MIXED TREE AND SPATIAL REPRESENTATION

OF DISSIMILARITY JUDGMENTS

Michel Wedel

Department of Business Administration
University of Groningen
PO Box 800
9700 AV Groningen
The Netherlands
m.wedel@eco.rug.nl

and

Tammo H.A. Bijmolt

Department of Business Administration
Tilburg University
PO Box 90153
5000 LE Tilburg
The Netherlands
t.h.a.bijmolt@kub.nl

October 1998

Abstract

Whereas previous research has shown that either tree or spatial representations of dissimilarity judgments may be appropriate, focussing on the comparative fit at the aggregate level, we investigate whether there is heterogeneity among subjects in the extent to which their dissimilarity judgments are better represented by ultrametric tree or spatial multidimensional scaling models. We develop a mixture model for the analysis of dissimilarity data, that is formulated in a stochastic context, and entails a representation and a measurement model component. The latter involves distributional assumptions on the measurement error, and enables estimation by maximum likelihood. The representation component allows dissimilarity judgments to be represented either by a tree structure or by a spatial configuration, or a mixture of both. In order to investigate the appropriateness of tree versus spatial representations, the model is applied to twenty empirical data sets. We compare the fit of our model with that of aggregate tree and spatial models, as well as with mixtures of pure trees and mixtures of pure spaces, respectively. We formulate some empirical generalizations on the relative importance of tree versus spatial structures in representing dissimilarity judgments at the individual level.

Key words:

Multidimensional scaling, tree models, mixture models, dissimilarity judgments

JEL codes:

C10, C91, M31

1. Introduction

Perceptions have been studied using graphical representations of dissimilarity judgments of stimuli that either take the form of trees or spaces. The assumption underlying the analysis of dissimilarity judgments is that subjects compare the stimuli on the basis of a number of attributes, that are either discrete features or continuous dimensions (Garner 1978; Johnson and Fornell 1987; Johnson, Lehmann, Fornell, and Horne 1992; Tversky 1977; Tversky and Gati 1978). Those attributes are recovered through the analysis of dissimilarity judgments with models that represent them as a tree (cf. Corter 1996; DeSarbo, Manrai, and Manrai 1993; Sattath and Tversky 1977) or as a space, respectively (cf. Carroll and Arabie 1996; Carroll and Green 1997; Green, Carmone, and Smith 1989). The choice between trees and spaces is based on a) prior theory on the attribute-types discerned by subjects for that particular type of stimuli, b) the basis of the relative fit of the two models, or c) diagnostic measures such as the skewness of the dissimilarity judgments (cf. Ghose 1998; Glazer and Nakamoto 1991; Pruzansky, Tversky, and Carroll 1982).

The question is, however, whether tree structures and spatial configurations should be considered as substitutes or as complements. Carroll (1976, p. 455) stated: “I am increasingly inclined to think of tree structures and spatial structures not so much as competing models as complementary ones, each of which captures certain aspects of a reality which is probably in fact much more complex than either model alone”. Or as formulated by Shepard (1980, p. 397): “It would be a mistake to ask which of these various scaling, tree-fitting, or clustering methods is based on *the* correct model. (..) Different models may be more appropriate for different sets of stimuli or types of data. Even for the same set of data, moreover, different methods of analysis may be better suited to binging out different, but equally informative aspects of the underlying structure." Recently, Ghose (1998) stated: "..items such as the nature of the stimuli and the way consumers process information influence the nature of the input data sets. Coupled with the dimensional- versus feature based structure of spaces and trees, this demonstrates that spaces and trees should be considered complementary approaches for representing data."

These insights have given rise to the development of mixed or hybrid models, i.e. models that contain a tree structure as well as a spatial configuration. In recent literature reviews, hybrid models have been mentioned as one of the important developments in the field of psychometric methods (e.g. Carroll and Arabie 1996; Carroll and Green 1997). However, despite the added value such approaches may have over single tree structure models or spatial MDS models, "much has been said but little done about such mixed or hybrid models" (Carroll and Arabie 1996).

In the literature, only a few hybrid models for dissimilarity judgments have been proposed. An important point to be made here is that although individual differences have been shown to occur both in processing the attributes and in the judgment of the dissimilarity between stimuli (cf. Bijmolt, Wedel, Pieters, and DeSarbo 1998; Johnson and Fornell 1987; Johnson et al. 1992), most previous hybrid models for the analysis of dissimilarity judgments do not account for heterogeneity between subjects (Carroll and Pruzansky 1980; Degerman 1970); the one model of Carroll and Chaturvedi (1995) being an exception.

In this paper we propose a stochastic mixture model of tree and spatial representations for the analysis of dissimilarity judgments, which allows for structural heterogeneity in perception. The mixture model accounts for heterogeneity between subjects in a parsimonious way, namely by identifying two unobserved classes. The dissimilarity judgments of subjects in the first latent class are represented by means of a tree structure, those of the subjects in the second latent class by means of a spatial structure. The stochastic nature of the model allows for assessing which representation is most appropriate, for example by testing the mixed structure of a tree and a space versus a single tree, a single space, two trees, or two spaces. Our model differs importantly from previous mixture models published in the classification, psychometric, and marketing literature, in that it accounts for structural heterogeneity among classes, where previous work has accommodated parametric heterogeneity, assuming classes to be structurally homogeneous (cf. Wedel and Kamakura 1998).

In the remainder of this paper we first discuss the theoretical background of alternative representations of dissimilarity judgments. Next, we present the mixture model of tree and spatial representations. The performance of the model to classify subjects to a tree or spatial representation is demonstrated through the analysis of synthetic data sets. We illustrate our model on cola taste data published by Schiffman, Reynolds, and Young (1981). In addition, we describe the results of analysis of twenty empirical data sets to assess the relative importance of tree structures and spatial configurations. We compare the model with aggregate tree and spatial models, and two class mixtures of pure trees and pure spaces, respectively. Finally, we formulate some empirical generalizations from those analyses, discuss the model and results, and provide directions for future research.

2. Background

2.1. Features versus dimensions

In most studies involving dissimilarity judgments it is assumed that subjects evaluate and compare stimuli on discrete features or continuous dimensions, exclusively (cf. Garner 1978; Johnson and Fornell 1987; Johnson et al. 1992; Tversky 1977; Tversky and Gati 1978). Discrete features are attributes with a small and limited number of values, e.g. whether a particular cola is diet or regular. Continuous dimensions are attributes on which the stimuli vary as a matter of degree, e.g. sweetness of the taste of colas. The way a respondent processes an attribute may affect whether that attribute is used as a discrete feature or as a continuous dimension in brand dissimilarity judgments (Garner 1978; Johnson et al. 1992; Tversky 1977). In the cola example, the continuous attribute cherry flavour, for example, could be used in the judgment process as the presence of absence of that flavour rather than as the degree of that flavour. Alternatively, a set of discrete features may be combined into a continuous dimension.

The processes by which subjects evaluate and compare stimuli to arrive at dissimilarity judgments may be affected by factors related to the stimuli, such as the format by which the stimuli are presented to the subjects (Bijmolt, et al.1998), and by factors related to the subjects, such as the experience and familiarity of the subject with the stimuli (Johnson et al. 1992). The existence of heterogeneity in perceptions has been widely recognized (Eagly and Chaiken 1993; Scott, Osgood, and Peterson 1979), and may be related to personality constructs as cognitve complexity (Bieri 1955) and the style of information processing of subjects, where some subjects have a more verbal and others a more visual style of processing information (Childers, Houston and Heckler 1985; Richardson 1977).

