Journal of Classification 23: 269-299 (2006)
DOI: 10.1007/s00357-006-0016-0

# CLUSCALE ("CLUstering and multidimensional SCAL[E]ing"): A Three-Way Hybrid Model Incorporating Overlapping Clustering and Multidimensional Scaling Structure

Anil Chaturvedi

Capital One, McLean, Virginia

J. Douglas Carroll

Rutgers University, Newark NJ

**Abstract:** Traditional techniques of perceptual mapping hypothesize that stimuli are differentiated in a common perceptual space of quantitative attributes. This paper enhances traditional perceptual mapping techniques such as multidimensional scaling (MDS) which assume only continuously valued dimensions by presenting a model and methodology called CLUSCALE for capturing stimulus differentiation due to perceptions that are qualitative, in addition to quantitative or continuously varying perceptual attributes or dimensions. It provides models and OLS parameter estimation procedures for both a two-way and a three-way version of this general model. Since the two-way version of the model and method has already been discussed by Chaturvedi and Carroll (2000), and a stochastic variant discussed by Navarro and Lee (2003), we shall deal in this paper almost entirely with the three-way version of this model. We recommend the use of the three-way approach over the two-way approach, since the three-way approach both accounts for and takes advantage of the heterogeneity in subjects' perceptions of stimuli to provide maximal information; i.e., it explicitly deals with individual differences among subjects.

**Keywords:** Multidimensional scaling; Perceptual mapping' 3-way' Three-way analysis; Data analysis; CLUSCALE; Multi-linear model; Tri-linear model; Cluster analysis; INDCLUS; INDSCAL; Discrete parameter modeling; Non-linear optimization; Mixed integer programming,; Hybrid model; Metric analysis; MDS; Analysis of covariance data; Analysis of correlation data; Overlapping clusters; Variance decomposition.

---

Author's Addresses: A. Chaturvedi, Accenture, One Freedom Square, 11951 Freedom Drive, Reston, VA 20190-5651, Tel: 703-947-3891, Fax: 703-947-2200, email: anil.chaturvedi@att.net; J.D. Carroll, Rutgers Business School, Rutgers University, Management Education Center, MEC #125, 111 Washington St, Newark NJ 07102, Tel: 973-353-5814, Fax: 973-353-5376, email: dcarroll@rci.rutgers.edu

## 1. Introduction

One area in which MDS and related areas of "perceptual mapping" have been extensively applied is marketing research. Perceptual maps have been a very useful tool for marketers in understanding product differentiation (Dickson and Ginter 1987), product positioning and product preferences. The general practice has been to construct perceptual maps assuming a relatively small set of *common perceptual dimensions* for all the existing as well as potential new products to determine optimal positioning for marketing strategies. Product differentiation is defined as a function of distances (usually Euclidean) between products in the space of a common set of perceptual dimensions.

Certain popular approaches that attempt to explain product differentiation in market research, such as two-way and three-way multidimensional scaling (MDS) (Arabie, Carroll and DeSarbo 1987; Carroll and Arabie 1997; Carroll and Green 1997; Eliashberg and Manrai 1992; Kruskal and Wish 1978; Manrai and Sinha 1989) and factor or components analysis – particularly three-way factor analytic models (Harshman and Lundy 1984) – typically treat products as bundles of common perceptual attributes. These techniques assume that all objects are differentiated via only differences in levels of common physical or perceptual attributes that are "quantitative" in nature, thereby ignoring the effects of additional differentiation that could arise due to "qualitative" – i.e., nominally scaled or categorical – perceptual features such as perceptions unique to objects (e.g., country of origin of the manufacturer; or classification of internet access services providers as DSL providers vs. broadband service providers, etc.) in addition to certain common, quantitative perceptions of reliability, ease-of-access, cost, etc. Though Carroll (1976) first introduced Hybrid models and Carroll and Winsberg (1995), Chaturvedi and Carroll (1998), and Chaturvedi and Carroll (2000) later presented models and methods to capture differentiation arising simultaneously from both quantitative dimensions and qualitative features, their current models could only deal with one class of qualitative features – those that were unique to each brand in the set of products being analyzed. It should be noted that Chaturvedi and Carroll (2000) in their introductory CLUSCALE talk, had provided the earliest empirical illustration of the 2-way CLUSCALE procedure – called 2-way HYCLUS then – (ahead of Navarro and Lee 2003, and standing for "Hybrid CLUStering", a name that was changed because of possible confusion with S.C. Johnson's Hierarchical Clustering Procedure called HICLUS).

In this paper, we develop a descriptive perceptual mapping methodology based on proximity data (e.g., product or other stimulus similarity or dissimilarity matrices, multiple correlation or covariance matrices, or other measures of association) to explain object differentiation due not only to quantitative perceptual dimensions, but also to qualitative perceptual features associated with select objects within a set. Our model also deals with the inherent heterogeneity with respect to the importance people place on these quantitative perceptual dimensions and qualitative perceptual features.

Traditional techniques of two-way and three-way multidimensional scaling (Carroll and Chang 1970; Kruskal 1964a,b; Torgerson 1958) or three-way factor analysis (Harshman and Lundy 1984) that are used in constructing perceptual maps in marketing, psychological, or other applications, do not readily accommodate product or other stimulus differences based on discrete stimulus attributes or features. For example, traditional perceptual maps based on a set of dimensions common to all political candidates may not represent adequately the feature defining how the candidates differ based on their political affiliation (e.g., Democratic vs. Republican), even though they may take a similar stand on many political, economic and other issues. In the marketing context, how does one quantify the amount of differentiation of American cars from cars made by foreign manufacturers, though American cars have similar perceived product attribute levels as competing models? In the final analysis, this can only be attributed to features that are qualitative in nature (e.g., American vs. Foreign), in addition to any differentiation that might arise due to quantitative perceptual dimensions.

This paper suggests that objects can be differentiated not only in a common quantitative perceptual space, but also in terms of a set of discrete or qualitative perceptual features. This group of discrete features could be potentially overlapping, inducing an overlapping cluster structure on the products (see Arabie and Carroll 1980; Carroll and Arabie 1983; Chaturvedi and Carroll 1994). Thus, differentiation occurs not only because of differences in the common perceptual space, but also because of differences in the qualitative perceptual features or (possibly overlapping) classes or clusters. We also posit that the quantitative dimensions and the qualitative, possibly overlapping, features or clusters need not be equally important to all individuals in differentiating or choosing among objects. Well-informed subjects who have knowledge of the objects or other stimulus attributes or experience with the stimuli may use a combination and/or a subset of either the quantitative dimensions or the qualitative features in differentiating the various stimuli. Thus, some subjects might attach more importance to a particular discrete perceptual feature or quantitative perceptual dimension than do others.

The CLUSCALE (simultaneous CLUstering and multidimensional SCAL[E]ing) approach developed in this paper assumes a three-way model based on proximity data on objects aimed at estimating not only the common quantitative object/stimulus and subject spaces defining a common quantitative perceptual structure, but also discrete object clusters or features and a discrete subject space (defining subjects' importance weights for these discrete, qualitative features) which together comprise the unique perceptual structure.

The paper is structured as follows: first, we will briefly introduce the CLUSCALE model, and describe its components – the INDSCAL and INDCLUS models. Then we will present a description of the CLUSCALE model, followed by a description of the estimation procedure – the CLUSCALE method. We then present an empirical application of CLUSCALE to a data set on cars, and finally, we will provide concluding remarks.

## 2. The Components of CLUSCALE: INDSCAL and INDCLUS

The three-way version of CLUSCALE combines the INDSCAL (Carroll and Chang 1970) and INDCLUS (Carroll and Arabie 1983) models for three-way multidimensional scaling and overlapping clustering respectively, based on scalar product matrices derived from multiple proximity matrices. It is actually a special case of a very general class of two-way, three way and higher-way models called, generically, CANDCLUS (for CANonical Decomposition CLUStering; see Carroll and Chaturvedi 1995). CLUSCALE is a model for *scalar-products*-like data. Before we describe CLUSCALE in detail, we will describe briefly the INDSCAL and INDCLUS models.

### 2.1 The INDSCAL Component of the Hybrid Model

INDSCAL determines a common quantitative perceptual space for a set of stimuli, given pair-wise proximity (dissimilarity/similarity) data on these stimuli from multiple data sources. Heterogeneity is incorporated by assuming that each subject (or other source of data) weights each perceptual dimension differently. Since the INDSCAL model and method is well known to MDS users, we will discuss only the most basic aspects of the INDSCAL/SINDSCAL model and method. For more details we refer the reader to Carroll and Chang (1970), Carroll (1972), Pruzansky (1975), and

Arabie, Carroll and DeSarbo (1987). The INDSCAL model for proximity data (similarities or dissimilarities) is of the form:

$$d_{ijk} = F\left(s_{ijk}\right) \cong \left[\sum_{r=1}^{R} w_{kr}\left(x_{ir} - x_{jr}\right)^{2}\right]^{1/2} \qquad (1)$$

where $F$ is generally a linear function since INDSCAL/SINDSCAL is a *metric* model and method (unless $F$ is explicitly stipulated to be some other function based on theoretical or other principles). In most cases, if the $s_{ijk}$'s are similarities $F$ will be a linear function with negative slope; if dissimilarities $F$ will be a positive linear function. It should be noted that the INDSCAL model differs from the classical two-way metric MDS model associated with Torgerson (1952, 1958) only in the introduction of a differential pattern of weights for each of a number ($K$) of subjects or other data sources. If we drop the $k$ subscripts and the weights $w_{kr}$ it is exactly equivalent to that model, so that this classical metric method is, in a sense, a special case of INDSCAL.

