Journal of Classification 17:243-271 (2000)
DOI: 10.1007/s003570000021

Mixed Tree and Spatial Representations of Dissimilarity Judgments

Michel Wedel

University of Groningen

Tammo H.A. Bijmolt

Tilburg University

Abstract: Whereas much previous research, focusing on the comparative fit at the aggregate level, has shown that either tree or spatial representations of dissimilarity judgments may be appropriate, we investigate whether there exist classes of subjects differing in the extent to which dissimilarity judgments are better represented by additive tree or spatial multidimensional scaling models. We develop a mixture model for the analysis of dissimilarity data that entails both representation and measurement model components. The latter involves distributional assumptions on the measurement error and enables estimation by maximum likelihood. The former component allows dissimilarity judgments to be represented either by an additive tree structure or by a spatial configuration, or a mixture of both. To investigate the appropriateness of additive tree versus spatial representations, the model is applied to twenty empirical data sets. We compare the fit of our model with that of aggregate additive tree and spatial models, as well as with mixtures of additive trees and mixtures of spatial configurations, respectively. We formulate some empirical generalizations on the relative importance of tree versus spatial structures in representing dissimilarity judgments.

Key words: Multidimensional scaling; Additive trees; Mixture models; Dissimilarity judgments

---

The authors are indebted to the editor and two anonymous reviewers for valuable comments. The computer programs and data sets used in this paper can be obtained from the authors.

Authors' Addresses: Michel Wedel, Department of Marketing and Marketing Research, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands, m.wedel@eco.rug.nl and Visiting Professor of Marketing, University of Michigan Business School, 701Tappan Street, Ann-Arbor, Michigan 48109-12341, USA; Tammo H.A. Bijmolt, Department of Marketing, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands, t.h.a.bijmolt@kub.nl

1. Introduction

Perceptions have been studied using graphical representations of dissimilarity judgments of stimuli that either take the form of trees or spatial configurations. 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 (see Barthélemy and Guénoche 1991; Carroll and Chang 1973; Corter 1996; DeSarbo, Manrai, and Manrai 1993; De Soete and Carroll 1996; Gascuel and Levy 1996; Sattath and Tversky 1977) or as a spatial configuration, respectively (see Carroll and Arabie 1998; Carroll and Green 1997; Green, Carmone, and Smith 1989). The choice between tree and spatial models is based (a) on prior theory on the attribute-types discerned by subjects for the stimuli in question, (b) on the basis of the relative fit of the two models, or (c) on such diagnostic measures as the skewness of the dissimilarity judgments (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 bringing 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 1998; Carroll and Green 1997). However, despite the added value such approaches may have over single tree structure models or spatial multidimensional scaling (MDS) models, "much has been said but little done about such mixed or hybrid models" (Carroll and Arabie 1998). In the literature, only a few hybrid models for dissimilarity judgments have been proposed. An important point to be made here is that although differences between individuals have been shown to occur both in processing attributes and in judging the dissimilarity of stimuli (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 among subjects (Carroll and Pruzansky 1980; Degerman 1970; Hubert and Arabie 1994; Hubert, Arabie, and Meulman 1997, 1998), the 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 parsimoniously by identifying two unobserved classes. The dissimilarity judgments of subjects in the first latent class are represented using an additive tree structure, and those of the subjects in the second latent class using a spatial structure. The stochastic formulation allows for maximum likelihood estimation and testing which representation is most appropriate. 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, whereas previous work has accommodated parametric heterogeneity, assuming classes to be structurally homogeneous (see Wedel and Kamakura 2000, Part 2).

In 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 in identifying whether a tree or spatial representation is more appropriate for a particular subject is demonstrated through the analysis of synthetic data sets. We illustrate our model on cola tasting data published by Schiffman, Reynolds, and Young (1981, Chapter 3). We describe the results of analysis of twenty empirical data sets to assess the relative importance of tree structures and spatial configurations across different stimulus types. We compare the model with aggregate tree and spatial models, and two-class mixtures of pure trees and pure spatial configurations, 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 exclusively compare stimuli on either discrete features or on continuous dimensions (Garner 1978; Johnson and Fornell 1987; Johnson et al. 1992; Tversky 1977; Tversky and Gati 1978). Discrete features are attributes with a limited number of values, for example diet versus regular cola. Continuous dimensions are attributes on which the stimuli vary as a matter of degree, for example the sweetness of the cola. 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 flavor, for example, could be used in the judgment process as the presence or absence of that flavor rather than as the degree of that flavor. Alternatively, a set of discrete features may be combined into a continuous dimension. Several discrete taste aspects of the cola could be integrated to form a quality 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 their experience and familiarity with the stimuli (Johnson et al. 1992). The existence of heterogeneity in subjects' perceptions of stimuli has been widely recognized, and may be related to personality factors such as cognitive complexity (Bieri 1955), a verbal versus a visual style of information processing (Childers, Houston, and Heckler 1985), and the need and ability to achieve cognitive structuring (Bar-Tal, Kishon-Rabin, and Tabak 1997).