2.2. Tree structures versus spatial configurations

It has been previously found that tree structure models outperform multidimensional scaling methods in fitting empirical customer perceptions of *conceptual* stimuli such as brands (Johnson and Fornell 1987; Johnson and Hudson 1996; Johnson et al. 1992; Pruzansky, Tversky, and Carroll 1982). On the other hand, the fit of multidimensional scaling has been found to be better relative to tree structure models for *perceptual* stimuli (Pruzansky, Tversky, and Carroll

1982) and *abstract* stimuli, such as product categories (Johnson et al. 1992). When considering perceived usefulness and interpretability, spatial configurations appear to outperform tree structures. Johnson and Horne (1992) found that subjects were better able to indicate their perception of a certain brand by representing that brand as a point in a space than as a branch in a tree structure. In addition, Johnson and Hudson (1996) revealed that users found spatial configurations more useful as compared to tree structures.

On the basis of characteristics of the dissimilarity data, that is before the tree or MDS models are fitted, one may decide whether a tree or a space is more appropriate. Ghose (1988) and Pruzansky, Tversky, and Carroll (1982) showed that the skewness of the data helped to discriminate between the two representations, whereas other measures, such as elongation, centrality, and reciprocity, performed less in that respect. The shape of a tree allows for many large distances between stimuli, whereas a low-dimensional space does not. Hence, dissimilarity data with a large negative skewness generally fits a tree structure better relative to a space.

2.3. Hybrid Models

As noted in the introduction, it is generally accepted that features and dimensions on the one hand and trees and spaces on the other are complements rather than substitutes (Carroll 1976; Ghose 1998; Shepard 1980). Despite the added value that approaches that combine trees and spaces may have over single tree structure models or spatial MDS models, only a few hybrid models for dissimilarity judgments have been proposed. Degerman (1970) developed a model which combined continuous dimensions with discrete dimensions, at which the stimuli could take on only a restricted number of values. Carroll and Pruzansky (Carroll 1976; Carroll and Pruzansky 1980) developed a hybrid model that combines multiple tree structures and a single spatial configuration. In their model, the dissimilarity between two stimuli corresponds to the sum of the distances derived from the trees and from the space. The hybrid models mentioned above represent the data at the aggregate level. The estimated model, that is both the tree structure and the spatial configuration, is assumed to hold for all subjects in the sample. However, evidence has been provided that subjects differ in the way they judge the dissimilarity between stimuli (cf. Bijmolt et al. 1998; Johnson and Fornell 1987; Johnson et al. 1992). Heterogeneity of subjects is not accommodated in the hybrid models of Degerman (1970) and Carroll and Pruzansky (Carroll 1976; Carroll and Pruzansky 1980). Carroll and Chaturvedi (1995), however, proposed a hybrid model that accommodates individual differences. The model, labelled CANDCLUS, combines the tree structure model INDCLUS (Carroll and Arabie 1983; Chaturvedi and Carroll 1994) with the spatial model INDSCAL (Carroll and Chang 1970). In the CANDCLUS model heterogeneity is accounted for by estimating subject-specific weights for the discrete and continuous attributes. However, this substantially increases the number of parameters to be estimated, especially if the number of subjects is large, as is often the case in empirical applications. An additional limitation of the CANDCLUS model, as well as of the other hybrid model described above is that they are deterministic. Whereas stochastic models postulate a probabilistic data generation mechanism that describes the uncertainties in the outcomes of the underlying process, allow for parametric statistical inference, and enable generalizations from the sample to the population, deterministic approaches do not allow for such inferences and describe only the particular data set at hand. However, in spirit our approach is in line with that of Carroll and Chaturvedi (1995).

3. Mixture of tree and spatial representations

3.1. The model for dissimilarity judgments

Let n=1,...,N denote subjects, i, j, k = 1,...,I denote stimuli, and s = 1, 2 denote S=2 classes of judgment processes. In particular, we assume s = 1 to represent a judgment process based on common discrete features represented by an ultrametric tree, and s = 2 a judgment process based on continuous dimensions represented by a spatial MDS model. The data, d_{ijn}, are the observed dissimilarities of stimuli i and j by subjects n. Here we deal with the stochastic nature of the respondents' decision process, by formulating a model that consists of a representation component and a measurement component; the latter making distributional assumptions on the error. The representation component of the judgment process pertains to tree respectively spatial representations, assumed to capture subjects dissimilarity judgments of stimuli. Making distributional assumptions enables us to adopt maximum likelihood (ML) estimation. Under certain regularity conditions, ML estimates for mixture models have important properties such as consistency of the estimates, not shared by models that include individual-specific parameters (cf. Amemiya 1985, p. 115, 123).

We assume S = 2 unobserved classes, with for s = 1 π_{Tree} and for s = 2 π_{MDS} denoting the prior probabilities of the tree and the MDS representations, respectively. We assume that a particular subject in the sample, when making a dissimilarity judgment, draws from each of these two processes with prior probabilities π_{Tree} and π_{MDS}, respectively. Given that the process of class s is used, we assume the P = I(I-1)/2 dissimilarity judgments for subject n to follow a log-normal distribution. The log-normal distribution is well suited to describe dissimilarity judgments, since its support is restricted to the positive domain and it accounts for the skewness of the judgments. Thus we have:

$$\phi_s(d_{ijn} \mid \delta_{ijs}, \sigma_s) = \frac{1}{d_{ijn}\sqrt{(2\pi\sigma_s^2)}} \exp\left[\frac{(\ln(d_{ijn}) - \delta_{ijs})^2}{2\sigma_s^2}\right] \quad . \tag{1}$$

Here δ_{ijs} is the expected value of d_{ijn} given class s, and σ_s^2 its variance.

For class s = 1, it is assumed that the dissimilarity judgments are derived from an ultrametric tree, so that each triple of expected dissimilarities satisfies the ultrametric inequality:

$$\delta_{ij1} \leq \max(\delta_{ik1},\delta_{jk1}), \quad \forall(i,j,k) \quad . \tag{2}$$

The ultrametric inequality ensures that for any three objects, labelled i, j, and k, for which (2) holds, i and j are less distant from each other than each of them is from k. It can be shown (cf. Corter 1996) that this inequality is identical to restricting the largest two of any three distances to be equal. The set of constraints in (2) imposes the ultrametric inequality for class s = 1 only.

For class s = 2, it is assumed that the dissimilarities are produced by a T = 2 two-dimensional spatial model where the location of stimulus i on dimension t is represented by x_{it}. We restrict the spatial configuration to two dimensions for reasons of ease of interpretation and because “a tree contains about the same amount of information, generally speaking, as a two-dimensional space” (Carroll 1976, p. 453); in Ghose’s (1998) and other comparisons of trees and spaces spaces were also restricted to be two-dimensional. Thus, for class s = 2:

$$\delta_{ij2} = \mu + \sum_{t=1}^{2} (x_{it} - x_{jt})^2 \quad . \tag{3}$$

We use the node-height convention to determine the number of parameters for the tree in s = 1. The node-height convention states that there are I-1 parameters corresponding to the heights of the I-1 higher order nodes in an ultrametric tree for I stimuli in each class (cf. Corter 1996, p. 16). The effective number of parameters estimated is thus (I-1) for s = 1. The MDS solution has 2I associated parameters in T = 2 dimensions, but is invariant to centering, scaling and rotation, which subtracts T(T+1)/2 = 3 parameters, so that 2I-3 effective parameters are estimated for s = 2. In addition, there are 2 variance parameters and 1 prior probability to be estimated, which adds up to M = 3I-1 effective parameters estimated for the model as a whole. The unconditional distribution of the dissimilarity judgments is formulated as:

$$\phi(d_{ijn}) = \sum_{s=1}^{2} \pi_s \phi_s(d_{ijn} \mid \delta_{ijs}, \sigma_s) \quad . \tag{4}$$