The INDSCAL/SINDSCAL methods fit both the "group stimulus space" and "subject/source space directly to estimated scalar products derived from the basic proximity data, based on what is called the "scalar products form" of the INDSCAL model. The first step in INDSCAL/SINDSCAL is to convert the *distance* form of the model given in equation (1) to a "scalar products form" first by estimating the function $F$. This is accomplished by solving the so-called "additive constant problem" after converting similarities into dissimilarities, and, if needed, by reversing the scale. After this, further data preprocessing steps are applied to convert the estimated distances to estimated scalar products (see Carroll and Chang 1970) – specifically, to $K$ matrices (the $k^{\text{th}}$ of which is an $I \times I$ generally symmetric matrix of scalar products between pairs of stimuli $I$ or other objects for subject/source $k$). After this conversion of the data to scalar products form, the scalar products $b_{ijk}$ between stimuli/objects $i$ and $j$ for subject/source $k$, can be shown to be of the form:

$$b_{ijk} \cong \sum_{r=1}^{R} w_{kr} x_{ir} x_{jr} \; . \qquad (2)$$

This scalar products form of the INDSCAL model can be estimated in many ways. In the INDSCAL *method,* this is done by the use of a procedure called CANDECOMP, for *CAN*onical *DECOMP*osition of *N*-way tables, which provides a generalization of singular value decomposition (SVD) to three-way or higher-way data (Carroll and Chang 1970; Carroll and Pruzansky 1984). The SINDSCAL procedure (Pruzansky 1975) makes use of the symmetric nature of the data and implements an efficient CANDECOMP based approach for parameter estimation.

### 2.2 The INDCLUS Component of the CLUSCALE Model

The INDCLUS model (Carroll and Arabie 1983; Arabie, Carroll, and DeSarbo 1987; Chaturvedi and Carroll 1994) also assumes three-way data in the form of multiple proximity matrices, but assumes that objects share a common set of unknown discrete features, which could be potentially overlapping. INDCLUS determines the unknown discrete features, while simultaneously accounting for differential weighting of the discrete features by each data source. INDCLUS is inherently formulated as a scalar products model, but one entailing discrete (0, 1) features rather than continuous dimensions, with continuous weights for subjects or other sources of data. INDCLUS is a three-way generalization of its two-way predecessor called "ADCLUS" (Shepard and Arabie 1979; Arabie and Carroll 1980).

Since ADCLUS/INDCLUS are models for similarity data – whose values are assumed to be non-negative, there is a problem of an additive constant that needs to be solved for when fitting these models to derived scalar products data. For illustration, consider the form of the ADCLUS model:

$$s_{ij} \cong \sum_{t=1}^{T} u_t p_{it} p_{jt} + c \tag{3}$$

where $s_{ij}$ is the similarity between stimuli or other objects *i* and *j*, $p_{it}$ and $p_{jt}$ are discrete (0, 1) variables indicating that *i* and/or *j* are/is respectively either a member of the $t^{\text{th}}$ cluster (indicated by $p = 1$) or not a member ($p = 0$), and where "$\cong$" means "equals, except for otherwise unspecified error terms", or, simply, "approximately equals". The "weights" $u_t$ are assumed to be non-negative, which means the entire summational expression on the right (a weighted sum of features possessed by *both i and j*) must itself necessarily be nonnegative. Since the $I \times I$ matrix $\mathbf{S} = (s_{ij})$ is a derived scalar products matrix and would contain both positive and negative values, a judicious choice of the additive constant *c* added to all elements would enable the predicted scalar products model to take on all real values – not just non-negative values. Note that c can be viewed either as such an additive constant converting interval scale scalar products estimates into (nonnegative) ratio scale estimates, or as the weight for a $(T+1)^{\text{st}}$ cluster comprising the *universal* set, in which *every* object is contained.

The three-way generalization of ADCLUS/MAPCLUS, called "INDCLUS" (Carroll and Arabie 1983), standing for INdividual Differences CLUStering, simply allows a distinct profile of cluster/feature weights for each subject/source for the $T$ clusters, as well as a different additive constant $(c_k)$ for each of the K subjects/sources. Thus, the INDCLUS model can be stated as:

$$s_{ijk} \cong \sum_{t=1}^{T} u_{kt} p_{it} p_{jt} + c_k \,. \tag{4}$$

## 3. CLUSCALE - A Hybrid Model Combining INDSCAL and INDCLUS

The CLUSCALE model is a hybrid model that combines the INDSCAL and INDCLUS models in an additive fashion. Thus, the form of the CLUSCALE model, which we write below assuming scalar products data is:

$$b_{ijk} \cong \sum_{r=1}^{R} w_{kr} x_{ir} x_{jr} + \sum_{t=1}^{T} u_{kt} p_{it} p_{jt} + c_k \tag{5}$$

where:

- R is the number of quantitative perceptual dimensions
- T is the number of qualitative or discrete features
- $b_{ijk}$ is the derived scalar product of stimuli $i$ and $j$ for subject $k$,
- $w_{kr}$ is the non-negative importance weight of the $k^{th}$ subject on the $r^{th}$ quantitative perceptual dimension,
- $x_{ir}$ is the coordinate of the $i^{th}$ stimulus on the $r^{th}$ quantitative perceptual dimension,
- $u_{kt}$ is a non-negative importance weight of the $k^{th}$ subject for the $t^{th}$ qualitative or discrete feature,
- $p_{it}$ is the binary variable representing the presence (value of 1) or absence (value of 0) of the $t^{th}$ qualitative or discrete feature for the $i^{th}$ stimulus.
- $c_k$ is the additive constant for subject $k$.

The CLUSCALE model can be re-stated in matrix form as:

$$\mathbf{B}_k \cong \mathbf{X}\mathbf{W}_k\mathbf{X}' + \mathbf{P}\mathbf{U}_k\mathbf{P}' + \mathbf{C}_k \tag{6}$$

where:

- $I$ is the number of stimuli
- $k$ denotes the $k^{th}$ subject, or other data source (e.g., $k^{th}$ market segment), for $k = 1, 2 \cdots K$, where $K$ is the total number of subjects or other data sources.

- $R$ is the number of quantitative perceptual dimensions
- $T$ is the number of qualitative/discrete features
- $\mathbf{B}_k$ is an $I \times I$ scalar product-like similarity matrix for subject/source $k$
- **X** is an $I \times R$ "stimuli × quantitative dimensions" matrix defining coordinates of the $R$ dimensions of the continuous perceptual space
- $\mathbf{W}_k$ is an $R \times R$ diagonal matrix of nonnegative importance weights for the $R$ quantitative dimensions for the $k^{th}$ subject or other source of data
- **P** is an $I \times T$ binary matrix of qualitative discrete features (or cluster membership variables) for $T$ (possibly overlapping) clusters
- $\mathbf{U}_k$ is an $T \times T$ diagonal matrix of nonnegative weights for $T$ clusters for the $k^{th}$ subject or other data source
- $\mathbf{C}_k$ is a matrix whose off-diagonal entries all consist of the additive constant $c_k$.

### 3.1 Model Properties

The CLUSCALE model is, in spirit, a model belonging to a general family of hybrid models envisioned first by Carroll (1976), that involved modeling of continuous and discrete structure in data. A lot of rich work has been done since Carroll's presidential address to the Psychometric Society in 1976. More specifically, the CLUSCALE model in Equations (5) and (6) is a distance model in both the INDSCAL and INDCLUS components. While not obvious, the predicted similarities $s_{ijk}$ (Equation 3) between objects $i$ and $j$ for individual $k$ corresponding to the INDCLUS component can be converted to distances by the linear transformation $d_{ijk} = C - s_{ijk}$ where $C$ is a constant, determined by the expression $C = \max (s_{ijk}\text{-}s_{ilk}\text{-}s_{ljk})$ $\forall$ i, j, l $\in$ 1, … , $I$; k = 1, …, $K$.

In terms of the CLUSCALE model identification, the sub-model corresponding to its INDSCAL sub-component solves for the rotational indeterminacy inherent in two-way MDS models. The central advantage of the model in Equations (5) and (6) above is that the quantitative perceptual dimensions corresponding to INDSCAL are not subject to the orthogonal rotational indeterminacy inherent in most classical perceptual mapping procedures, but have a fixed orientation. The quantitative dimensions of CLUSCALE do suffer from the same scale indeterminacy as INDSCAL in the object and subject spaces and are subject to scale indeterminacies among the quantitative dimension coordinates and the importances that the subjects attach to these dimensions. These are resolved – up to possible reflection of coordinate axes – by constraining the sum of squares of the object coordinates on each quantitative perceptual dimension to equal one. CLUSCALE also entails an indeterminacy enabling permutations of the order of both the quantitative dimensions and the qualitative features, generally resolved by ordering them based on relative variance accounted for (VAF). There are no other theoretical or empirical indeterminacies that we have observed in the CLUSCALE model.

It should be noted that fitting this hybrid model combining INDSCAL and INDCLUS is *not* equivalent to fitting the INDSCAL and INDCLUS models *separately* to these data and then simply defining an additive combination of the two resulting models. In general, the combined hybrid model can be expected to require fewer total components (the $R$ dimensions plus the $T$ discrete features) than the sum of those for these separate solutions. This will occur because the continuous dimensions and discrete features in those two separate solutions will generally tend to be somewhat correlated. Often this correlation is of a quite simple nature – a quantitative dimension in one solution may correlate strongly with a qualitative feature in the other – but equally often it may be much more complex, involving correlations between complex linear (or even nonlinear) combinations of dimensions and features in the two.

One may ask the question: Why are discrete features necessary at all? Why could we not just replace all discrete features with continuous dimensions, even if they might be a bit more "noisy"? The best answer to this question is as follows: if there *is* indeed a truly discrete feature or property (e.g., male-female, domestic-foreign, electronic-mechanical, power-manual, animal, vegetable or mineral, etc.), its most effective and parsimonious representation can only be through a discrete feature or property than by a continuous dimensional one. We believe that the number of degrees of freedom for a discrete feature is *less than* that for a continuous spatial dimension. Another way to look at this is that it is more appropriate and mathematically more efficient, effective and meaningful to represent a truly discrete feature, property or cluster structure discretely rather than continuously. We believe that, in time, theoretically justifiable measures will be devised to reliably establish when and under what conditions a discrete representation is, in fact, more parsimonious and meaningful then a more continuous/spatial dimensional one. For the moment, though, we must leave this as a yet unsolved theoretical/statistical problem.