2.2. Tree Structures versus Spatial Configurations

It has been previously found that tree structure models outperform MDS 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 MDS models 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 spatial configuration than as a branch in a tree structure. In addition, Johnson and Hudson (1996) revealed that users found spatial configurations more useful when compared to tree structures.

Before the tree or MDS models are fitted, one may attempt to decide whether a tree or a spatial configuration is more appropriate on the basis of characteristics of the dissimilarity data. 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 well in that respect. The shape of a tree allows for many large distances between stimuli, whereas a low-dimensional spatial configuration does not. Hence, dissimilarity data with a large negative skewness generally fit a tree structure better relative to a spatial configuration.

2.3. Hybrid Models

As noted in the introduction, it is generally accepted that features and dimensions on the one hand and trees and spatial configurations on the other are complements rather than substitutes (Carroll 1976; Ghose 1998; Shepard 1980). Despite the added value of 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. Carroll and Pruzansky (Carroll 1976; Carroll and Pruzansky 1980) developed hybrid models that combine single and multiple tree structures and a single multidimensional spatial configuration. In their model, the dissimilarity between two stimuli corresponds to the sum of the distances derived from the trees and from the spatial configuration. Hubert, Arabie, and Meulman (Hubert and Arabie 1994; Hubert, Arabie, and Meulman 1997, 1998) developed procedures to represent a dissimilarity matrix through distances among a set of entities under certain types of constraints. Their procedures allow for approximating the dissimilarity matrix by the sum of several distance matrices. Specifying the appropriate sets of constraints for each of these distance matrices, results in a hybrid tree and space model as a special case of the Hubert, Arabie, and Meulman model. These hybrid models represent the data at the aggregate level, and 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 pairs of stimuli (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), Carroll and Pruzansky (Carroll 1976; Carroll and Pruzansky 1980), and Hubert, Arabie, and Meulman (Hubert and Arabie 1994; Hubert, Arabie, and Meulman 1997, 1998). Carroll and Chaturvedi (1995) proposed CANDCLUS, a general class of multilinear models and methods for the analysis of multi-way data with continuous, discrete, or mixed attributes. One of its important special cases is a hybrid model that accommodates individual differences through combining aspects of the overlapping clustering 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 both the discrete and continuous attributes. However, this approach 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 do not allow for parametric statistical inference and describe only the particular data set at hand. Stochastic models that explicitly postulate a probabilistic data generation mechanism describing uncertainties in the outcomes of the underlying process allow for parametric statistical inference, and enable generalizations from the sample to the population. However, in spirit our approach is consonant with that of Carroll and Chaturvedi (1995).

3. Mixture of Tree and Spatial Representations

3.1. The Model for Dissimilarity Judgments

Let $n = 1,\ldots,N$ denote subjects, $i, j, k = 1,\ldots,I$ denote stimuli, and $c = 1, 2$ denote $C = 2$ classes of judgment processes. In particular, we assume $c = 1$ to correspond to a judgment process based on common discrete features represented by an additive tree, and $c = 2$ to a judgment process based on continuous dimensions represented by a spatial MDS model. The data, δ_{ijn}, are the observed dissimilarities of the pair of stimuli i and j by subject n.

Throughout this paper, we will use the Greek symbol δ for observed data and Roman symbols (e.g. *c, d, i, p,* and *s)* for reconstructed, output data and model parameters. We consider the stochastic nature of the respondents' decision processes, by formulating a model that consists of a representation component and a measurement component, the former pertaining to additive tree respectively spatial representations assumed to capture subjects' dissimilarity judgments of stimuli. The measurement model involves distributional assumptions on the error, which enable us to adopt maximum likelihood (ML) estimation. Under certain regularity conditions, ML estimates for mixture models have such important properties as consistency of the estimates, not shared by models that include individual-specific parameters (Amemiya 1985, p. 115, 123).

We assume C = 2 unobserved classes, where Class 1 corresponds to an additive (or path length) tree structure and Class 2 to a spatial configuration. Because mixture models are invariant under interchanging of the class labels (McLachlan and Basford 1988, p. 12), this assignment is arbitrary. We assume that a particular subject in the sample, when making a dissimilarity judgment, draws from each of these two processes with prior probabilities, denoted as p_1 and p_2, respectively. Next, we assume the I(I-1)/2 dissimilarity judgments for subject *n,* using the process of Class *c,* to follow a normal distribution. Thus we have :

$$\phi_c(\delta_{ijn} \mid d_{ijc}, s_c) = \frac{1}{\sqrt{(2\pi s_c^2)}} \exp\left[-\frac{(\delta_{ijn} - d_{ijc})^2}{2s_c^2} \right]. \tag{1}$$

Here d_{ijc} is the expected value of δ_{ijn} in Class c, and s_c^2 its variance.