3.2. Estimation

The likelihood:

$$L = \prod_{n=1}^{N} \prod_{i \leq j} \sum_{s=1}^{2} \pi_s \phi_s(d_{ijn} \mid \delta_{ijs}, \sigma_s) \tag{5}$$

is maximized under the constraints on the fitted distances provided by (2) and (3) using an EM algorithm (cf. Wedel and Kamakura 1998). We provide the main features. The algorithm maximizes the likelihood in a series of major EM iterations, and minor iterations within each M-step for s = 1,2. The E-step of the algorithm involves taking the expectation of the complete log-likelihood with respect to unobserved 0/1 class membership indicators, which amounts to replacing these indicators with their expected values. These expected values equal the posterior probabilities, π_{ns}, that subject n belongs to class s, calculated at the current parameter estimates by means of Bayes' Theorem, see equation (6) in the next section. Each M step for s = 1 is started using unconstrained estimation. After convergence the ultrametric constraints in (2) are approximately enforced by using the triple reduction method, which involves a repeated sequential averaging of the largest two pairs of each triple (Roux 1987). From the starting values of the distances thus obtained, a Sequential Quadratic Programming constrained estimation algorithm is applied, using the Broyden, Fletcher, Goldfarb and Shanno (Scales 1985) Quasi Newton method. We use the SQP algorithm implemented in GAUSS (Aptech 1995). In each M-step for s = 2 is initialized by a metric MDS based on a singular value decomposition of the distance matrix. For s = 2 an unconstrained maximization algorithm, using the Polak-Ribiere (Scales 1985) Conjugate Gradients method is used in each M-step. We approximate the required derivatives for both s = 1 and s = 2 numerically using forward differences. In each subsequent M step the parameter estimates from the previous steps are used as starting values, for s = 1,2. The convergence criterion used on the average log-likelihood is 10^{-6}. For further details on the EM algorithm we refer to Dempster, Laird, and Rubin (1977) or Wedel and Kamakura (1998). The EM algorithm is started from equal posterior probabilities ($\pi_{ns} = 0.5$; n = 1,...N; s = 1,2), so that each subject has an equal a-priori probability of belonging to the ultrametric tree and the spatial class.

3.3. Evaluation

Once the parameters of the model are estimated, the posterior probabilities, π_{ns}, that subject n has drawn upon process s (tree or space), can be calculated by means of Bayes' Theorem. For the tree class the posteriors equal:

$$\pi_{n,TREE} = \frac{\pi_{TREE}\,\phi_{n1}}{\pi_{TREE}\,\phi_{n1} + \pi_{MDS}\,\phi_{n2}} \quad , \tag{6}$$

and for the MDS class $\pi_{n,MDS} = 1-\pi_{n,TREE.}$ These posterior probabilities are important quantities in our study, since they enable us to assess post-hoc whether subject n has used the tree representation ($\pi_{n,MDS} = 0$, $\pi_{n,TREE} = 1$), or the spatial representation ($\pi_{n,MDS} = 1$, $\pi_{n,TREE} = 0$), or a mixture of both. The π_{ns} provide a probabilistic allocation of the objects to the ultrametric tree and spatial MDS classes, and thus enable one to judge a-posteriori which judgment strategy a particular subject employs. We investigate this using an entropy measure E_2:

$$E_2 = 1 - \sum_{s=1}^{2}\sum_{n=1}^{N} -\hat{\pi}_{ns}\ln\hat{\pi}_{ns} / N\ln(2) \quad . \tag{7}$$

The entropy measure assesses the separation of the two classes and can thus be interpreted as the extent to which subjects use a single judgement process. Values close to one indicate that subjects use a single strategy, i.e. a subjects dissimilarity judgment process can be represented by either a tree or a spatial representation. Values close to zero indicate that there is not enough information in the data to distinguish between the two processes for a particular subject, so that the available data indicate that subjects use a mix of the two strategies, i.e. for each judgment, they draw with non zero probabilities from both processes to arrive at their dissimilarity judgment.

To assess the fit of each model and to compare this across alternative model formulations, we compute AIC (Akaike 1974) and the R^2 fit measures, the latter being defined as:

$$R^2 = 1 - \frac{\sum_{s=1}^{S}\sum_{n=1}^{N}\pi_{ns}\sum_{i,j}(\ln(d_{ijn}) - \delta_{ijs})^2}{\sum_{n=1}^{N}\sum_{i,j}(\ln(d_{ijn}) - \overline{\ln(d_{ijn})})^2} \quad , \tag{8}$$

where $\overline{\ln(d_{ijn})}$ equals the average dissimilarity judgment across all subjects and pairs of stimuli.

The estimated prior and posterior probabilities, the entropy, the R^2 fit measure, and AIC are the statistics by which we evaluate the empirical results to draw generalizeable conclusions on the use of discrete versus continuous dimensions in the dissimilarity judgment of stimuli.

3.4. Analysis of synthetic data sets

In order to check the performance of the algorithm, we generated three synthetic data sets with S = 2 classes, I = 5 stimuli and N = 20 subjects. The first data set, A, is generated on the basis of one single ultrametric tree for all subjects, where the distances conform to (2). The distances, satisfying the ultrametric inequality, were taken from subsets of the stimuli in the Table 5.3 in DeSarbo, Manrai, and Manrai (1993). Random error drawn from N(0, 0.5) was added to these true distances. This data set was analysed with the above mixture of tree and MDS configurations. The estimation procedure assigned all subjects correctly to class 1, the tree structure, with a posterior probability of 1.0000. The second class, the MDS class, was empty, all posteriors equalling 0.0000. Consequently $\pi_{TREE} = 1.0$ $\pi_{MDS} = 0.0$, and $E_2 = 1.0$, indicating all subjects using the tree representation.

The second data set, B, was generated on the basis of a single MDS model, where the distances conform to (3). The stimulus coordinates were drawn from a N(0,2) distribution, and μ=20 was used. Random error drawn from N(0, 0.5) was added to these distances computed on the basis of these parameter values for all subjects. Data set B was analysed with the mixture of tree and MDS model. The algorithm correctly assigned all subjects to class 2, the MDS class, with a posterior probability of 1.0000. Class 1, with the tree structure , was empty, all posteriors equalling 0.0000. Consequently $\pi_{TREE} = 0.0$ $\pi_{MDS} = 1.0$, and E_2=1.0, indicating that all subjects use the spatial configuration.

The third data set, C, was generated on the basis of a tree model for class 1, and an MDS model for class 2. Each of the two classes comprised 10 subjects. The distances for the two classes were generated as for data sets A, respectively B, above. Data set C was analysed with the mixture model. The EM estimation algorithm assigned subjects 1 through 10 to class 1, with the tree structure, with a posterior probability of 1.0000, and subjects 11 through 20 to Class 2, with the spatial structure, all posteriors equalling 1.0000. Consequently $\pi_{TREE} = 0.5$ $\pi_{MDS} = 0.5$, while again E_2=1.0, indicating that all subjects use a single strategy, but that half of the subjects fit the spatial, and the other half the tree structure.

Thus, from these analyses of synthetic data, it appears that the tree model is capable of identifying the true decision process from the data, even if the number stimuli (I=5) is relatively small, which theoretically leads to a weak posterior update in the E-step of the algorithm. Both a pure tree structure, a pure spatial structure and a mixed structure were correctly identified, while the posterior probabilities and the entropy statistic indicate that the classification of subjects into both processes is quite good.