## 4. The CLUSCALE Method: Estimation Procedure

A three-way methodology based on the CANDCLUS *method* of Carroll and Chaturvedi (1995) is used to fit the CLUSCALE model. It should be noted that the CANDCLUS *method* combines the CANDECOMP method of Carroll and Chang (1970), with a methodological principle based on a row-wise separability property (Chaturvedi and Carroll 1994) of Lp-norm based loss functions for the CANDECOMP family of models, and develops a generalized procedure to estimate not just quantitative dimensions and weights, but also discrete factors and weights for any of the *N* ways based on any Lp-norm based loss function.

Specifically, in equation (6) above, matrices **X**, **W**, **P**, **U**, and **C** are all unknown. Only the data matrices $\mathbf{B}_k$ are known. Determining the OLS parameter combines continuous optimization with a discrete, non-linear, integer programming problem. We alternate between estimating the *R* quantitative perceptual dimensions in the *R*-Step and the *T* qualitative discrete features (possibly overlapping) in the *T*-step, until convergence occurs to at least a locally optimal solution. By starting from a number of "rational" and/or random starting configurations, as is routine in MDS and related work, we aim to find the best solution for a particular combination of values of *R* and *T,* and generally do. If the same solution is found several times from different starts, we can assume without serious doubt that the solution is the best possible (i.e., the globally optimal) solution.

### 4.1 R-Step: Estimating the Parameters of the R Quantitative Perceptual Dimensions

The parameters associated with the *R* quantitative dimensions (first part of equations 5 and 6 involving matrices **X** and **W**) are estimated by iterating between estimating the object/stimulus space (matrix **X**) and subject space ($\mathbf{W}_k$'s) in an alternating least squares fashion, using the following equation:

$$\overline{\mathbf{B}}_k = \mathbf{B}_k - \hat{P}\hat{U}_k\hat{\mathbf{P}}' - \hat{\mathbf{C}}_{\mathbf{k}} = \hat{\mathbf{X}}\hat{\mathbf{W}}_k\hat{\mathbf{X}}' + \text{error} \tag{7}$$

As stated earlier in this paper, this methodology is too well documented to be described here again in detail.

### 4.2 T-Step: Estimating the Parameters of the T Qualitative or Discrete Features

The parameters associated with the *T* discrete (possibly overlapping) features are estimated using the "one-cluster-at-a-time" SINDCLUS methodology of Chaturvedi and Carroll (1994), which we will describe in greater detail here, since it is less well known than the CANDECOMP method. Assuming that the parameters for the quantitative, perceptual space ($\mathbf{X}$, $\mathbf{W}_k$) are known, at least conditionally, we reformulate (6) as:

$$\overline{\mathbf{B}}_k = \mathbf{B}_k - \hat{\mathbf{X}}\hat{\mathbf{W}}_k\hat{\mathbf{X}}' = \mathbf{P}\mathbf{U}_k\mathbf{Q}' + \mathbf{C}_k + \text{error} \tag{8}$$

Note that this is a relaxation of (6), since the CLUSCALE model assumes $\mathbf{P}$=$\mathbf{Q}$. We do not impose this constraint in the estimation procedure. We typically find that for symmetric data, the iterative procedure converges to $\mathbf{P}$=$\mathbf{Q}$ at the locally optimal solutions. Ten Berge and Kiers (2005) provide details on an estimation procedure that does impose the constraint of $\mathbf{P}$=$\mathbf{Q}$. The estimation problem is to determine the ordinary least squares estimates of binary $\mathbf{P}$ and $\mathbf{Q}$, diagonal but continuous $\mathbf{U}_k$, and the matrix $\mathbf{C}_k$ whose off-diagonal elements are $c_k$. This is a 0-1 non-linear integer programming problem. Assuming random starts for $\mathbf{Q}$, $\mathbf{U}_k$, and $\mathbf{C}_k$ , we use the following four-steps in an iterative fashion until the algorithm converges to at least a locally optimal solution.

Step *T*1. Estimate $\mathbf{P}$, given current $\mathbf{U}_k$, current $\mathbf{Q}$, and current $\mathbf{C}_k$
Step *T*2. Estimate $\mathbf{Q}$, given $\mathbf{P}$ from step *T*1, current $\mathbf{U}_K$, and current $\mathbf{C}_k$
Step *T*3. Estimate $\mathbf{U}_k$, given $\mathbf{P}$ from step *T*1, $\mathbf{Q}$ from step *T*2, and current $\mathbf{C}_k$
Step *T*4. Estimate $\mathbf{C}_k$, given $\mathbf{P}$ from step *T*1, $\mathbf{U}_k$, from step *T*2, and $\mathbf{Q}$ from step *T*3

We repeat steps *T*1 through *T*4 until convergence to at least a locally optimal solution occurs. Upon convergence, estimated $\mathbf{P}$ and $\mathbf{Q}$ matrices are usually equal for symmetric data.

*Step T1: Estimate* $\mathbf{P}$*, given current* $\mathbf{U}_k$*, current* $\mathbf{Q}$*, and current* $\mathbf{C}_k$
Let

- $\mathbf{G}$ be an $I$ x $KI$ matrix $[\overline{\mathbf{B}}_1 \mid \overline{\mathbf{B}}_2 \mid \overline{\mathbf{B}}_3 \mid \ldots \mid \overline{\mathbf{B}}_k]$
- $\mathbf{u}_t$ be a vector of weights for the $t^{\text{th}}$ cluster (or feature)
- $\mathbf{p}_t$ be an $I \times 1$ binary vector representing the $t^{\text{th}}$ column of Matrix $\mathbf{P}$
- $\mathbf{q}_t$ be an $I \times 1$ binary vector representing the $t^{\text{th}}$ column of Matrix $\mathbf{Q}$
- $\mathbf{D}$ be a $(T+1) \times KI$ matrix (including the universal cluster), where the $t^{\text{th}}$ row of $\mathbf{D}$ is $\mathbf{d}_t$ where $\mathbf{d}_t = \mathbf{u}_t \otimes \mathbf{q}_t$ (where $\otimes$ denotes the Kronecker product)
- $\mathbf{P}$ be an $I \times (T+1)$ binary matrix of stimuli x clusters
- $\mathbf{P}_{-t}$ is matrix $\mathbf{P}$ with the $t^{\text{th}}$ column dropped
- $\mathbf{D}_{-t}$ is matrix $\mathbf{D}$ with the $t^{\text{th}}$ row dropped

Assuming parameters for all clusters except the $t^{th}$ cluster are known, and assuming $\mathbf{d}_t$ is known the estimation problem becomes estimating $\mathbf{p}_t$ in

$$\mathbf{G} - \mathbf{P}_{-t}\mathbf{D}_{-t} = \mathbf{p}_t\mathbf{d}'_t + Error$$

or equivalently, estimating $\mathbf{p}_t$ in

$$\mathbf{G}_t^* = \mathbf{p}_t\mathbf{d}'_t + Error$$

where $\mathbf{G}_t^* \equiv \mathbf{G} - \mathbf{P}_{-t}\mathbf{D}_{-t}$.

This is done by using the Elementary Binary Least Squares procedures (EBLSP) described in Chaturvedi and Carroll (1994). This procedure is then repeated for $t = 1, \ldots, T$.

*Step $T_2$. Estimate* **Q**, *given* **P** *from step $T_1$, current* $\mathbf{U}_k$, *and current* $\mathbf{C}_k$

Let

- **H** be an $I \times KI$ matrix $[\overline{\mathbf{B}}_1 \mid \overline{\mathbf{B}}_2 \mid \overline{\mathbf{B}}_3 \mid \ldots \mid \overline{\mathbf{B}}_k]$
- $\mathbf{u}_t$ be a $K \times 1$ vector of weights for the $t^{th}$ cluster
- $\mathbf{p}_t$ be an $I \times 1$ binary vector representing the $t^{th}$ column of Matrix **P**
- $\mathbf{q}_t$ be an $I \times 1$ binary vector representing the $t^{th}$ column of Matrix **Q**
- **E** be a $(T+1) \times KI$ matrix (including the universal cluster), where the $t^{th}$ row of **E** is $\mathbf{e}_t$ where $\mathbf{e}_t = \mathbf{u}_t \otimes \mathbf{p}_t$.
- **Q** be an $I \times (T+1)$ binary matrix of stimuli × clusters
- $\mathbf{Q}_{-t}$ is matrix **Q** with the $t^{th}$ column dropped
- $\mathbf{E}_{-t}$ is matrix **E** with the $t^{th}$ row dropped

Assuming parameters for all clusters except the $t^{th}$ cluster are known, and assuming $\mathbf{e}_t$ is known the estimation problem becomes estimating $\mathbf{q}_t$ in

$$\mathbf{H} - \mathbf{Q}_{-t}\mathbf{E}_{-t} = \mathbf{q}_t\mathbf{e}'_t + \text{ Error}$$

or equivalently, estimating $\mathbf{q}_t$ in

$$\mathbf{H}_t^* = \mathbf{q}_t\mathbf{e}'_t + \text{ Error}$$

where $\mathbf{H}_t^* \equiv \mathbf{H} - \mathbf{Q}_{-t}\mathbf{E}_{-t}$.

This is also done by using the EBLSP procedure. This procedure is then repeated for $t = 1, 2, \cdots T$.