For Class 1, it is assumed that the dissimilarity judgments are produced by the distances in an additive tree (Corter 1996, p. 16-26; Gascuel and Levy 1996; Sattath and Tversky 1977), in which each quadruple of distances satisfies the additive or four-point inequality:

$$d_{ij1} + d_{kl1} \leq \min\left(d_{ik1} + d_{jl1}, d_{il1} + d_{jk1}\right) \ \forall \{i, j, k, l\}. \tag{2}$$

This inequality is identical to restricting the largest two of the three sums of distances between pairs of the four objects to be equal. The set of constraints in (2) imposes the additive inequality for Class 1 only.

For Class 2, it is assumed that the dissimilarities are produced by a constant term, *a*, and the Euclidean distances of 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 throughout this paper for ease of interpretation and because a tree apparently contains about the same amount of information as a two-dimensional spatial configuration (Carroll 1976, p. 453). In Ghose's (1998) and other comparisons of trees and spatial configurations, the spatial configurations were also restricted to be two-dimensional. Thus, for Class 2:

$$d_{ij2} = a + \sqrt{\left(\sum_{t=1}^{2}(x_{it} - x_{jt})^2\right)} \quad . \tag{3}$$

The effective number of parameters for the additive tree in $c = 1$ is 2I-3 (Corter 1996, p. 51); the MDS solution has 2I-2 effective parameters (DeSarbo, Manrai, and Manrai 1994, p. 199). In addition, there are 2 variance parameters and 1 prior probability to be estimated, adding up to K = 4I-2 effective parameters estimated for the model as a whole. The unconditional distribution of the dissimilarity judgments for subject n is formulated as:

$$\phi_n(\delta_{ijn}) = \sum_{c=1}^{2} p_c \phi_{nc}(\delta_{ijn} | d_{ijc}, s_c) = \sum_{c=1}^{2} p_c \prod_{i<j} \phi_c\left(\delta_{ijn} | d_{ijc}, s_c\right) \quad . \tag{4}$$

3.2. Estimation

The likelihood:

$$L = \prod_{n=1}^{N} \sum_{c=1}^{2} p_c \phi_{nc}(\delta_{ijn} | d_{ijc}, s_c) \tag{5}$$

is maximized under the constraints on the fitted distances provided by (2) and (3) using an EM algorithm (Dempster, Laird, and Rubin 1977). We provide the main features. That algorithm maximizes the likelihood through a series of major iterations, each consisting of an expectation step (E-step) and a maximization step (M-step), and minor iterations within each M-step for c = 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, p_{nc}, that subject n belongs to Class c, calculated at the current parameter estimates by means of Bayes's Theorem (see Equation (6) in Section 3.3).

Each M step for $c = 1$ is started using unconstrained estimation of the distances. We then parameterize the additive tree as an ultrametric tree plus a constant for each stimulus $d_{ij1} = d_i^* + d_j^* + d_{ij1}^*$ (Carroll 1976; DeSarbo, Manrai, and Manrai 1993). Doing so allows us to impose the (fewer) ultrametric constrains: $d_{ij1}^* \leq \max(d_{ik}^*, d_{ij1}^*)$, $\forall (i, j, k)$ on d_{ij1}^* rather than the additive constraints on d_{ij1}. This reduction is first done approximately by using the triple reduction method (TRM). TRM involves a repeated sequential averaging of the largest of two pairs of each triple (Roux 1987). The resulting distances do not yet completely conform to the ultrametric inequality. Subsequently, a Sequential Quadratic Programming method, involving a quasi-Newton maximization algorithm (Scales 1985, Section 3.5), is applied to enforce the ultrametric constraints on d^*_{ij1} exactly. This step completes the M-step for $c = 1$.

The M-step for $c = 2$ is initialized by a metric MDS based on a singular value decomposition of the double-centered matrix of dissimilarities averaged across subjects weighted with the class probabilities. Then an unconstrained maximization algorithm, involving a quasi-Newton maximization algorithm (Scales 1985, Section 3.5), is used, after imposing identifying constraints on the parameters. We approximate the required derivatives for both $c = 1$ and $c = 2$ numerically using forward differences. In each M step the parameter estimates from the previous steps are used as starting values, for $c = 1, 2$. The convergence criterion used on the average log-likelihood is 10^{-5}. For further details on the EM algorithm we refer to Dempster, Laird, and Rubin (1977) or Wedel and Kamakura (2000, Appendices A1 and A2). The EM algorithm is started from equal posterior probabilities ($p_{nc} = 0.5$; $n = 1,...,N$; $c = 1, 2$), so that each subject has an equal à priori probability of belonging to the additive tree and the spatial class.