4. An illustrative application

To provide an example of alternative representations of dissimilarity judgments, we analyse the data published by Schiffman, Reynolds, and Young (1981, pp. 33-34). In a sensory experiment, 10 subjects (nonsmokers, aged 18-21 years) tasted ten different brands of cola: Diet Pepsi, Royal CLub Cola, Yukon, Dr. Pepper, Shasta, Coca Cola, Diet Dr. Pepper, Tab, Pepsi Cola, and Diet Rite. Each subject provided 45 dissimilarity judgments by means of paired comparisons on a graphical anchored line-scale. The judgments were transcribed on a scale from 0-100 representing same (near 0), and different (near 100). In addition, ratings on thirteen taste attributes, e.g. bitterness, sweetness, and fruitiness, were collected from the same subjects.

To assess what structure best represents the dissimilarity judgments, we estimate the following five models:

1. An S = 1 ultrametric tree, TREE(1), i.e. the model provided by equations (1) and (2) with S=1.This model corresponds to traditional ultrametric tree models.

2. An S = 2 ultrametric tree, TREE(2), i.e. a model provided by equations (1) and (2) with S=2. This model accounts for heterogeneity in the tree structure across unobserved classes. The ultrametric restrictions in classes 1 and 2 may differ, so that this model simultaneously identifies latent classes of subjects, as well as an ultrametric tree-topology for each class (Wedel and DeSarbo 1998).

3. An S = 1 and T = 2 MDS model: MDS (1), i.e. a model provided by equations (1) and (3) with S = 1 class only and T = 2 latent dimensions. This model corresponds to a traditional MDS model for paired comparison data.

4. An S = 2 MDS model: MDS(2). i.e. a model provided by equations (1) and (3) with S = 2 and T = 2. This model accounts for heterogeneity in the spatial representation of the stimuli across unobserved classes. The positions of the stimuli in the two dimensional spaces in classes 1 and 2 may differ, reflecting different perceptual orientations. This model simultaneously identifies latent classes of subjects, as well as an spatial MDS structure for each class.This model itself seems not to have been published previously.

5. The S = 2 mixture of ultrametric tree and spatial T = 2 MDS model, that identifies latent classes of subjects that potentially differ in the type of representation of the stimuli as described in the methods section.

We estimate models 2 and 4, with two trees and two spaces respectively, in order to inspect whether the mixture of tree and space model identifies classes that actually differ in the representational structure underlying the dissimilarity judgments, or that parametric heterogeneity would have been sufficient.

The ultrametric tree structure for the ten cola brands is presented in Figure 1. One branch of the tree contains three diet colas, which have relatively low distances (the distance to their least common ancestor node): Diet Pepsi, Diet Rite and Tab. Under the feature matching model, the path-length from the root of the tree to the least common ancestor node of these three diet cola's is a measure of the importance of the features shared by these stimuli. The interpretation of the common ancestor node of these three colas as diet/regular feature is hampered somewhat by the fact that the fourth diet cola in the stimulus set, Diet Dr. Pepper is joined to the regular Dr. Pepper, albeit with a relatively large distance. The common ancestor node of these two brands seems to be a brand taste feature: Dr Pepper versus other brands, which can be interpreted as the presence or absence of the characteristic cherry flavour. The sub-tree in the middle of the ultrametric tree in Figure 1 shows a set of nodes that can be interpreted as representing brand-specific features, distinguishing the five remaining non diet brands.

Figure 2 presents the spatial representation of the ten cola brands. We try to fit the thirteen taste attributes into this configuration in order to label the dimensions. However, only two attributes, namely fruitiness and fresh versus stale, reach a satisfactory fit (rho above 0.80), whereas the other attributes are clearly not useful for interpreting the dimensions (rho below 0.60). Colas in the upper part of Figure 2 can be interpreted as more fruity and fresh, whereas colas in the lower part are less fruity and more stale. Hence, the vertical dimension separates the Dr. Pepper brands at the top from the other brands, with Yukon taking an intermediate position, which is apparently due to the specific cherry-taste of the Dr. Pepper brand. None of the thirteen taste attributes correlates high with the horizontal dimension. It seems to separate the diet versus the regular brands: on the left hand side of the horizontal axis one observes Diet Dr. Pepper, Diet Pepsi, Tab, and Diet Rite, on the right hand side the regular brands Yukon, RC Cola, Pepsi, Shasta and Coca Cola. Dr. Pepper seems to take a somewhat intermediate position. Among the regular cola brands on the right hand side of the plot there seems to be fairly little distinction, among the diet colas Diet Dr. Pepper stands out. To summarize, the two dimensions in the plot can be interpreted as a diet versus regular and a cherry taste dimension, respectively.

Comparing the tree and the space (Figures 1 and 2), one has to conclude that the interpretation of both representations is very similar. In both representations the same three clusters of brands emerge: Dr. Pepper and Diet Dr. Pepper; Diet Pepsi, Diet Rite, and Tab; and Yukon, RC Cola, Pepsi, Shasta, and Coca Cola. Both the tree and the space derive the taste distinction between diet and regular colas, and between brands with and without cherry flavour. These attributes are best interpreted as discrete features. No clear continuous dimensions are found in the spatial representation. Since, features correspond to the behavioral model underlying tree structures, the tree representation seems to derive a clearer structure among the cola brands.

[INSERT FIGURES 1 AND 2 HERE]

To illustrate the insights provided by our mixed model of a tree and a space, we provide the results of an analysis of the Schiffman, Reynolds, and Young (1981) cola data with our procedure in Figure 3. The mixture model results show that there are two well separated classes (all posterior probabilities of membership are very close to either zero or one), each comprising of 5 subjects. The subjects in the tree-class seem to identify specific brand tastes: Dr. Pepper and Diet Dr Pepper, respectively Coca Cola and its diet version Tab are in separate sub-trees. Note that particularly the Dr. Pepper brands stand out as indicated by the length of the path from the root to the node. However, the exception is that Diet Pepsi and Regular Pepsi are not joined in a specific sub-tree, which shows Pepsi has not been able to let its diet version taste similar to its regular version. Nevertheless, we conclude that the specific brand tastes are the dominant features determining dissimilarity judgments in this class. In the space-class, the vertical dimension still separates out the diet and the regular colas, although the positions of the brands on this dimension are more continuously dispersed than in the aggregate space (Figure 2), with Diet Pepsi clearly having the weakest diet flavour, and Diet Rite and Tab the strongest. However, in this case the horizontal dimension is very clearly a continuous dimension on which the brands are well dispersed. This continuous dimension underlies both the diet and the regular versions of the brands, the Dr. Pepper brands being on the one extreme and Coca Cola brands on the other extreme of the dimension. Again, we fit the thirteen attributes into the configuration: now 7 attributes reach a satisfactory fit (rho above 0.80); all of them having a high correlation with the horizontal dimension.This dimension reflects sweetness and bitterness, with the brands on the left hand side being more sweet and less bitter, sour, and chemical. It is interesting to note that the subjects in the MDS class all have the ability to taste PTC (a chemical compound that tastes bitter), while subjects in the tree class do not have this ability. The posterior memberships exactly correspond with this trait, which is determined by one allelle on the human genome. Thus, the perceptual process being based on continuous dimensions, such as sweetness and bitterness, seems to be determined here by one single genetic factor.