*Step T3. Estimate* $\mathbf{U}_k$, *given* **P** *from step* $T_1$, **Q** *from step T2, and current* $\mathbf{C}_k$

Let

- **J** be a $K \times I^2$ matrix that contains the $I^2$ elements of Matrices $\overline{\mathbf{B}}_\mathbf{k}$ in the $k^{th}$ row
- $\mathbf{u}_t$ be a $K \times 1$ vector of weights for the $t^{th}$ cluster

- $\mathbf{p}_t$ be an $I \times 1$ binary vector representing the $t^{\text{th}}$ column of Matrix $\mathbf{P}$
- $\mathbf{q}_t$ be an $I \times 1$ binary vector representing the $t^{\text{th}}$ column of Matrix $\mathbf{Q}$
- $\mathbf{F}$ be a $(T+1) \times I^2$ matrix (including the universal cluster), where the $t^{\text{th}}$ row of $\mathbf{F}$ is row vector $\mathbf{f}'_t$ where $\mathbf{f}_t = \mathbf{p}_t \otimes \mathbf{q}_t$ .
- $\mathbf{U}$ be a $K \times (T+1)$ binary matrix of stimuli × clusters
- $\mathbf{f}_t = \mathbf{p}_t \otimes \mathbf{q}_t$
- $\mathbf{U}_{-t}$ is matrix $\mathbf{U}$ with the $t^{\text{th}}$ column dropped
- $\mathbf{F}_{-t}$ is matrix $\mathbf{F}$ with the $t^{\text{th}}$ row dropped

Assuming parameters for all clusters except the $t^{\text{th}}$ cluster are known, and assuming $\mathbf{f}_t$ is known the estimation problem becomes estimating $\mathbf{u}_t$ in

$$\mathbf{J} - \mathbf{U}_{-t}\mathbf{F}_{-t} = \mathbf{u}_t\mathbf{f}_t' + Error$$

or equivalently, estimating $\mathbf{u}_t$ in

$$\mathbf{J}_t^* = \mathbf{u}_t\mathbf{f}_t' + \text{Error}$$

where $\mathbf{J}_t^* \equiv \mathbf{J} - \mathbf{U}_{-t}\mathbf{F}_{-t}$ .

We use a closed form expression to find the OLS estimates of these continuous parameters. It should be noted, too, that non-negativity constraints can easily be imposed on the elements of the vector $\mathbf{u}_t$ by simply "zeroing out" any values whose estimates have negative values in this "one component at a time" approach. This procedure is then repeated for $t = 1, \ldots, T$.

*Step T4. Estimate* $\mathbf{C}_k$*, given* $\mathbf{P}$ *from step* $T_1$*,* $\mathbf{U}_k$*, from step* $T_2$*, and* $\mathbf{Q}$ *from step* $T_3$

This step is analogous to step T3, with both $\mathbf{p}_t$ and $\mathbf{q}_t$ known. A closed form solution yields the OLS solution in this case also.

It should be noted that for data based on a single proximity matrix ($K$=1), the CLUSCALE algorithm described above gets modified slightly. The $R$-step of the CLUSCALE algorithm uses the singular value decomposition (SVD) for estimating the parameters associated with the quantitative perceptual dimensions rather than the CANDECOMP methodology presented in Carroll and Chang (1970) for reasons detailed in that 1970 paper. The $T$-step of the CLUSCALE algorithm remains the same as when there are multiple data sources ($K$ >1 subjects or other data sources). This slightly modified algorithm is called the two-way CLUSCALE method throughout the remainder of this paper. It is important to note, however, that while, in the case of three-way data, because of the well established "dimensional uniqueness" property of INDSCAL, the dimensions in the spatial component of CLUSCALE are uniquely identified, in the two-way case the spatial component is subject to the same rotational indeterminacy that is characteristic of other two-way MDS methods based on the Euclidean metric.

## 5. Partitioning Variance

One issue that arises is of apportioning the total VAF in the CLUSCALE model into a contribution of the continuous, spatial structure and that of the discrete, cluster-like structure. This is a difficult question, closely related to that of attributing variance to correlated independent variables in a multiple regression context. In fact, we would argue that the latter is a general question of which the former is a special case.

Let us just consider the regression case. As mentioned above, the problem of uniquely associating VAF to individual independent variables is one that has no unique solution, although many possible solutions have been discussed in the statistical literature. A general discussion of this problem is provided by Green, Carroll and DeSarbo (1978), while a particularly elegant symmetrical solution (independent of order in which the variables are entered into the regression equation, and having the property that the individual variable VAF’s sum to the total VAF) is offered, which results in a measure of predictor variable importance called “delta squared”.

As has been discussed by Carroll and Chang (1970), Carroll and Chaturvedi (1995) and earlier in the current paper, once the continuous dimensions and/or the discrete feature vectors have been discussed, (whether one is dealing with the INDSCAL, INDCLUS or hybrid CLUSCALE model) the subject weights are estimated by a special case of multiple linear regression. The independent variables in this formal regression problem are what are sometimes called “columnwise Kronecker products” of the dimension and/or feature vectors (the former only in the case of INDSCAL, the latter for INDCLUS, and the two combined for CLUSCALE).

Specifically, using the delta squared measure for each dimension and/or feature, one has a consistent means of attributing variance for each individual subject uniquely to a) dimensions in an INDSCAL solution, b) features or clusters in a CLUSCALE solution or c) dimensions *and* features in a CLUSCALE solution. Note that the measures are defined separately for each subject, since it is that subject’s weights for dimensions and/or clusters that are calculated via regression. Thus, if an overall partition of VAF for dimensions and/or features, the delta squared for each for each subject must be summed over subjects, separately for each dimension and/or feature, to derive a partition of total VAF in the overall model being fit.

Given the additive property of the delta squared measure, the CLUSCALE solution VAF can be partitioned into separate additive VAFs for the continuous dimensions and the discrete features. This can be accomplished by summing the total delta squared measure of VAF for each dimension over the R dimensions to get a total VAF for the continuous structure in the model fit while, analogously, summing these measures over the T features to define a VAF for the discrete structure. The additive property of delta squared guarantees that these two separate VAF components will add to the total VAF for the overall CLUSCALE model that has been fit! This provides a very general and unique basis for partitioning variance between the quantitative dimensions and the qualitative features. This will be discussed later in the paper.

## 6. Application to Car Data

Chaturvedi (1993) provides pair-wise dissimilarity data collected from 24 MBA students at a university on the east coast of USA. Sixteen cars were chosen for the study. The sixteen cars used in the study were: Alpha Romeo 164L, BMW 325I, Buick Riviera, Ferrari 348TB, Ford Mustang, Honda Accord, Honda Prelude, Hyundai Sonata, Lincoln Town Car, Mercedes Benz 190E, Plymouth Sundance, Pontiac Firebird, Rolls Royce, Toyota Celica, Toyota Lexus SC 400, and Volkswagen Golf. The order of presentation of the 120 pairs of cars was randomized across the 24 subjects/consumers in order to eliminate order effects.

### 6.1 Application of Two-Way CLUSCALE to the "Average" Proximity Matrix from 24 Consumers

Two-way CLUSCALE was first applied to an average proximity matrix derived from the 24 proximity matrices corresponding to the 24 consumers. After a large number of CLUSCALE runs (ranging from 1through 5 quantitative dimensions and 1 through 5 discrete perceptual features), we chose a solution with $R$=2 quantitative perceptual dimensions and $T$=2 qualitative perceptual features, accounting for 82% of the variance in the averaged data, based on the criteria of (a) a "scree-test"; and (b) interpretability of the solution. Since, the focus of this paper is entirely on the three-way CLUSCALE model and methods, we do not describe the two-way CLUSCALE solution in detail. We will only remark that, as would be expected in the two-way case, an orthogonal rotation of the quantitative two-way MDS component of the two-way CANDCLUS model provided a solution essentially *identical* to the dimensions in the "group stimulus space" of the three-way CLUSCALE solution, to be discussed below. The qualitative discrete feature representation in the two-way analysis did not correspond very well, however, with that in the three-way CLUSCALE solution – primarily because, we believe, of the greater number of such features extracted in the three-way analysis ($T$=3). We believe the reason for this difference in number of features in the two- and three-way CLUSCALE analyses was probably the fact that the greater richness of the data analyzed in the three-way case (and of the model being fitted) enabled reliable fitting of more such discrete features/properties. Those fitted in the two-way case (only two, $T$=2 in this case) may have been some form of composite of the three such features fitted for the three-way model.

### 6.2 Application of Three-Way CLUSCALE to the 24 Proximity Matrices

Three-way CLUSCALE was then applied to the 24 proximity matrices corresponding to the 24 consumers. As with the application of two-way CLUSCALE, we obtained a variety of CLUSCALE solutions ranging from 1 through 5 quantitative perceptual dimensions, and 1 through 5 qualitative perceptual features. Table 1 presents the VAF (Variance-Accounted-For) for these solutions. We found a solution with $R$=2 quantitative perceptual dimensions and $T$=3 qualitative perceptual features, accounting for 72.1% of the variance in the data, to be most interpretable. The other solutions had only parts of the solutions interpretable (either some or all of the quantitative dimensions or qualitative features).

Figure 1 presents the two quantitative perceptual dimensions extracted using three-way CLUSCALE. Since the orientation of the brands with respect to the axes is fixed (no rotational indeterminacy in the solution), a good interpretation of the axes would provide "face validity" to the quantitative dimensions of the CLUSCALE solution. In this case, the two dimensions were interpreted as the "Price" dimension, and a bi-polar "Luxuriousness-Sportiness" dimension. The price dimension clearly rank orders the "Super-luxury" cars such as Rolls-Royce and Ferrari from the "Luxury-cars" such as Mercedes Benz, BMW, Alpha Romeo, Lincoln Town Car, and Toyota Lexus, followed by the only full-sized car in the set – Buick Riviera. The mid-sized cars Pontiac Firebird, Honda Accord, Honda Prelude, Toyota Celica, and Ford Mustang follow next. Hyundai Sonata, Plymouth Sundance, and Volkswagen Golf, the small sized, low cost cars, come at the end. The bipolar Luxuriousness-Sportiness dimension clearly separates the pure sports or sporty cars such as Ferrari, Ford Mustang, Pontiac Firebird, Alpha Romeo, etc. from the big, luxurious sedans such as Lincoln Town Car, Buick Riviera, etc. from the non- sporty, non-luxury cars such as Honda

Table 1: CLUSCALE, INDSCAL, and INDCLUS Solutions: Variance Explained (VAF) for different numbers of dimensions and features

| | Number of Clusters | | | | | |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 |
| # Dimensions | | | | | | |
| 0 | | 19.9 | 24.2 | 32.3 | 37.4 | 40.0 |
| 1 | 37.8 | 62.5 | 64.6 | 66.8 | 68.1 | 69.9 |
| 2 | 45.4 | 69.0 | 70.7 | 72.1 | 72.5 | 72.9 |
| 3 | 52.1 | 72.9 | 73.7 | 74.0 | 74.5 | 75.0 |
| 4 | 57.2 | 74.9 | 75.3 | 75.6 | 75.8 | 76.3 |
| 5 | 60.3 | 76.5 | 76.9 | 77.1 | 77.2 | 77.5 |

