller's. *Proceedings of the 81st Annual Convention of the American Psychological Association*, *8*, 1097–1098.
- Carroll, J. D., Clark, L. A., & DeSarbo, W. S. (1984). The representation of three-way proximity data by single and multiple tree structure models. *Journal of Classification*, *1*, 25–74.
- Carroll, J. D., & Pruzansky, S. (1975, August). *Fitting of hierarchical tree structure (HTS) models, mixtures of HTS models, and hybrid models, via mathematical programming and alternating least squares*. Paper presented at U.S.-Japan Seminar of Multidimensional Scaling, University of California at San Diego, La Jolla, California.
- Carroll, J. D., & Pruzansky, S. (1980). Discrete and hybrid scaling models. In E. D. Lantermann & H. Feger (Eds.), *Similarity and choice* (pp. 48–69). Bern: Hans Huber.
- Chandler, J. P. (1969). STEPIT—Finds local minima of a smooth function of several parameters, *Behavioral Sciences*, *14*, 81–82.
- Corter, J. E. (1987). Similarity, confusability and the density hypothesis. *Journal of Experimental Psychology: General*, *116*, 238–249.
- Corter, J. E. (1988). Testing the density hypothesis: Reply to Krumhansl. *Journal of Experimental Psychology: General*, *117*, 105–106.
- Cunningham, J. P. (1978). Free trees and bidirectional trees as a representation of psychological distance. *Journal of Mathematical Psychology*, *17*, 165–188.
- Currin, I. S. (1982). Predictive testing of consumer choice models not subject to independence of irrelevant alternatives. *Journal of Marketing Research*, *19*, 208–222.
- Debreu, G. E. (1960). Review of R. D. Luce, Individual choice behavior: A theoretical analysis. *American Economic Review*, *50*, 186–188.
- DeSarbo, W. S. (1982). GENCLUS: New models for general nonhierarchical clustering analysis. *Psychometrika*, *47*, 449–475.
- De Soete, G., DeSarbo, W. S., Furnas, G. W., & Carroll, J. D. (1984). The estimation of ultrametric and path length trees from rectangular proximity data. *Psychometrika*, *49*, 289–310.
- Fiacco, A. V., & McCormick, G. P. (1968). Nonlinear programming: Sequential unconstrained minimization techniques. New York: Wiley.
- Gower, J. (1977). The analysis of asymmetric data and orthogonality. In J. P. Barra et al. (Eds.), *Recent developments in statistics* (pp. 109–123). New York: North Holland.
- Green, P. E., & Tull, D. S. (1978). *Research for marketing decisions* (4th ed). Englewood Cliffs, N. J.: Prentice-Hall.
- Gregg, V. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and Recognition* (pp. 183–216). New York: Wiley.
- Hartigan, J. A. (1967). Representation of similarity matrices by trees. *Journal of the American Statistical Association*, *62*, 1140–1158.
- Huber, J., Payne, J., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research*, *9*, 90–98.
- Jardine, C. J., Jardine, N., & Sibson, R. (1967). The structure and construction of taxonomic hierarchies. *Mathematical Biosciences*, *1*, 173–179.
- Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*, 241–254.
- Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, *85*, 445–463.



- Krumhansl, C. L. (1982). Density versus feature weights as predictors of visual identification: Comment on Appelman and Mayzner. *Journal of Experimental Psychology: General*, *111*, 101–108.
- Krumhansl, C. L. (1988). Testing the density hypothesis: Comment on Corter. *Journal of Experimental Psychology: General*, *117*, 101–104.
- Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*, 1–27.
- Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*, 115–129.
- Kruskal, J. B. & Wish, M. (1978). *Multidimensional Scaling*. Beverly Hills, CA: Sage.
- Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. In M. R. Rozenzweig & L. W. Porter (Eds.), *Annual Review of Psychology* (Vol. 32, pp. 89–115). Palo Alto, CA: Annual Reviews.
- Murphy, G. L., & Smith, E. E. (1982). Basic-level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior*, *21*, 1–20.
- Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, *72*, 407–418.
- 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.
- Rao, V. R., Sabavala, D. J., & Langfeld, P. A. (1977). *Alternative measures for partitioning analysis based on brand switching data*. Unpublished manuscript, Cornell University.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.
- Rosenberg, S., & Kim, M. P. (1975). The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, *10*, 489–502.
- Rothkopf, E. Z. (1957). A measure of stimulus similarity and errors in some paired-associated learning tasks. *Journal of Experimental Psychology*, *53*, 94–101.
- 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.
- Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function. Part I. *Psychometrika*, *27*, 125–140.
- Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function. Part II. *Psychometrika*, *27*, 219–246.
- Shepard, R. N. (1963). Analysis of proximities as a technique for the study of information processing in man. *Human Factors*, *5*, 33–48.
- Sjöberg, L. A. (1972). A cognitive theory of similarity. *Goteborg Psychological Reports*, *2*, 10–30.
- Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.
- Tulving, E. (1974). Cue dependent forgetting. *American Scientist*, *62*, 74–82.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.
- Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General*, *113*, 169–193.
- Tversky, A., & Russo, J. E. (1969). Substitutability and similarity in binary choices. *Journal of Mathematical Psychology*, *6*, 1–12.
- Watkins, M. J. (1979). Engrams as cuegrams and forgetting as cue overload: A cuing approach to the structure of memory. In C. R. Puff (Ed.), *Memory organization and structure* (pp. 347–372). New York: Academic Press.

*Manuscript received 7/8/87*

*Final version received 9/12/88*