[INSERT FIGURE 3 HERE]

Comparing the mixed tree-space solution in Figure 3 with the aggregate tree and space in Figures 1 and 2, it is obvious that the sample is heterogeneous with respect to the representation underlying the dissimilarity judgments. The aggregate level analyses mask important characteristics that are recovered by the tree and spatial structures identified at the latent class level. In the mixed model, the continuous dimension sweetness is found, which is not recoved in the separate tree and space analyses, for example. Furthermore, the fit of the tree-space mixture (R^2=44.8%) is much better than that of either the aggregate space (R^2=24.4%) or aggregate tree (R^2=24.5%) solutions presented above for the cola data. In addition, the mixed tree and space model fit better than either a mixture of two trees (R^2 = 40.8%), or a mixture of two spaces (R^2 = 40.9%). The empirical application demonstrates the insights obtained with our procedure and that incorrect and incomplete conclusions may be drawn from aggregate level solutions if the true underlying perceptual structure is heterogeneous.

5. Analysis of twenty empirical data sets

Following Pruzansky, Tversky and Carroll (1982), Johnson et al. (1992), and Ghose (1998), we analyse multiple paired comparisons data sets. We restrict the analysis to data sets that pertain to this type of dissimilarity judgments, and we do not consider for example derived dissimilarity data (e.g. computed from attribute ratings), brand switching data, co-occurence data, or any dissimilarity judgments other than paired comparisons (e.g. traidic combinations or free sorting). The reason is that we are interested in deriving the processes underlying paired comparisons, and in particular in identifying individual differences in those processes. The analyses enable us to draw conclusions on the type of representation model that best describes the decision process underlying dissimilarity judgments, and whether or not this depends on factors related to the stimuli, the subjects, and the measurements.

Table 1 lists the twenty data sets and their characteristics: the type of stimuli, the number of stimuli, the type of subjects, the number of subjects, and the number of points of the dissimilarity rating scale. The stimuli in most applications are commercial stimuli, in particular brands of fast moving consumer goods (fmcg), durables, services, and media. Such stimuli are usually referred to as conceptual (e.g. Johnson and Fornell 1987; Johnson et al. 1992; Pruzansky, Tversky, and Carroll 1982). In addition, we analyse several data sets on non-commercial stimuli, namely locations and emotions. The number of stimuli in the data sets ranged from 8 to 15, the number of subjects ranged from 10 to 60. Dissimilarity judgments were provided on 7 to 11 point scales. The two data sets on emotions have been published previously, the other data sets are primary data.

[INSERT TABLE 1 ABOUT HERE]

The dissimilarity data were standardized by subject before the analyses, to prevent the solutions becoming confounded with the effects of response strategies (cf. Bijmolt et al. 1998). Most data sets do not contain missing values, a few data sets have a small percentage of missings. Before standardisation, the missing values were imputed by mean substitution for each individual. As in the application to the cola taste data, for each of the twenty data sets we estimate five models, namely a single tree model, a two trees model, a single space model, a two spaces model, and the mixed tree and space model, in order to examine which structure best represents the dissimilarity judgments.

In line with the previous literature in this area (Johnson et al. 1992, Pruzansky, Tversky and Carroll 1982, Ghose 1998), we report the percentages of variance explained, R^2, and AIC for each of those five models. Tables 2 and 3 present the results, where the model that explains the largest percentage of variance, respectively has the lowest AIC, is indicated in bold face type for each data set.

[INSERT TABLES 2 AND 3 ABOUT HERE]

Model selection based on the highest R^2 fit statistic and the lowest AIC (in Table 2 and

3 respectively) clearly shows that heterogeneity between subjects exists. For each of the twenty data sets, using either of both selection criteria, a model with two latent classes is identified as most appropriate. Each of the three mixture models, that is the model with two trees, the model with two spaces and the model with a tree and a space, is selected about equally often.The pattern of which models are identified is highly similar for AIC and the R^2 statistic. The AIC favors the model with two trees somewhat more often and the two spaces model somewhat less often as compared to the R^2 statistic. This finding can be explained by the fact that AIC corrects for the number of parameters estimated, whereas the R^2 statistics does not, while a tree is somewhat more parsimonious than a two-dimensional space. We examined whether the type of stimuli, the number of stimuli, the type of subjects, the number of subjects, and the number of scale points affect the model indicated as most appropriate by AIC or R^2. F-tests and Chi-square tests did not show any significant effect (all p-values > 0.05). Hence, we conclude that the relative fit of these models is not related to the factors mentioned above. However, note that in a number of cases our mixed model reaches a substantially higher fit as compared to the four alternative models. This supports the need for the possibility to model structural heterogeneity in stimulus representation.

Next, we more closely examine the results of the mixture of space and tree model for the twenty data sets. The proportions of the tree and space classes, the entropy statistic that indicates the separation of the classes, and the R^2 of the solution are reported in Table 4. In addition, we compute the average skewness of log-dissimilarities for the tree and the MDS class, since that statistic seems to be the most important data indicator of the appropriateness of a tree or space (Ghose 1998).

[INSERT TABLE 4 AND 5 ABOUT HERE]

The estimates of the prior probabilities show that across all twenty data sets, the tree structure and the spatial configuration are about equally important. For eight data sets the tree structure is more important (i.e. has a larger π_{TREE}), whereas for twelve data sets the spatial configuration is more important (i.e. has a larger π_{MDS}). The average proportions, 0.482 and 0.518 for the tree and the space respectively, do not differ significantly (t=-0.45; df=19; p=0.66). For most sets of stimuli, the size of both components is substantial. The contribution of the tree structure (spatial configuration) ranges from 0.8176 (0.1824) for restaurants to 0.1500 (0.8500) for cities.

In eleven of the twenty applications, the classes are very well separated ($E_2 > 0.900$), indicating that nearly all subjects in these data sets either have the tree structure or the spatial configuration. In particular, looking at the entropy and the prior probabilities, restaurants and bars seem to be almost entirely judged on the basis of features, while cities appear to be perceived predominantly in terms of continuous dimensions. For several sets of stimuli, most notably recreation facilities and women's magazines, the entropy measure is medium (around 0.7), while the prior probabilities for the tree and spatial configurations are around 0.5, which indicates that individual subjects tend to perceive these stimuli in terms of both discrete features and continuous dimensions.

The explained variance R^2, as defined in equation (8), varies substantially across the data sets, with a minimum of 0.0409 for cars and a maximum of 0.5725 for TV channels. A number of R^2 values are rather low, which is to (a) our standardization and log-transformation applied before fitting the data, (b) the fact that a large number of observations, NI(I-1)/2, is represented by a relatively small number of parameters, 3I-1. However, the relative values of R^2 are very well interpretable and parallel those of AIC (Tables 2 and 3).

We performed three analyses of covariance to examine whether the proportion of tree versus space latent classes, the entropy measure, and the R^2 fit statistic (see Table 4) are affected by the type of stimuli (commercial versus non-commercial stimuli), the number of stimuli, the type of subjects (students versus consumers/managers), the number of subjects, and the number of scale points (see Table 1). The type of stimuli turns out to have a significant effect on the proportion of the tree class versus the space latent class (F=4.41; df=1; p=0.05). The average proportion of the tree class is higher for the sets of commercial/conceptual stimuli (0.531) than for the non-commercial stimuli (0.334). This result is in line with results of previous studies, which reported tree structure models to outperform multidimensional scaling methods in fitting conceptual stimuli such as brands and the opposite for perceptual stimuli (Pruzansky, Tversky, and Carroll 1982) and more abstract stimuli such as product categories (Johnson et al. 1992). Note, however, that previous studies draw these conclusions on the basis of aggregate analyses, that did not account for heterogeneity among subjects. We did not find a significant effect (all p-values > 0.05) of the type of stimuli on the entropy and the explained variance measures. Furthermore, there seems to be no relationship (again all p-values > 0.05) between on the one hand the proportion of the tree structure component versus that of the MDS component, the entropy measure, and the R^2 fit measure and on the other hand the number of stimuli, the type of subjects, the number of subjects, and the number of values of the rating scale. Hence, with respect to these three criteria, the outcomes of the mixed tree and space model is rather robust against factors in the study design. In the majority of the applications, the results of our model enabled an fairly strict classification of subjects into the latent classes, while both classes are substantial in size. Thus, differences between subjects in the representation of the stimuli and the judgement process may be more important than differences caused by design factors such as the number of stimuli and the number of points of the rating scale.