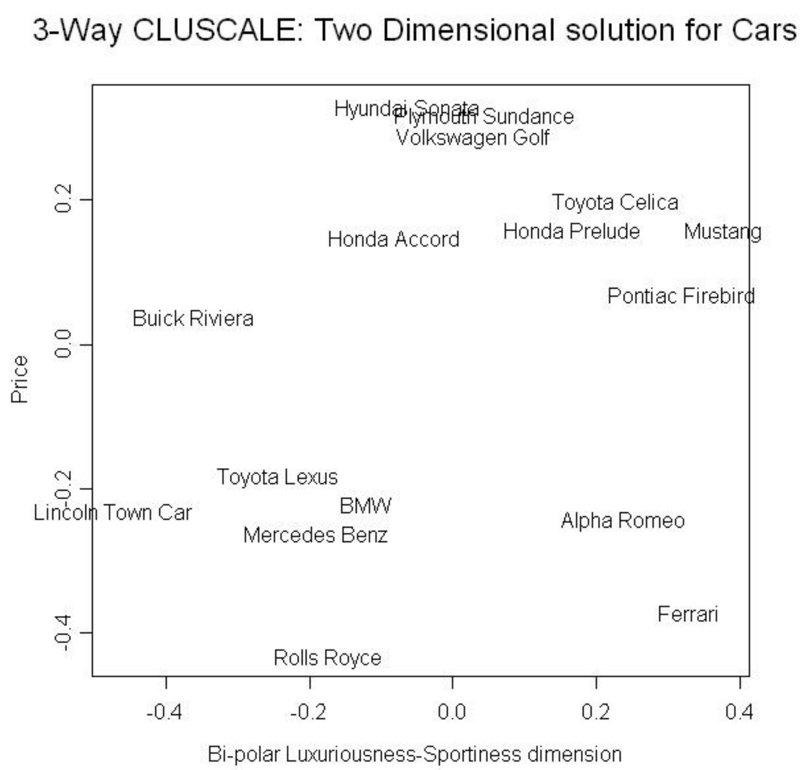


Figure 1. Three-Way CLUSCALE: 2 Dim. Solution

Table 2: CLUSCALE: Discrete Features
Three-Way Solution: 3 Overlapping Discrete Features

| | Very Expensive and Very cheap cars | American Cars | Popular Continental European and Japanese Cars |
|---|---|---|---|
| **Alpha Romeo** | 0 | 0 | 1 |
| **BMW** | 0 | 0 | 1 |
| **Buick Riviera** | 0 | 1 | 0 |
| **Ferrari** | 1 | 0 | 0 |
| **Ford Mustang** | 0 | 1 | 0 |
| **Honda Accord** | 0 | 0 | 1 |
| **Honda Prelude** | 0 | 0 | 1 |
| **Hyndai Sonata** | 1 | 0 | 0 |
| **Lincoln Town Car** | 0 | 1 | 0 |
| **Mercedes Benz** | 0 | 0 | 1 |
| **Plymouth Sundance** | 1 | 1 | 0 |
| **Pontiac Firebird** | 0 | 1 | 0 |
| **Rolls Royce** | 1 | 0 | 0 |
| **Toyota Celica** | 0 | 0 | 1 |
| **Toyota Lexus** | 0 | 0 | 1 |
| **Volkswagen Golf** | 1 | 0 | 1 |

Accord, Hyundai Sonata, Plymouth Sundance, and Volkswagen Golf, which are placed near the origin of this axis.

Table 2 presents the solution for the three discrete perceptual features extracted from the cars data. Two of the discrete features are clearly interpreted as an "American cars" feature, and a "Popular Continental European and Japanese cars" feature. The "American cars" feature includes all American cars in the set of cars presented to the subjects, and only the American cars (Buick Riviera, Ford Mustang, Lincoln Town Car, Plymouth Sundance, and Pontiac Firebird). The "popularly or affordably priced Continental European and Japanese cars" feature includes all European cars (except Ferrari), and all Japanese cars. We modify "European" by "continental", as opposed to British or Irish, to exclude Rolls Royce, which is a British car, and so not a "*Continental* European" car. Rolls Royce could also conceivably be excluded from this cluster/feature for similar reasons as those given below for Ferrari not being included in this cluster. Alternatively, we could *drop* the modifier "Continental" and argue that Rolls Royce is excluded not because it's not a Continental car, but because it's not affordably priced. In order to test this alternative hypothesis, we'd need to include some inexpensive British or Irish cars in the choice set. We use the phrase “popular or affordably priced” in the feature above to mean cars that have wide ownership in the car market. Ferrari, by virtue of its high price and niche positioning in the car market, has a narrow ownership (low market penetration), and is relatively “less popular” in that sense. Hence, when we say that Ferrari is not a popular car, it simply means that not too many people might have the wherewithal to buy this car, and hence, might never consider buying it, so that this car may not be “popular” among these consumers. The third feature that we derived, we termed “very expensive and very cheap cars.” This feature was very stable since it showed up across multiple solutions, as we varied the values of $R$ (the number of continuous perceptual dimensions) and $T$ (the number of discrete perceptual features). This feature includes the two super luxury cars – Ferrari and Rolls Royce, and the three small sized/economy cars – Hyundai Sonata, Plymouth Sundance, and Volkswagen Golf. It could also be characterized as the complement of a feature called “moderately priced cars,” or as “not in choice consideration set” because the cars are either too extremely high priced to be affordable, or too low priced to offer good “perceived quality.” It could also mean, for a similar reason, that this feature includes cars with which these particular consumers tested had a low degree of familiarity because these are cars they “will not consider buying at all.”

Table 3 presents the importance weights of the 24 consumers for the two normalized quantitative perceptual dimensions, the three normalized qualitative perceptual features, and the additive constant (representing a weight for a normalized qualitative perceptual feature corresponding to the universal set with all the brands in it).[1] This table captures the heterogeneity that exists in the marketplace. Consumer 3, for example, differentiates cars based primarily on price (weight of 0.32), and the qualitative perceptual features: “American cars” (weight of 0.25), and “Popular European and Japanese cars” (weight of 0.14). Consumer 8, on the contrary, has relatively higher weights across all five perceptual dimensions and features, representing a pretty well informed and discriminating consumer. Consumer 15 is more like Consumer 3, using the price dimension and the “American Car” feature primarily for differentiation of cars.

1. All important weights have been rescaled for unit length quantitative dimensions and qualitative features. All the quantitative dimensions and qualitative features have been normalized to unit length. This was accomplished by dividing each quantitative dimension and qualitative feature vector by the square root of the sum of squares of the stimulus coordinates, whether these are continuous values (for the quantitative dimensions) or (0,1) binary indicator variables encoding cluster membership/feature possession in the case of discrete features.

Table 3: Three-Way CLUSCALE: Consumer Importances for the 2 Dimensions and 3 discrete clusters + Constant

| | Dimensions | | Discrete Features | | | |
|---|---|---|---|---|---|---|
| **Consumer** | **Bipolar Luxuriousness/ Sportiness** | **Price** | **Very Expensive and Very Cheap Cars** | **American Cars** | **Popular European and Japanese Cars** | **Constant** |
| 1 | 0.16 | 0.05 | 0.03 | 0.07 | 0.07 | -0.10 |
| 2 | 0.12 | 0.07 | 0.03 | 0.02 | 0.01 | -0.06 |
| 3 | 0.10 | 0.32 | 0.08 | 0.25 | 0.14 | -0.19 |
| 4 | 0.06 | 0.38 | 0.09 | 0.07 | 0.08 | -0.10 |
| 5 | 0.07 | 0.40 | 0.06 | 0.05 | 0.04 | -0.07 |
| 6 | 0.27 | 0.28 | 0.05 | 0.13 | 0.12 | -0.13 |
| 7 | 0.28 | 0.19 | 0.01 | 0.04 | 0.05 | -0.05 |
| 8 | 0.29 | 0.38 | 0.18 | 0.23 | 0.24 | -0.26 |
| 9 | 0.14 | 0.23 | 0.03 | 0.03 | 0.05 | -0.05 |
| 10 | 0.22 | 0.25 | 0.10 | 0.06 | 0.10 | -0.13 |
| 11 | 0.19 | 0.38 | 0.05 | 0.06 | 0.04 | -0.06 |
| 12 | 0.13 | 0.41 | 0.06 | 0.04 | 0.04 | -0.06 |
| 13 | 0.10 | 0.31 | 0.04 | 0.05 | 0.05 | -0.07 |
| 14 | 0.17 | 0.18 | 0.04 | 0.02 | 0.02 | -0.10 |
| 15 | 0.12 | 0.32 | 0.05 | 0.15 | 0.10 | -0.14 |
| 16 | 0.10 | 0.35 | 0.04 | 0.06 | 0.02 | -0.05 |
| 17 | 0.21 | 0.29 | 0.00 | 0.13 | 0.07 | -0.08 |
| 18 | 0.08 | 0.40 | 0.06 | 0.08 | 0.08 | -0.10 |
| 19 | 0.15 | 0.11 | 0.03 | 0.09 | 0.13 | -0.12 |
| 20 | 0.21 | 0.44 | 0.12 | 0.16 | 0.17 | -0.18 |
| 21 | 0.13 | 0.35 | 0.02 | 0.03 | 0.02 | -0.05 |
| 22 | 0.22 | 0.29 | 0.04 | 0.06 | 0.07 | -0.09 |
| 23 | 0.26 | 0.25 | 0.13 | 0.10 | 0.15 | -0.18 |
| 24 | 0.25 | 0.30 | 0.13 | 0.08 | 0.07 | -0.11 |

In Table 4, we present the additive decomposition of the VAF from the CLUSCALE model to VAFs attributable to the quantitative dimensions and the qualitative features. We use the approach discussed earlier for partitioning VAF based on summing VAF for the continuous spatial and discrete feature-like structure. This, as argued earlier, results in a unique partitioning of the total VAF. Table 4 reveals the VAF decomposition across all 24 subjects, and across the order of entry of the individual components in the decomposition (the quantitative dimensions and the qualitative features). It can be seen that the order of entry of the two components does not change the VAF drastically. As a final step in partitioning VAF, we average the two sets of VAFs (entering first and entering second) for each of the two components of the CLUSCALE solution to arrive at the additively decomposed VAFs components for each subject and overall.