3.3. Evaluation

Once the parameters of the model are estimated, the posterior probabilities, p_{nc}, that subject n has drawn upon process c (tree versus space), can be calculated using Bayes's Theorem. For the tree class the posteriors equal:

$$p_{n,1} = \frac{p_1 \phi_{n1}(\delta_{ijn} | d_{ij1}, s_1)}{p_1 \phi_{n1}(\delta_{ijn} | d_{ij1}, s_1) + p_2 \phi_{n2}(\delta_{ijn} | d_{ij2}, s_2)}, \tag{6}$$

where ϕ_{nc} is implicitly defined in (4). For the MDS class: $p_{n,2} = 1 - p_{n,1}$ These posterior probabilities are important quantities in our study, because they enable us to assess post hoc whether subject n has used the additive tree representation ($p_{n,1} = 1, p_{n,2} = 0$), or the spatial representation ($p_{n,1} = 0, p_{n,2} = 1$), or a mixture of both. The p_{ns}'s provide a probabilistic allocation of the subjects to the additive tree and spatial MDS classes, and thus enable one to judge à posteriori which judgment strategy a particular subject employs. We investigate this using an entropy measure E_2 (see Wedel and Kamakura 2000,

p. 92):

$$E_2 = 1 - \sum_{c=1}^{2}\sum_{n=1}^{N} \frac{-\hat{p}_{nc} \ln \hat{p}_{nc}}{N \ln(2)} \quad . \tag{7}$$

The entropy measure E_2 assesses the separation of the two classes and can thus be interpreted as the extent to which subjects use a single judgment process. Values close to unity indicate that subjects use a single strategy: i.e., a subject's 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, subjects draw with nonzero probabilities from both processes to arrive at their dissimilarity judgments.

To assess the fit of each model and to compare this fit across alternative model formulations, we compute an R^2 fit measure, being defined as:

$$R^2 = 1 - \frac{\sum_{c=1}^{C}\sum_{n=1}^{N} p_{nc} \sum_{i,j} (\delta_{ijn} - d_{ijc})^2}{\sum_{n=1}^{N}\sum_{i,j} (\delta_{ijn} - \overline{\delta_{ijn}})^2} \quad , \tag{8}$$

where $\overline{\delta_{ijn}}$ equals the average of the dissimilarity judgment across all subjects and pairs of stimuli. In addition, we compute AIC = $-2 \ln L + 2K$ (Akaike 1974). 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 generalizable conclusions on the use of discrete versus continuous dimensions in the pairwise dissimilarity judgments of stimuli.

3.4. Analysis of Synthetic Data Sets

To check the performance of the algorithm described above, we generated three synthetic data sets with C = 2 classes, I = 5 stimuli, and N = 20 subjects. The first data set was generated on the basis of one single tree for all subjects. The distances were generated according to Equations (1) and (2), where the true distances were taken from subsets of the stimuli in Table 5.3 in DeSarbo, Manrai, and Manrai (1993), and random error was drawn from N(0, 0.5). This data set was analyzed with the above mixture of tree and MDS configurations, starting from posterior probabilities of 0.5 for all subjects. The estimation procedure assigned all subjects correctly with a posterior probability of 1.0000 to Class 1, the tree structure. The second MDS class, was empty, all posteriors equaling 0.0000. Consequently $p_1 = 1.0$ $p_2 = 0.0$, and $E_2 = 1.0$, indicating that all subjects use the tree representation.

The second data set was generated on the basis of a single MDS model, where the distances were generated to conform to Equations (1) and (3). The stimulus coordinates were drawn from a N(0,2) distribution, random error from a N(0, 0.5), and $a = 20$ was used. This data set was analyzed with the mixture 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 being equal to 0.0000. Consequently $p_1 = 0.0$, $p_2 = 1.0$, and $E_2 = 1.0$, correctly indicating that all subjects use the spatial configuration.

The third data set was generated on the basis of an additive 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 the first and second data set above, respectively. The third data set was analyzed with the mixture model. The EM estimation algorithm correctly 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 equaling 1.0000. Consequently $p_1 = 0.5$, $p_2 = 0.5$, while again $E_2 = 1.0$, indicating that all subjects use a single strategy, but that half the subjects fit the spatial, and the other half the tree structure.

Thus, from these analyses of synthetic data, it appears that the proposed mixture model is capable of identifying the true decision process from the data, even if the number of 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 analyze the data published by Schiffman, Reynolds, and Young (1981, p. 33-34). In a sensory experiment, 10 subjects (nonsmokers, aged 18-21 years) tasted ten different brands of cola: Diet Pepsi, RC Cola, Yukon, Dr. Pepper, Shasta, Coca-Cola, Diet Dr. Pepper, Tab, Pepsi Cola, and

Diet Rite.[1] Each subject provided 45 pairwise dissimilarity judgments 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, for example bitterness, sweetness, and fruitiness, were collected from the same subjects.