Finally, the skewness of the dissimilarity data is lower in the tree class than in the space latent class for forteen of the twenty data sets (last two columns of Table 4). On average, the skewness differs significantly between the two classes (t=-2.17; df=19; p=0.04), with averages of -0.96 and -0.80 for the tree classes and the space classes, respectively. This finding corresponds to findings of previous studies, e.g. Ghose (1998), which have show that skewness discriminates between trees and spaces, the former having a more negative skewness.

6. Conclusion and discussion

We proposed a mixture model of a tree structure and a spatial configuration for the analysis of dissimilarity judgments. The mixture model accounts for heterogeneity between subjects in the extent to which they use a feature-based or a dimension-based representation of stimuli through a mixture model specification, where the dissimilarity judgments of one class are modelled as distances in an ultrametric tree and the dissimilarity judgments of the other class are modelled as distances in a Euclidean space. Thereby, the model accommodates structural heterogeneity among classes, which is importantly different from parametric heterogeniety accommodated in previous mixture models. Through the analysis of synthetic data sets, we showed that the model adequately recovers known tree and space structures that underly dissimilarity data. The results of the mixture model were illustrated in an application to previously published data from a sensory experiment with colas. Analysing this data set by means of a tree model and a space model separately, yielded highly similar representations. Both representations appeared to be dominated by discrete features, whereas one might expect more continuous dimensions to show up in the space model. Analysis with the mixed tree and space model yielded a much richer structure revealing amongst others clear continuous dimensions. Hence, the empirical application of our model to the cola taste data demonstrated that it yields well-interpretable, useful solutions, whereas pure tree or space models, ignoring structural heterogeneity, may lead to erroneous conclusions.

In the application of the mixed model to twenty empirical data sets, we found heterogeneity across subjects for each data set, since the two-class models outperformed the single class models. Each of the three mixture models, that a two-trees model, a two-spaces model, and a model with one tree and one space, was identified as best for a number of data sets. Which of these three models is most appropriate did not seem to be related to design factors such as the type and number of subjects, the type and number of stimuli, and the number of scale values. However, due to the fact that "only" twenty data sets were analyzed, the power of the tests may have been only moderate.

When examining the mixed tree and space model as proposed in this paper, in general, the two latent classes turned out to be both substantial and separated rather well. Hence, there are substantial and clear differences between individual subjects with respect to whether a feature-based or a dimension-based representation fits there dissimilarity judgments better. Hybrid models that do not deal with heterogeneity among subjects with respect to the representation of stimuli and the decision process may lead to erroneous results. Furthermore, the fact that substantial individual differences may exist casts some doubt about the conclusions drawn in previous studies examining the choice between tree and space models while not accounting for such heterogeneity.

There turned out to be little to no relationship between on the one hand the relative importance of the tree versus the space, the extent to which subjects use a single or a mixed judgment strategy, and the total fit of the model, and on the other hand study design factors like the number of stimuli, the type and number of subjects, and the number of points of the rating scale. In addition the type of stimuli did not affect the class separation and the fit either. These findings are reassuring, since in applying our model one may assume that the outcomes are relatively robust against the study design. However, we found that the importance tree structure relative to the spatial configuration was significantly higher for commercial, conceptual stimuli as compared to the non-commercial stimuli (locations and emotions). This result is in line with previous studies, which also showed that tree structure models outperform multidimensional scaling methods in fitting dissimilarity judgments between conceptual stimuli such as brands (Johnson and Fornell 1987; Johnson and Hudson 1996; Johnson et al. 1992; Pruzansky, Tversky, and Carroll 1982). However, those studies examine the fit at the aggregate level, while we take individual differences into account.

Further research is needed in a number of directions. First, whereas previous studies have primarily focussed on characteristics of the stimuli and the task as causes of differential ability of trees versus spaces to represent dissimilarity data, this study revealed that individual differences may be much more important in that respect. Therefore, future research should address subject-related factors, such as cognitive complexity (Bieri 1955) and style of processing (Childers, Houston, and Heckler 1985), as possible drivers of the adequacy of tree and spatial structures to fit dissimilarity judgments at the individual- or class-level. Second, research into the psychological processes underlying dissimilarity judgments is needed. As demonstrated by Glazer and Nakamoto (1991), an observed pattern of dissimilarity judgments may not always accurately reveal what is the correct model from a cognitive perspective. They show that the relative fit of alternative tree structures (ultrametric and additive trees) and spatial configurations (Euclidean and city-block distances) is occasionally not very strong related to the true psychological processes that underlies the data. Hence, care should be taken in considering the results of tree structure models, MDS models, or mixed models as evidence of the true underlying psychological process. If the main interest is to reveal the true processes underlying dissimilarity judgments, alternative approaches should be used in conjunction with statistical modelling procedures as the one described in this paper. One could examine the underlying psychological processes and the judgment task through studies in the line with for example Bijmolt et al. (1998) using a process-tracing perspective, that is through the analysis of verbal protocols of dissimilarity judgments. Such studies may in particular focus on the nature of the attributes used by respondents while comparing stimuli as well as on the characteristics of the respondents, the stimuli, and the judgment task, that affect subjects' perceptual representation of stimuli.

References

AKAIKE, H. (1974), "A New Look at Statistical Model Identification," *IEEE Transactions on Automatic Control, AC-19*, 716-723.

AMEMIYA, T. (1985), *Advanced Econometrics*, Cambridge, MA: Harvard University Press.

APTECH (1995), *Constrained Maximum Likelihood, GAUSS manual*,Maple Valley: Aptech systems.

BIERI, J. (1955), "Cognitive Complexity-Simplicity and Predictive Behavior," *Journal of Abnormal and Social Psychology*, *51*, 263-268.

BIJMOLT, T.H.A., WEDEL, M., PIETERS, R.G.M., and DeSarbo, W.S. (1998), "Judgments of Brand Similarity," *International Journal of Research in Marketing*, *15*, 249-268.

CARROLL, J.D. (1976), "Spatial, Non-Spatial and Hybrid Models for Scaling," *Psychometrika, 41*, 439-463.

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

CARROLL, J.D., and ARABIE, P. (1996), "Multidimensional Scaling", in *Handbook of Perception and Cognition*, *Volume 3: Measurement*, *Judgment and Decision Making*, M.H. Birnbaum (ed.), San Diego: 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.

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

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. (1980), "Discrete and Hybrid Scaling Models," in *Similarity and Choice,* E.D. Lanterman and H. Feger (eds), Bern: Hans Huber, 108-139.

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

CHILDERS, T.L., HOUSTON M.J., and HECKLER S. (1985), "Measurement of Individual Differences in Visual Versus Verbal Information Processing," *Journal of Consumer Research, 12*, 125-134.