**Table 4: Decomposition of CLUSCALE Solution VAF Into Additive Components**

| Subject | Combined VAF | Quantitative Dimensions entering First | Quantitative Diemenions entering second | Average VAF for Quantitative Dimensions | Qualitative Features entering first | Qualitative Features entering Second | Average VAF for Qualitative Features |
|---|---|---|---|---|---|---|---|
| 1 | 0.33 | 0.02 | 0.01 | 0.02 | 0.32 | 0.31 | 0.31 |
| 2 | 0.16 | 0.13 | 0.12 | 0.13 | 0.04 | 0.02 | 0.03 |
| 3 | 0.76 | 0.31 | 0.29 | 0.30 | 0.47 | 0.45 | 0.46 |
| 4 | 0.95 | 0.68 | 0.67 | 0.67 | 0.29 | 0.27 | 0.28 |
| 5 | 0.79 | 0.56 | 0.55 | 0.55 | 0.24 | 0.23 | 0.24 |
| 6 | 0.77 | 0.48 | 0.47 | 0.47 | 0.31 | 0.30 | 0.30 |
| 7 | 0.75 | 0.53 | 0.52 | 0.53 | 0.22 | 0.22 | 0.22 |
| 8 | 0.88 | 0.45 | 0.43 | 0.44 | 0.45 | 0.44 | 0.44 |
| 9 | 0.80 | 0.56 | 0.56 | 0.56 | 0.24 | 0.24 | 0.24 |
| 10 | 0.88 | 0.52 | 0.50 | 0.51 | 0.37 | 0.35 | 0.36 |
| 11 | 0.96 | 0.73 | 0.72 | 0.72 | 0.24 | 0.23 | 0.23 |
| 12 | 0.74 | 0.52 | 0.51 | 0.52 | 0.23 | 0.22 | 0.22 |
| 13 | 0.83 | 0.58 | 0.57 | 0.58 | 0.26 | 0.25 | 0.25 |
| 14 | 0.55 | 0.28 | 0.22 | 0.25 | 0.33 | 0.27 | 0.30 |
| 15 | 0.76 | 0.42 | 0.40 | 0.41 | 0.36 | 0.34 | 0.35 |
| 16 | 0.97 | 0.74 | 0.73 | 0.74 | 0.24 | 0.23 | 0.24 |
| 17 | 0.61 | 0.36 | 0.35 | 0.35 | 0.26 | 0.25 | 0.26 |
| 18 | 0.71 | 0.45 | 0.44 | 0.45 | 0.27 | 0.26 | 0.26 |
| 19 | 0.37 | 0.05 | 0.04 | 0.05 | 0.32 | 0.32 | 0.32 |
| 20 | 0.96 | 0.62 | 0.60 | 0.61 | 0.36 | 0.34 | 0.35 |
| 21 | 0.65 | 0.44 | 0.43 | 0.43 | 0.22 | 0.21 | 0.22 |
| 22 | 0.79 | 0.54 | 0.52 | 0.53 | 0.26 | 0.25 | 0.26 |
| 23 | 0.56 | 0.23 | 0.22 | 0.22 | 0.34 | 0.33 | 0.34 |
| 24 | 0.75 | 0.45 | 0.44 | 0.44 | 0.31 | 0.30 | 0.31 |
| Average | **0.72** | 0.44 | 0.43 | **0.44** | 0.29 | 0.28 | **0.28** |

### 6.3 Analysis of the Car Data Via the INDCLUS Model (Using SINDCLUS Program) and the Extended INDSCAL (or EXSCAL) Model, Using the METric EXSCAL (or METEXSCAL) Program

We also analyzed these data separately via the INDSCAL and INDCLUS models, in order to compare the results and make a more convincing case for the use of the hybrid CLUSCALE model combining both the continuous dimensions characteristic of INDSCAL and the discrete features or clusters characteristic of INDCLUS.

We applied a *generalization* of INDSCAL called EXSCAL (Carroll and Winsberg 1995), which stands for Extended INDSCAL, and assumes "specific" as well as "common" dimensions underlying the stimuli (which is, in some sense analogous to the discrete features assumed in the INDCLUS model). The version of EXSCAL we used is a metric one called METEXSCAL (Chaturvedi 1993; Chaturvedi and Carroll 1998) rather than the "quasi-nonmetric" approach proposed by Carroll and Winsberg (1995) because we feel this method is more directly comparable to the CLUSCALE model, since CLUSCALE is *also* a metric approach. The METEXSCAL model is also more general than EXSCAL, since it allows completely unique "specificities" for each combination of a subject and a stimulus, whereas EXSCAL constrains the specificities to be a product of a general specificity for the stimulus multiplied by a weight (analogous to an INDSCAL subject weight) for the particular individual subject. For these reasons we feel METEXSCAL is more appropriate for this comparative analysis than EXSCAL, *or* than the INDSCAL model without specificities. Also, other analyses have demonstrated that the INDSCAL analysis results in stimulus dimensions and subject weights for these dimensions that are very similar to those in the METEXSCAL analysis.

As for the INDCLUS analysis, we used a method called SINDCLUS devised by Chaturvedi and Carroll (1994) which uses a somewhat different algorithm than Carroll and Arabie (1983) used to fit this model. The SINDCLUS algorithm is much faster and more efficient than the earlier INDCLUS approach, and has been shown essentially always to produce the same or better solutions for a given set of data, especially for larger data sets.

## 7. The METEXSCAL Analysis

The METEXSCAL analysis reported here was first reported by Chaturvedi (1993) in his dissertation, and later also described by Chaturvedi and Carroll (1998). After attempting the analysis in dimensionalities ranging from 1 to 5, the decision was made, based on both VAF as a function of dimensionality *and* on interpretability of the solutions, that the most appropriate dimensionality was four. The four common dimensions were interpreted as "Luxuriousness" (Dimension 1), "Sportiness" (Dimension 2), "Fuel Economy" (Dimension 3) and "Acceleration" (Dimension 4).

The basis for this interpretation of the dimensions, as well as the justification for choosing the four-dimensional structure, is provided by Chaturvedi (1993) and by Chaturvedi and Carroll (1998), and will not be repeated here. The two dimensional plots of each of dimensions 2, 3 and 4 plotted against dimension 1 are given in Figures 2, 3, and 4 respectively. The importance weights (or saliences) of these dimensions for each of the 24 subjects are provided in Table 5.

This four dimensional METEXSCAL solution accounted for 68.7% of the variance in the derived scalar products, but this required a total of $(n-2)R + mR + mn = 536$ parameters (for $n=16$ stimuli [cars], $m = 24$ subjects and $R = 4$ dimensions). Chaturvedi (1993) also fit an ordinary INDSCAL model (with no specificities) to these data, which accounted for 57.2% of the variance (with a total of 152 parameters for the four common dimensions and their subject importance weights). A pseudo – $F$ test to compare the METEXSCAL model to the INDSCAL model (both in four dimensions) indicated that, despite the large number of additional parameters, METEXSCAL actually fit these data statistically significantly better than INDSCAL (as indicated by a $p$ value associated with the pseudo – $F$ statistic of $p < 0.001$, which, even though this statistic cannot be assumed to be distributed precisely as $F$ with the appropriate degrees of freedom, nevertheless suggests strongly that METEXSCAL fits significantly better than INDSCAL). A similar pseudo – $F$ test was also used to ascertain that the four dimensional METEXSCAL model fit significantly better than the three-dimensional model did. In contrast to these two versions of the INDSCAL model, with and without specificities, the CLUSCALE solution accounted for 72.1% of the variance with two quantitative (spatial, or INDSCAL-like) dimensions and three qualitative (or discrete, INDCLUS-like) features. The number of parameters fit by this ($R = 2$ and $T = 3$) CLUSCALE solution was $R(m + n - 2) + Tn + (T + 1)m = 220$ (using a very conservative formula that assumes that one binary parameter equals one continuous one, which in fact, is highly unlikely). Summarizing these comparative results we have the four dimensional METEXSCAL model accounting for 68.7% of the variance with 536 parameters, INDSCAL (also in four dimensions) accounting for 57.2% of the variance with 152 parameters, while three-way CLUSCALE (with two continuous dimensions and three discrete features) accounts for 72.1% of the variance with 220