To assess which structure best represents the dissimilarity judgments, we compare the proposed model to a number of alternative models. Two candidate models are the aggregate additive tree model and the aggregate spatial model. In addition, we estimate a two-class parametric mixture of trees and a two-class parametric mixture of spatial configurations. These benchmark models allow us to assess whether the proposed mixture of tree and spatial structure merely represents parametric heterogeneity in the sample. If the proposed mixed model outperforms the two-class tree model and the two-class spatial model, there is support for the existence of structural heterogeneity. Hence, we estimate the following five models:

1. The C = 1 additive tree, TREE(1), i.e., the model provided by Equations (1) and (2) with C = 1. This model corresponds to traditional additive tree models, estimated with ML. The number of parameters estimated equals K = 2I-2.
2. The C = 2 additive tree, TREE(2), i.e., a model provided by Equations (1) and (2) with C = 2. This model accounts for heterogeneity in the tree structure across unobserved classes. The additive restrictions in Classes 1 and 2 may differ, so that this model simultaneously identifies latent classes of subjects, as well as an additive tree structure for each class. This model has not been published previously and presents an extension of the mixture model for ultrametric trees by Wedel and DeSarbo (1998). The number of parameters estimated equals MK= 4I-3.
3. The C = 1 and T = 2 MDS model: MDS (1), i.e., a model provided by Equations (1) and (3) with C = 1 class only and T = 2 latent dimensions. This model corresponds to a traditional metric MDS model for pairwise dissimilarity data, estimated by ML. The number of parameters estimated equals K = 2I-1.
4. The C = 2 MDS model: MDS(2), i.e., a model provided by Equations (1) and (3) with C = 2 and T = 2. This model accounts for heterogeneity in the spatial representation of the stimuli across

1. Unlike the diet colas currently on the market, the diet versions in this data set were still saccharine-based, which affects the results of the taste test presented.

unobserved classes. The positions of the stimuli in the two-dimensional spatial configurations in Classes 1 and 2 may differ, reflecting different perceptual orientations. This model simultaneously identifies latent classes of subjects, as well as a spatial MDS structure for each class and was proposed by DeSoete, Meulman, and Heiser (1992) and described by DeSarbo, Manrai, and Manrai (1994, p. 198-200). The number of parameters estimated equals K = 4I-1.

5. The C = 2 mixture of additive 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 section 3. The number of parameters estimated equals K = 4I-2.

The fit of the additive tree and MDS mixture (AIC = 1021.2; R^2 = 0.509) is much better than that of either the aggregate tree (AIC = 1148.1; R^2 = 0.291) or aggregate MDS (AIC = 1136.0; R^2 = 0.313) solutions for the cola data. In addition, the mixed tree and MDS model fits better than either a mixture of two trees (AIC = 1034.1; R^2 = 0.494), or a mixture of two spatial configurations (AIC = 1060.3; R^2 = 0.463). Hence, for the cola data the mixed additive tree and MDS model is to be preferred.

The additive tree structure for the ten cola brands is presented in Figure 1. The additive tree can be thought to represent the distinctive features of the colas (Corter 1996). One branch of the tree contains three diet colas, which have relatively low interpoint distances: Diet Pepsi, Diet Rite, and Tab. The arc-length separating these diet colas from the others represents the summed weights of the distinctive features that these brands share. The interpretation of the arc as diet/regular feature is hampered 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 arc connecting these two brands to the others seems to be a distinctive brand taste feature: Dr. Pepper versus other brands, which can be interpreted as the presence/absence of the characteristic cherry flavor. The subtree in the upper part of the additive tree in Figure 1 shows a set of nodes that can be interpreted as representing distinctive brand features, distinguishing the five remaining nondiet brands.

Figure 2 presents the spatial representation of the ten cola brands. We try to fit the thirteen taste attributes into this configuration so as to label the dimensions. However, only two attributes, namely fruitiness and sweetness,