CORTER, J.E. (1996), *Tree Models of Similarity and Association*, London:Sage.

DEGERMAN, R. (1970). "Multidimensional Analysis of Complex Structure Mixtures of Class and Quantitative Variation," *Psychometrika, 35*, 475-491.

DEMPSTER, A. P., LAIRD, N. M., and RUBIN, R. B. (1977), "Maximum Likelihood From Incomplete Data Via the EM-algorithm," *Journal of the Royal Statistical Society, B39*, 1-38.

DESARBO, W.S., MANRAI, A.K., and MANRAI, L.A. (1993), "Non-Spatial Tree Models for the Assessment of Competitive Market Structure: An Integrated Review of the Marketing and Psychometric Literature," in: *Marketing, Handbooks of OR&MS vol. 5*, J. Eliashberg and G.L. Lilien (eds), 193-257.

EAGLY, A.H., and CHAIKEN, S. (1993), *The Psychology of Attitudes*, Fort Worth: HBJ Publishers.

GARNER, W.R. (1978), "Aspects of a Stimulus: Features, Dimensions, and Configurations," in: *Cognition and Categorization*, E. Rosch and B.B. Lloyd (eds.), Hillsdale:Lawrence Erlbaum, 99-

133.

GHOSE, S. (1998), “Distance Representations of Consumer Perceptions: Evaluating Appropriateness by Using Diagnostics,” *Journal of Marketing Research*, *35*, 137-153.

GLAZER, R., and NAKAMOTO, K. (1991), “Cognitive Geometry: An Analysis of Structure Underlying Representations of Similarity,” *Marketing Science, 10*, 205-228.

GREEN, P.E., CARMONE, F.J., and SMITH, S.M. (1989), *Multidimensional Scaling: Concepts and Applications*, Boston: Allyn and Bacon.

JOHNSON, M.D., and FORNELL, C. (1987), “The Nature and Methodological Implications of the Cognitive Representation of Products,” *Journal of Consumer Research, 14,* 214-228.

JOHNSON, M.D., and HORNE, D.A. (1992), “An Examination of the Validity of Direct Product Perceptions,” *Psychology & Marketing, 9*, 221-235.

JOHNSON, M.D., and HUDSON, E.J. (1996), “On the Perceived Usefulness of Scaling Techniques in Market Analysis,” *Psychology & Marketing*, *13*, 653-675.

JOHNSON, M.D., LEHMANN, D.R., FORNELL, C., and HORNE, D.R. (1992), “Attribute Abstraction, Feature-Dimensionality, and the Scaling of Product Similarities,” *International Journal of Research in Marketing*, *9*, 131-147.

PRUZANSKY, S., TVERSKY, A., and CARROLL, J.D. (1982), “Spatial Versus Tree Representations of Proximity Data,” *Psychometrika, 47*, 3-24.

RICHARDSON, A. (1977) “Verbalizer-Visualizer: A Cognitive Style Dimension,” *Journal of Mental Imagery, 1*, 109-126.

ROUX, M. (1987), “Techniques of Approximation for Building Two Tree Structures,” in: *Proc. Franco-Japanese Scientific Seminar: Recent Developments in Clustering and Data-Analysis,* Tokyo, 127-146.

SATTATH, S., and TVERSKY, A. (1977), “Additive Similarity Trees,” *Psychometrika*, *42*, 319-

345.

SCALES, L.E. (1985), *Introduction to Non-Linear Optimization.* London: MacMillan.

SCHIFFMAN, S.S., REYNOLDS, M.L., and YOUNG, F.W. (1981), *Introduction to Multidimensional Scaling: Theory, Methods, and Applications*, New York: Academic Press.

SCOTT, W.A., OSGOOD, D.W., and PETERSON, C. (1979), *Cognitive Structure: Theory and Measurement of Individual Differences*, New York: Wiley.

SHEPARD, R.N. (1980), "Multidimensional Scaling, Tree-Fitting, and Clustering," *Science*, *210*, 390-398.

TVERSKY, A. (1977), "Features of Similarity," *Psychological Review, 84,* 327-352.

TVERSKY, A., and GATI, I. (1978), "Studies of Similarity". in: *Cognition and Categorization*, E. Rosch and B.B. Lloyd (eds.), Hillsdale: Lawrence Erlbaum, 79-98.

WEDEL, M., and DESARBO, W.S. (1998), "Mixtures of (Constrained) Ultrametric Trees," *Psychometrika*, forthcoming.

WEDEL, M., and KAMAKURA, W.A. (1998), *Market Segmentation: Conceptual and Methodological Foundations*, Dordrecht: Kluwer.

Table 1: Characteristics of the Twenty Data Sets

Stimuli	Type of Stimuli	Number of Stimuli	Type of Subjects	Number of Subjects	Scale Points
Soft drinks	FMCG[1]	12	Students	60	7
Candy bars	FMCG	12	Students	50	9
Shampoos	FMCG	10	Consumers	47	7
Beer	FMCG	9	Students	20	7
Cars	Durables	12	Consumers	48	9
Audio	Durables	9	Consumers	20	11
Supermarkets	Services	12	Students	50	9
Recreation facilities	Services	12	Students	50	9
Banks	Services	12	Students	50	9
Restaurants	Services	10	Consumers	32	9
Supporting facilities	Services	10	Managers	15	7
Bars	Services	9	Students	20	11
Weekly magazines	Media	12	Students	60	7
Women’s magazines	Media	10	Consumers	40	9
TV-Stations	Media	9	Consumers	20	11
Countries	Locations	15	Students	14	9
Capitals	Locations	9	Students	13	11
Cities	Locations	9	Consumers	20	7
Emotions-1	Emotion	14	Students	15	9
Emotions-2	Emotion	8	Students	14	9

[1] FMCG = fast moving consumer goods

Table 2: R^2 Fit Statistics for the Twenty Data Sets

Stimuli	TREE(1)	TREE(2)	MDS(1)	MDS(2)	MIX
Soft drinks	0.2573	**0.3031**	0.2265	0.2641	0.2776
Candy bars	0.2349	**0.2922**	0.1873	0.2307	0.2742
Shampoos	0.0031	0.0319	0.0046	0.0529	**0.0541**
Beer	0.0786	0.1902	0.1110	**0.1930**	0.1681
Cars	0.0068	0.0402	0.0089	0.0377	**0.0409**
Audio	0.0610	0.1446	0.0932	**0.1896**	0.1704
Supermarkets	0.1241	0.1783	0.1401	**0.1931**	0.1927
Recreation facilities	0.0488	0.1247	0.0612	**0.1320**	0.0962
Banks	0.1408	0.2563	0.1265	0.2111	**0.2575**
Restaurants	0.0233	0.1242	0.0304	**0.1590**	0.1343
Supporting facilities	0.1985	0.2566	0.2588	0.2993	**0.3195**
Bars	0.1912	0.2681	0.1739	0.2754	**0.2932**
Weekly magazines	0.1387	0.2804	0.1387	0.2210	**0.2820**
Women's magazines	0.0037	0.0307	0.0053	**0.0512**	0.0454
TV-Stations	0.5291	**0.5956**	0.4360	0.4965	0.5725
Countries	0.0236	0.0642	0.0159	**0.0788**	0.0741
Capitals	0.1185	**0.2716**	0.1180	0.2271	0.2670
Cities	0.2872	0.3556	0.3303	**0.4181**	0.3918
Emotions-1	0.4677	0.4919	0.4741	0.4922	**0.5298**
Emotions-2	0.2110	0.3153	0.2824	**0.3511**	0.3392