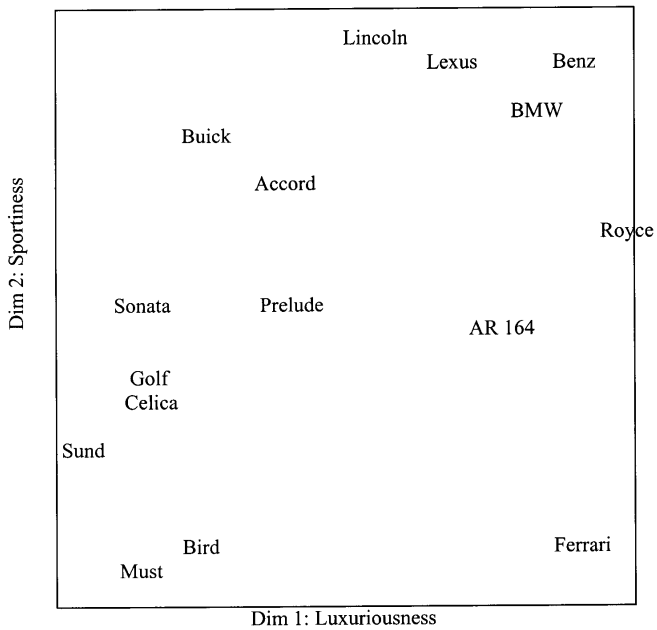


Figure 2. METEXSCAL Solution: Dimension 1 vs. 2 for cars

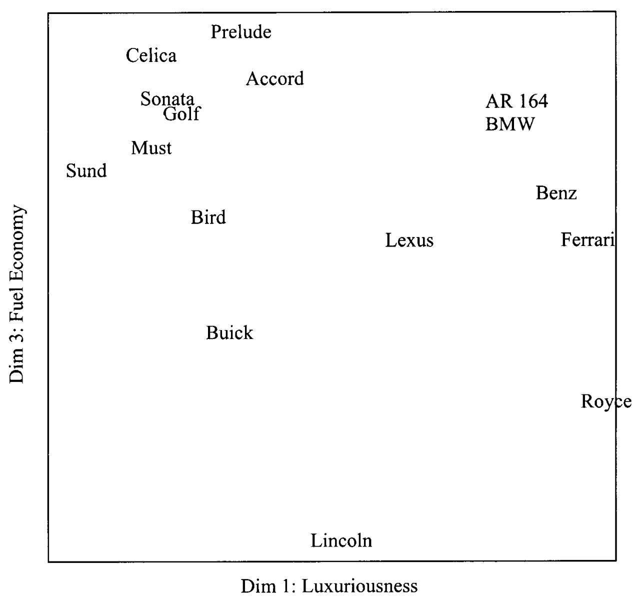


Figure 3. METEXSCAL Solution: Dimension 1 vs. 3 for cars

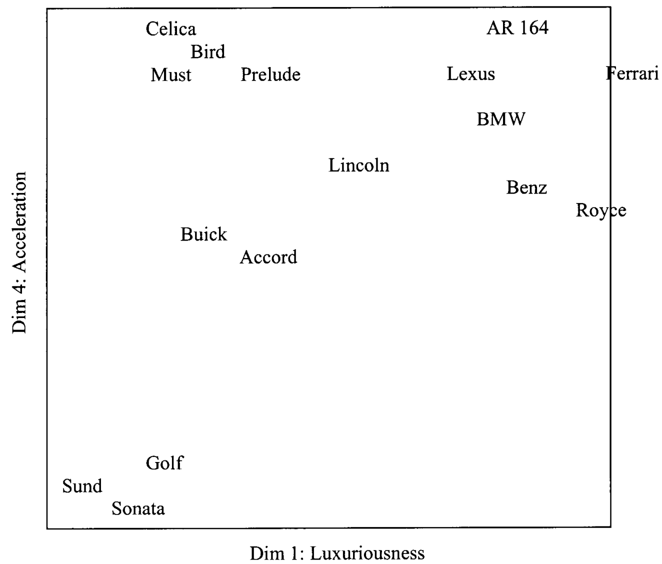


Figure 4. METEXSCAL Solution: Dimension 1 vs. 4 for cars

parameters (a conservative upper bound, we would argue). Since METEXSCAL fit the data significantly better than INDSCAL, it would appear that the only relevant statistical comparison of these results should be between the METEXSCAL and three-way CLUSCALE model. Since the CLUSCALE model accounts for over 3.4% more variance than does the METEXSCAL model, with only about 40% of the number of parameters, we don't even have to attempt a statistical test, since it is clear that the CLUSCALE model (with $R$=2 and $T$=3) does fit these data quite significantly better than does the METEXSCAL (*or* the ordinary INDSCAL) model with R=4 dimensions.

One question that remains to be answered : What about ignoring the interpretable structure present in dimensions 3 and 4 and the specificities of the METEXSCAL solution if we select the CLUSCALE solution over the METEXSCAL solution? There are two explanations. The first explanation lies in looking at the degree of superior fit achieved by the CLUSCALE solution compared to the METEXSCAL solution. Since the CLUSCALE solution out-performs the METEXSCAL solution comprehensively on statistical grounds, we have to choose the better-fitting (more parsimonious) *and* interpretable solution of CLUSCALE over the not-as-well fitting but interpretable solution of METEXSCAL. The second explanation for preferring the CLUSCALE solution is that it reveals structure (American

Table 5. Importance Weights in Metexscal Common Space for Cars

| SUBJECT | LUXURIOUSNESS | SPORTINESS | FUEL-ECONOMY | ACCELERATION |
|---|---|---|---|---|
| 1 | 0.00 | 0.16 | 0.38 | 0.11 |
| 2 | 0.02 | 0.04 | 0.19 | 0.02 |
| 3 | 0.50 | 0.06 | 0.17 | 0.05 |
| 4 | 0.69 | 0.21 | 0.08 | 0.10 |
| 5 | 0.56 | 0.14 | 0.09 | 0.02 |
| 6 | 0.37 | 0.39 | 0.11 | 0.14 |
| 7 | 0.24 | 0.56 | 0.23 | 0.00 |
| 8 | 0.30 | 0.34 | 0.37 | 0.08 |
| 9 | 0.34 | 0.23 | 0.40 | 0.32 |
| 10 | 0.23 | 0.33 | 0.44 | 0.24 |
| 11 | 0.44 | 0.26 | 0.32 | 0.23 |
| 12 | 0.34 | 0.09 | 0.44 | 0.15 |
| 13 | 0.49 | 0.11 | 0.35 | 0.14 |
| 14 | 0.13 | 0.16 | 0.21 | 0.14 |
| 15 | 0.47 | 0.14 | 0.14 | 0.00 |
| 16 | 0.68 | 0.21 | 0.17 | 0.15 |
| 17 | 0.43 | 0.30 | 0.12 | 0.00 |
| 18 | 0.54 | 0.04 | 0.23 | 0.03 |
| 19 | 0.02 | 0.15 | 0.53 | 0.00 |
| 20 | 0.47 | 0.29 | 0.24 | 0.21 |
| 21 | 0.44 | 0.16 | 0.25 | 0.00 |
| 22 | 0.37 | 0.37 | 0.22 | 0.01 |
| 23 | 0.16 | 0.26 | 0.12 | 0.25 |
| 24 | 0.26 | 0.26 | 0.29 | 0.42 |

cars vs. foreign cars) that has been shown to be extremely important in people's choice of cars (O'Boyle 1996), and that can more reasonably be expected be more salient in determining people's similarity judgments than dimensions 3 and 4 of METEXSCAL. However, METEXSCAL does provide certain interpretations that might not be captured by CLUSCALE such as the relatively salient weights for subjects 2 and 19 for Fuel Economy, and for Subject 24 for Acceleration (Table 5). Moreover, METEXSCAL clearly separates Luxuriousness and Sportiness, which might be advantageous sometimes.

## 8. The INDCLUS Analysis

The SINDCLUS method (Chaturvedi and Carroll, 1994) of fitting the INDCLUS (Carroll and Arabie, 1983) model was then applied to the 24 proximity matrices corresponding to the 24 subjects. We obtained a variety of INDCLUS solutions ranging from two through eight clusters. Table 1 presents the variance accounted for as the number of extracted clusters was varied from one through five. We found only the two-cluster solution to be interpretable. The solutions with three to eight clusters/features all yielded two clusters which had the same interpretations as those in the two-cluster solution, but the remaining clusters in these solutions were not interpretable at all. The details of the two cluster INDCLUS solution are presented in Table 6.

The two-cluster solution was a partitioning solution, separating the Luxury cars (Alpha Romeo, BMW, Ferrari, Lincoln Town Car, Mercedes Benz and Toyota Lexus) from the Non-Luxury cars (Ford Mustang, Honda Accord, Honda Prelude, Toyota Celica, Pontiac Firebird, Plymouth Sundance, and Volkswagen Golf). Buick Riviera was the only car that was not a part of either cluster.

In order to minimize the chance of merely locally optimal SINDCLUS solutions, we tried up to 300 random starts for each solution. Clearly, the INDCLUS model was failing to capture all the discrete structure present in the data. Since this two-feature INDCLUS solution accounts for only 24.2% of the variance, this is clearly not, by itself, providing a satisfactory account of these data. Although, as mentioned above, the third, as well as all the other "higher order" features were not at all interpretable, we can nevertheless consider what the VAF was for these solutions. This measure of fit (VAF) was only 32.3% for the three feature solution, and even with as many as eight (8) features only reached a value of 45.5%. So, if we follow the often stated "upper bound" that the number of clusters in an overlapping clustering model should be about half the number of objects, we still find that, at best, we account for less than half the variance (45.5%), with a number of parameters (counted in the same conservative manner as described earlier) equal to $nT + m\ (T + 1) = (16)\ (8) + 24\ (9) = 344$!

Based on the pattern of results obtained via these METEXSCAL, INDSCAL and INDCLUS analyses of these data, we conclude that the three-way CLUSCALE model and the solutions obtained for these illustrative data by optimally fitting this hybrid model to them *can not* in any sense be inferred from these separate analyses. In fact, the two dimensions obtained via three-way CLUSCALE do not resemble closely any of the dimensions of those fit via METEXSCAL or INDSCAL.

Rather, they appear to be meaningful composites of these and (perhaps) of some of the features exhibited in the INDCLUS analysis. For example, the first dimension is a *bipolar* one *contrasting* luxuriousness and sportiness – combining in a meaningful and interpretable manner these two

**Table 6: Two Cluster INDCLUS Solution**

| *Cars* | **Luxury** | **Non-Luxury** |
|---|---|---|
| Alpha Romeo | 1 | 0 |
| BMW 325 I | 1 | 0 |
| Buick Riviera | 0 | 0 |
| Ferrari | 1 | 0 |
| Ford Mustang | 0 | 1 |
| Honda Accord | 0 | 1 |
| Honda Prelude | 0 | 1 |
| Hyundai Sonata | 0 | 1 |
| Lincoln Town Car | 1 | 0 |
| Mercedes Benz 190 | 1 | 0 |
| Plymouth Sundance | 0 | 1 |
| Pontiac Firebird | 0 | 1 |
| Rolls Royce | 1 | 0 |
| Toytoa Celica | 0 | 1 |
| Tyota Lexus | 1 | 0 |
| Volkswagen Golf | 0 | 1 |

separate dimensions that emerged from the METEXSCAL and INDSCAL analyses. The three features in this hybrid three-way CLUSCALE solution do not resemble the two highly complementary features of "luxury" and "non-luxury" found in the INDCLUS analysis in any way. They are based on entirely different aspects of the stimuli – e.g., the American vs. "foreign" cars, cars that are moderately priced vs. either very *in*expensive or very *ex*pensive ones, and Continental European or Japanese cars vs. the opposite.