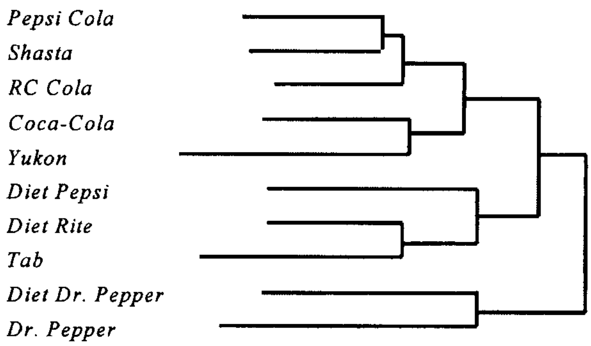


Figure 1. Additive Tree for the Schiffman et al. (1981) Cola Data

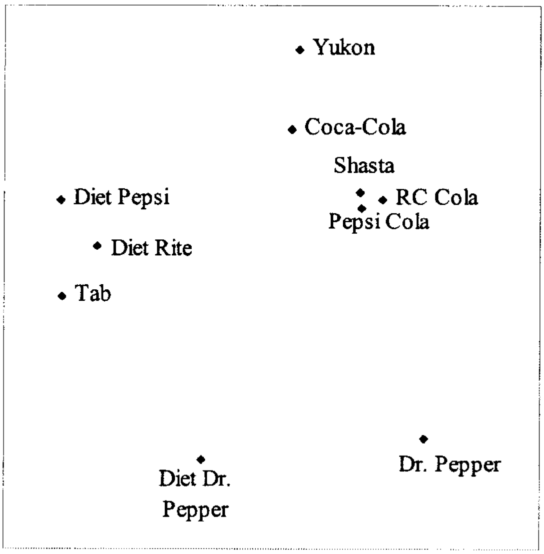


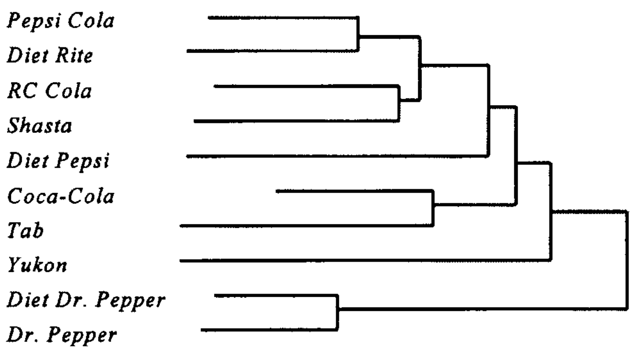
Figure 2. T=2 Dimensional Spatial Configuration for the Schiffman et al. (1981) Cola Data

reach a good fit (rho above 0.80) and two attributes, namely bitterness and stale versus fresh, a satisfactory fit (rho between 0.70 and 0.80). Seven attributes are clearly not useful for interpreting the dimensions (rho below 0.60). The four attributes that can be used to label the dimensions are highly correlated, and define only a single direction in the plot. Colas at the lower right of Figure 2 can be interpreted as more fruity, sweet, and fresh, and less bitter, whereas colas in the upper left are less fruity and sweet and more stale and bitter. The vertical dimension separates the Dr. Pepper brands at the top from the other brands, which apparently results from the specific cherry taste of the Dr. Pepper brand. The horizontal dimension separates the diet versus the nondiet brands: on the left side of the horizontal axis one observes Diet Dr. Pepper, Diet Pepsi, Tab, and Diet Rite, on the right side, the nondiet brands Yukon, RC Cola, Pepsi, Shasta, and Coca-Cola. Among the nondiet cola brands on the upper right side of the plot there seems to be fairly little distinction, but 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.

Comparing the additive tree and the spatial representations (respectively Figures 1 and 2), one has to conclude that the interpretation of both is very similar: the same three groups 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 additive tree and the spatial configuration depict the taste distinction between diet and regular colas, and between brands with and without cherry flavor. These attributes are best interpreted as discrete features. No clear, continuous dimensions are found in the spatial representation. Because discrete features correspond to the behavioral model underlying tree structures, the tree representation seems to depict a clearer structure among the cola brands.

To illustrate the insights provided by our mixed model of a tree and a spatial configuration, we provide in Figure 3 the results of an analysis of the Schiffman, Reynolds, and Young (1981) cola data with our procedure. The mixture model results show that there are two well separated classes: all posterior probabilities of membership deviate less than 0.001 from either zero or one, and the entropy measure E_2 equals 0.999. Each class comprises 5 subjects. The subjects in the tree-class seem to identify specific brand tastes: Dr. Pepper and Diet Dr. Pepper, and their respective counterparts Coca-Cola and its diet version Tab, are in separate subtrees. Note that the Dr. Pepper brands stand out particularly, as indicated by the length of the arc connecting. them to the other brands. However, the exception is that Diet Pepsi and Pepsi are not joined in a specific subtree, which shows Pepsi was not able to produce a diet version tasting similar to its nondiet version. Nevertheless, we conclude that the specific brand tastes are the distinctive features determining dissimilarity judgments in this class. In the MDS class, the vertical dimension separates the diet and the nondiet colas. 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, with the Dr. Pepper brands being on the one extreme and Coca-Cola brands on the other extreme of the dimension. We