[1] MDS(S): S-class MDS solution; TREE(S): S-class Tree solution; MIX: 2-class mixed Tree-MDS solution

Table 3: AIC Statistics for the Twenty Data Sets

Stimuli	TREE(1)	TREE(2)	MDS(1)	MDS(2)	MIX
Soft drinks	10011.52	**9762.40**	10182.64	10007.98	9942.48
Candy bars	8443.14	**8196.28**	8652.27	8484.36	8284.94
Shampoos	5958.03	5906.15	5962.72	5868.41	**5863.09**
Beers	1974.04	**1890.21**	1955.26	1901.60	1915.95
Cars	8932.47	**8836.94**	8935.83	8865.40	8845.56
Audio	1987.64	1923.18	1969.50	**1904.56**	1914.31
Supermarkets	8889.39	8685.64	8838.66	8650.01	**8640.42**
Recreation facilities	9161.46	8900.56	9128.35	**8841.98**	9034.39
Banks	8825.83	**8304.11**	8890.14	8556.79	8324.79
Restaurants	3910.83	3713.10	3908.29	**3693.24**	3715.63
Supporting facilities	1761.08	1721.34	1716.24	1680.72	**1663.11**
Bars	1880.18	1813.76	1902.43	1823.53	**1797.89**
Weekly magazines	5071.00	5031.20	5076.21	**5007.08**	5016.03
Women's magazines	10505.94	**9802.89**	10608.06	10179.57	9830.28
TV-Stations	1490.77	**1361.22**	1627.64	1537.74	1397.15
Countries	4149.04	**4091.09**	4150.44	4093.64	4095.62
Capitals	1264.92	**1173.11**	1272.17	1227.30	1182.40
Cities	1789.20	1715.12	1751.28	**1661.11**	1685.13
Emotions-1	3011.82	2959.54	3007.32	2984.46	**2868.50**
Emotions-2	1013.27	963.73	982.11	**955.79**	958.43

[1] MDS(S): S-class MDS solution; TREE(S): S-class Tree solution; MIX: 2-class mixed Tree-MDS solution

Table 4: Mixed Tree en Space Model Results for the Twenty Data Sets

Data set	π_{Tree}	π_{MDS}	Entropy	R^2	S_{TREE}	S_{MDS}
Soft drinks	0.6638	0.3362	0.8045	0.2776	-1.6828	-0.9689
Candy bars	0.7144	0.2856	0.9022	0.2742	-1.2334	-1.0651
Shampoos	0.2643	0.7357	0.8417	0.0541	-0.9362	-0.7243
Beer	0.4222	0.5778	0.7262	0.1618	-0.8210	-1.1412
Cars	0.4498	0.5502	0.7570	0.0409	-1.3516	-1.2533
Audio	0.3413	0.6587	0.9248	0.1704	-0.5981	-0.6624
Supermarkets	0.4174	0.5826	0.8642	0.1927	-0.3039	-0.6984
Recreation facilities	0.5175	0.4825	0.6976	0.0962	-0.8302	-0.9328
Banks	0.5378	0.4622	0.9783	0.2575	-0.9112	-0.5615
Restaurants	0.8176	0.1824	0.8689	0.1343	-0.5576	-1.0028
Supporting facilities	0.3971	0.6029	0.9649	0.3195	-1.0412	-0.5341
Bars	0.7999	0.2001	0.9985	0.2932	-0.7921	-0.6235
Weekly magazines	0.5702	0.4298	0.9824	0.2820	-0.9501	-0.5731
Women’s magazines	0.3570	0.6430	0.6861	0.0454	-0.8983	-0.6574
TV Channels	0.6985	0.3015	0.9885	0.5725	-1.3249	-1.1679
Countries	0.2951	0.7049	0.9549	0.0741	-0.7915	-0.9939
Capitals	0.4614	0.5386	0.9980	0.2670	-0.9552	-0.6123
Cities	0.1500	0.8500	0.9997	0.3918	-1.2944	-0.6914
Emotions-1	0.4021	0.5979	0.9855	0.5298	-1.1521	-0.7176
Emotions-2	0.3593	0.6407	0.8579	0.3392	-0.7655	-0.4529

FIGURE 1

Ultrametric Tree for the Schiffman et al (1981) Cola Data

FIGURE 2

T=2 dimensional Space for the Schiffman et al (1981) Cola Data

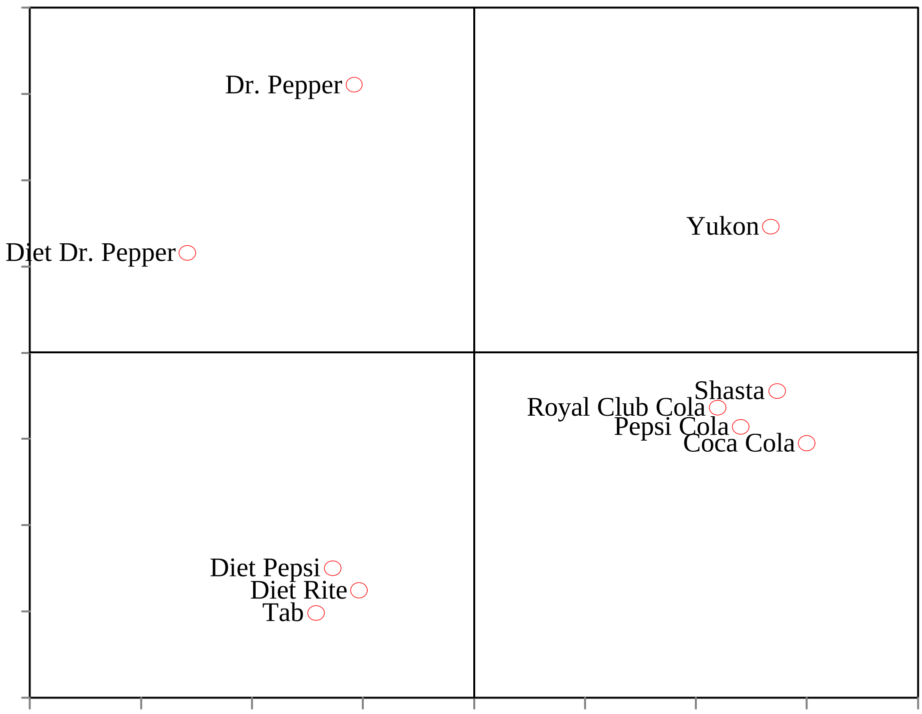
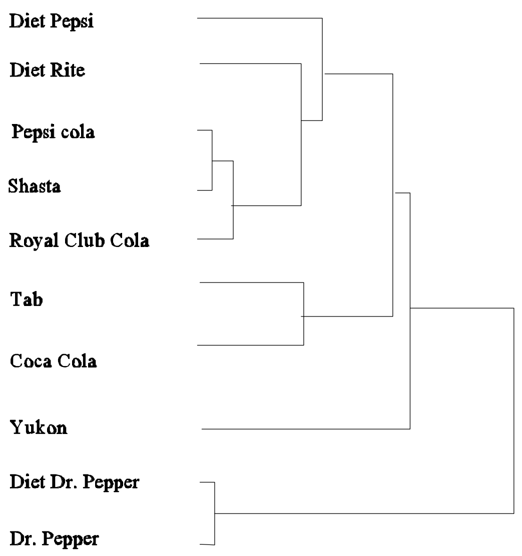


FIGURE 3

S=2 Mixed Tree- Space Solution for the Schiffman et al (1981) Cola Data

Class 1: Tree Structure



Class 2: Spatial configuration