These results make it very clear, we feel, that the hybrid three-way CLUSCALE model, as well as the corresponding *method* for fitting it, is a genuinely *different* model from any simple combination of separate INDSCAL and INDCLUS models for these data, while, consistent with this fact, fitting the INDSCAL model followed by independently fitting the INDCLUS model to the same data yields little if any insight into the precise nature of the optimal hybrid CLUSCALE model for these data. This CLUSCALE solution is clearly a case in which "the whole is more than the sum of its parts".

## 9. Conclusions

This paper has presented a new perceptual mapping technique CLUSCALE that enhances traditional techniques for perceptual mapping, such as two-way multidimensional scaling and factor analysis in two important ways – (a) by explaining product differentiation that might exist due to perceptual features that are qualitative (i.e., possessed, in an all or none fashion, by a subset of the brands being analyzed), in addition to quantitative or continuously varying perceptual dimensions, and (b) by incorporating heterogeneity in the form of different profiles of weights for different consumers/subjects or consumer segments for these derived quantitative perceptual dimensions and qualitative perceptual features.

The paper also presents a methodology for least squares parameter estimation, which can be used in psychological, marketing research, and other social and behavioral science applications. For example, advertising agencies, and advertising tracking agencies which periodically monitor a brand's or a category's performance, can now track shifts in consumers' perceptions regarding various brands over time by using multiple longitudinal measurements from the same respondents. More generally, this methodology can be used to develop quite powerful perceptual maps of stimuli such as the brands or products in this marketing example reflecting both the effects and nature of continuous dimensions on which the brands or products vary continuously and of qualitative features representing all or none attributes or features which a brand or product either possesses or does not possess, inducing an overlapping cluster structure on the brands, products or other stimuli. Differential weights for both the continuous dimensions and the qualitative features allow the user (e.g., the marketing manager in the current example) to assess the degree and nature of stimulus homogeneity – e.g., market homogeneity vs. heterogeneity—while the nature of the three-way data analyzed and the CLUSCALE model fit leads to this representation being rotationally uniquely determined .

We would argue that these attributes of the CLUSCALE model, considered as a whole, make it a very comprehensive model and method for perceptual mapping, based on three-way or individual differences (direct or derived) proximity data. When combined with appropriate preferential choice data on the same stimuli (e.g., brands or products), using preference mapping (Carroll 1980) combined with conjoint analysis (Green and Srinivasan 1990) to account for the preference data in terms of the recovered quantitative dimensions and qualitative features, respectively, this should lead to a very methodologically complete and interpretively satisfying approach to representation of a group of subjects' perceptions *and* preferences for a set of stimuli, or in the marketing context, for market structure analysis.

## References

ARABIE, P., and CARROLL, J.D. (1980), "MAPCLUS: A Mathematical Programming Approach to Fitting the ADCLUS Model", *Psychometrika*, *45*, 211–235.

ARABIE, P., CARROLL, J.D., and DESARBO, W.S. (1987), *Three Way Scaling and Clustering*, Newbury Park, California: Sage.

CARROLL, J.D. (1972), "Individual Differences and Multidimensional Scaling", in *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*, (Vol. 1, pp. 105–155), Eds. R.N. Shepard, A.K. Romney and S.B. Nerlove, New York: Seminar Press. [Reprinted (1984) in *Key Texts on Multidimensional Scaling* (pp. 267–301), Eds. P. Davies and A.P.M. Coxon, Portsmouth, NH: Heinemann.]

CARROLL, J.D. (1976), "Spatial, Non-spatial and Hybrid Models for Scaling" (Presidential Address for Psychometric Society), *Psychometrika*, *41*, 439–463.

CARROLL, J.D. (1980), "Models and Methods for Multidimensional Analysis of Preferential Choice (Or Other Dominance) Data", in *Similarity and Choice* (pp. 234–289), Eds. E.D. Lantermann and H. Feger, Bern: Hans Huber.

CARROLL, J.D., and ARABIE, P. (1983), "INDCLUS: An Individual Differences Generalization of the ADCLUS Model and the MAPCLUS Algorithm", *Psychometrika*, *48*, 157–169.

CARROLL, J.D., and ARABIE, P. (1997), "Multidimensional Scaling", in *Handbook of Perception and Cognition, Volume 3: Measurement, Judgment, and Decision Making* (pp. 179–250), Ed. M.H. Birnbaum, San Diego, CA: Academic Press.

CARROLL, J.D., and CHANG, J.J. (1970), "Analysis of Individual Differences in Multidimensional Scaling via an N-way Generalization of Eckart-Young Decomposition", *Psychometrika*, *35*, 283–319. [Reprinted (1984) in *Key Texts in Multidimensional Scaling* (pp. 229–252), Eds. P. Davies and A.P.M. Coxon, Portsmouth, NH: Heinemann.]

CARROLL, J.D., and CHATURVEDI, A.D. (1995), "A General Approach to Clustering and Scaling of Two-way, Three-way, or Higher-way Data", in *Geometric Representations of Perceptual Phenomena* (pp. 295–318), Eds. R.D. Luce, M. D'Zmura, D. Hoffman, G.J. Iverson and A.K. Romney, Mahwah, NJ: Erlbaum.

CARROLL, J.D., and GREEN, P.E. (1997), "Psychometric Methods in Marketing Research: Part II, Multidimensional Scaling", *Journal of Marketing Research*, 34, 193–204.

CARROLL, J.D., and PRUZANSKY, S. (1984), "The CANDECOMP-CANDELINC Family of Models and Methods for Multidimensional Data Analysis", in *Research Methods for Multimode Data Analysis* (pp. 372–402), Eds. H.G. Law, C.W. Snyder, J.A. Hattie and R.P. McDonald, New York: Praeger.

CARROLL, J.D., and WINSBERG.S. (1995), "Fitting an Extended INDSCAL Model to Three-way Proximity Data", *Journal of Classification*, *12*, 57–71.

CHATURVEDI, A.D. (1993), "Perceived Product Uniqueness in Product Differentiation and Product Choice", *9316296*, Ann Arbor, Michigan: University Microfilms International.

CHATURVEDI, A.D., and CARROLL, J.D. (1994), "An Alternating Combinatorial Approach to Fitting the INDCLUS and Generalized INDCLUS Models", *Journal of Classification*, *11*, 155–170.

CHATURVEDI, A.D., and CARROLL, J.D. (1998), "A Perceptual Mapping Procedure for the Analysis of Proximity Data to Determine Common and Unique Product-Market Structures", *European Journal of Operational Research*, *111*, 268–284.

CHATURVEDI, A.D., and CARROLL, J.D. (2000), "Two-way and Three-Way HYCLUS: A Hybrid Individual Differences Model Combining INDSCAL and INDCLUS Structures" presented at Classification Society of North America (CSNA) meeting in Montreal, Canada, June 2000.

DICKSON, R.P., and GINTER, J.L. (1987), "Market Segmentation, Product Differentiation, and Marketing Strategy", *Journal of Marketing*, *51*, 1–10.

ELIASHBERG, J.E., and MANRAI, A.K. (1992), "Optimal Positioning of New Product Concepts: Some Analytical Implications and Empirical Results", *European Journal of Operational Research*, *63*, 376–397.

GREEN, P.E., CARROLL, J.D., and DESARBO, W.S. (1978), "A New Measure of Predictor Variable Importance in Regression", *Journal of Marketing Research*, *15*, 356-360.

GREEN, P.E., and SRINIVASAN, V. (1990), "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice", *Journal of Marketing, 54,* 3–19.

HARSHMAN, R.A., and LUNDY, M.E. (1984), "The PARAFAC Model for Three-way Factor Analysis and Multidimensional Scaling", in *Research Methods for Multimode Data Analysis* (pp. 122–215), Eds. H.G. Law, C.W. Snyder, J.A. Hattie and R.P. McDonald, New York: Praeger.

KRUSKAL, J.B. (1964a), "Multidimensional Scaling by Optimizing Goodness of Fit to a Non-metric Hypothesis", *Psychometrika*, *29*, 1–27.

KRUSKAL, J.B. (1964b), "Nonmetric Multidimensional Scaling: A Numerical Method", *Psychometrika*, *29*, 115–129.

KRUSKAL, J.B., and WISH, M. (1978), *Multidimensional Scaling*, Newbury Park, CA: Sage.

MANRAI, A.K., and SINHA, P.K. (1989), "Elimination-by-cutoffs", *Marketing Science*, *8*, 132–152.

NAVARRO, D.J., and LEE, M.L. (2003), "Combining Dimensions and Features in Similarity-based Representations", in *Advances in Neural Information Processing Systems*, *15*, pp. 67-74, Eds. S. Becker, S. Thrun, and K. Obermayer, Cambridge: MIT Press.

O'BOYLE (1996), "The Culture of Decision-Making", *R &D Innovator, 5 (12).*

PRUZANSKY, S. (1975), *How to Use SINDSCAL: A Computer Program for Individual Differences in Multidimensional Scaling*, Murray Hill, NJ: AT&T Bell Laboratories.

SHEPARD, R.N., and ARABIE, P. (1979), "Additive Clustering: Representation of Similarities as Combinations of Discrete Overlapping Properties", *Psychological Review*, *86*, 87–123.

TEN BERGE, J.M.F., and KIERS, H.A.L. (2005), "A Comparison of Two Methods for Fitting the INDCLUS Model", *Journal of Classification, 22*, 273-286.

TORGERSON, W.S. (1952), "Multidimensional Scaling: I. Theory and Method", *Psychometrika*, *17*, 401–419.

TORGERSON, W.S. (1958), *Theory and Methods of Scaling*, New York: John Wiley.