Panel (a), Class 1: Additive Tree Structure



Panel (b), Class 2: Spatial Configuration

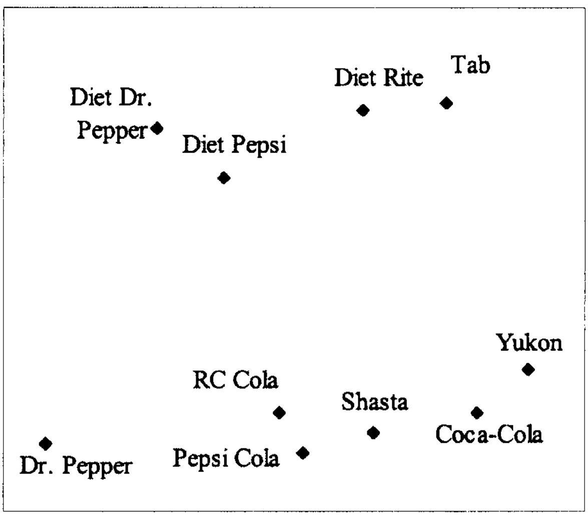


Figure 3. C=2 Mixed Tree-Space Solution for the Schiffman et al. (1981) Cola Data

fitted the thirteen attributes into the configuration of this MDS class; now six attributes reach a good fit (rho above 0.80) and three a satisfactory fit (rho between 0.70 and 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 side being perceived more sweet and fresh and less bitter, sour, stale, and chemical. It is interesting to note that the subjects in the MDS class all have the ability to taste phenylthiocarbamide (PTC; a chemical compound that tastes bitter), while subjects in the tree class do not have this ability. This ability is determined by one gene on chromosome seven (Srb, Owen, and Edgar 1965, p. 401). For the subjects in this sample, whose perceptual process appears based on continuous dimensions, such as sweetness and bitterness, or instead on discrete features, the judgments seem to be determined here by a single genetic factor. Once subjects have the ability to taste bitter, the continuous sweetness/bitterness dimension dominates their perceptions. Subjects lacking this ability base their judgments on discrete brand-taste features.

Comparing the mixed additive tree and spatial solution in Figure 3 with the aggregate tree and spatial configuration 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 evel. In the mixed model, the continuous sweetness dimension is evident, which is not recovered in the separate additive tree and MDS analyses, for example. Hence, 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 subjects' true underlying perceptual structures are heterogeneous.

5. Analysis of Twenty Empirical Data Sets

Following studies by Pruzansky, Tversky, and Carroll (1982), Johnson et al. (1992), and Ghose (1998), we analyze twenty data sets of pairwise dissimilarities to investigate the adequacy of tree versus spatial representations. We restrict the analysis to data sets that pertain to pairwise dissimilarity judgments[2], and do not consider other types of dissimilarity judgments (for example triadic combinations or free sorting), derived dissimilarity data (for example computed from attribute ratings), brand switching data, or co-occurrence data. The reason is that we are interested in identifying the processes underlying pairwise dissimilarity judgments, and in particular in identifying differences in those processes. The analyses enable us to draw conclusions on whether or not the type of representation model

2. Strictly speaking, the dissimilarity judgments in our twenty empirical data sets are ordinal. However, given the large number of scale points (7 to 11), this type of data can arguably be treated as interval scale (see e.g. Bijmolt and Wedel 1999), which is commonly done in empirical applications.

that best describes the dissimilarity judgments 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 thirteen data sets are commercial stimuli, in particular brands of fast moving consumer goods (fmcg), durables, services, and media. In addition, we analyze seven data sets on non-commercial stimuli, namely locations and emotions. The number of stimuli in the data sets ranges from 8 to 15, the number of subjects ranged from 10 to 60. Dissimilarity judgments are provided on 7-, 9-, or 11-point scales. The two data sets on emotions have been published previously and accompany the MULTISCALE program (Ramsay 1982, 1991). The other data sets are collected by the authors, and some of these data sets have been used in other studies (Bijmolt and Wedel 1995; Bijmolt, et al. 1998; Bijmolt, DeSarbo, and Wedel 1998).

Before running the analyses, the dissimilarity data are standardized to zero mean and unit variance by subject, to prevent the solutions becoming confounded with the effects of response strategies (Bijmolt et al. 1998). Most data sets do not contain missing values, but a few data sets have a small percentage of missing values that are imputed by mean substitution for each individual before standardization. As in the application to the cola taste data, for each of the twenty data sets we estimate the same five models as in Section 4, namely a single additive tree model, a two additive trees model, a single spatial configuration model, a two spatial configurations model, and the mixed tree and spatial configuration model. Following the previous literature in this area (Johnson et al. 1992; Pruzansky, Tversky, and Carroll 1982; Ghose 1998) we report the fit for each of those five models, in particular the AIC criterion and the percentage of variance explained[3]. Tables 2 and 3 present the results, where the model that has the lowest AIC value, or respectively, explains the largest percentage of variance for a particular data set is underlined. Because the R^2 and AIC statistics yield highly similar conclusions (by identifying the same models as best for all data sets except for supporting facilities), we only discuss the results based on the latter.

3. Note that the Likelihood ratio test cannot be used to compare the models. The tree and spatial models are not nested, while the one-class specification of a tree or spatial model is on the boundary of the parameter space of the corresponding two-class specification (Titterington, Smith, and Makov 1985, p. 4; Aitkin and Rubin 1985).

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 Channels	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

Model selection based on the lowest value of AIC clearly shows that heterogeneity is important. For nineteen out of the twenty data sets, a model with two latent classes yields the lowest AIC. For one data set, namely supporting facilities, the single spatial configuration model is indicated as most appropriate. The two-tree model does best seven times, the two spatial configurations model four times, and the model with both a tree and a spatial representation eight times. This result supports the need for modeling heterogeneity in stimulus representation, be it either parametric or structural.

We examined whether the type of stimuli, the number of stimuli, the

Table 2: AIC statistics for the Twenty Data Sets

Stimuli	TREE(1)	TREE(2)	MDS(1)	MDS(2)	MIX
Soft drinks	9944.1	9721.7	9885.0	10445.6	9827.9
Candy bars	8484.9	8276.2	8511.0	8899.8	8358.9
Shampoos	5980.7	5889.6	5980.8	5953.0	5885.8
Beer	1964.5	1904.9	1966.1	2023.6	1921.8
Cars	8941.8	8857.9	8947.9	8871.1	8931.0
Audio	1983.0	1945.3	1977.3	1938.9	1944.2
Supermarkets	8747.0	8507.2	8696.3	8466.8	8499.0
Recreation facilities	9162.4	8739.0	9141.2	8962.0	9012.6
Banks	8828.7	8348.3	8813.9	8412.1	8375.3
Restaurants	3963.9	3939.5	3986.7	3960.1	3921.8
Supporting facilities	1781.5	1742.0	1728.6	1764.5	1746.1
Bars	1890.3	1824.2	1891.0	1831.4	1802.1
Weekly magazines	10516.7	9845.6	10497.8	10043.6	9792.4
Women's magazines	5092.8	5018.1	5095.2	5063.0	5011.8
TV Channels	1501.6	1370.8	1507.5	1549.3	1418.3
Countries	4173.0	4144.9	4162.8	4180.1	4135.2
Capitals	1289.7	1262.9	1292.7	1242.6	1270.8
Cities	1732.6	1662.0	1712.1	1667.9	1643.8
Emotions-1	2891.4	2865.8	2846.8	2883.3	2837.5
Emotions-2	1038.3	1026.6	995.1	970.3	991.0

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

type f subjects, the number of subjects, and the number of scale points, as shown in Table 1, affect the model indicated as most appropriate by AIC. Bivariate analyses (analysis of variance with F-tests and cross tabulations with chi-square tests) did not show any significant relationship (all p-values > 0.05). In addition, examining the effects of all factors simultaneously in a discriminant analysis (see Stevens 1996, Chapter 7) showed no significant differences between the alternative models (Wilks' $\lambda = 0.55$, $\chi^2 = 8.98$, d.f. = 10, p = 0.53). Hence, we tend to conclude that the relative fit of the three types of mixture models is not related to the factors mentioned above (the results are supported by a similar analysis of the R^2 statistics).

Table 3: The R^2 statistics for the Twenty Data Sets

Stimuli	TREE(1)	TREE(2)	MDS(1)	MDS(2)	MIX
Soft drinks	0.275	0.323	0.286	0.185	0.305
Candy bars	0.232	0.289	0.227	0.142	0.267
Shampoos	0.004	0.061	0.005	0.036	0.063
Beer	0.118	0.223	0.118	0.091	0.189
Cars	0.013	0.053	0.012	0.049	0.030
Audio	0.095	0.180	0.104	0.192	0.183
Supermarkets	0.169	0.237	0.182	0.248	0.239
Recreation facilities	0.057	0.155	0.064	0.124	0.111
Banks	0.148	0.256	0.152	0.247	0.248
Restaurants	0.083	0.119	0.070	0.112	0.133
Supporting facilities	0.205	0.290	0.267	0.271	0.272
Bars	0.204	0.301	0.205	0.299	0.326
Weekly magazines	0.163	0.278	0.167	0.238	0.290
Women's magazines	0.006	0.066	0.005	0.044	0.067
TV Channels	0.536	0.610	0.533	0.515	0.606
Countries	0.027	0.082	0.035	0.059	0.089
Capitals	0.115	0.223	0.113	0.244	0.207
Cities	0.360	0.446	0.380	0.435	0.457
Emotions-1	0.526	0.552	0.542	0.547	0.561
Emotions-2	0.200	0.259	0.287	0.363	0.346

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

Next, we examine the results of the hybrid mixture model for the twenty data sets more closely. The AIC and R^2 statistics of the hybrid mixture model are shown in Tables 2 and 3, and the entropy statistic and the proportions of the tree and MDS classes are reported in Table 4. In addition, we report the average skewness of the dissimilarities for the tree and the MDS class in Table 4, because those statistics seem to be the most important indicators of the appropriateness of a tree or spatial configuration (Ghose 1998). We use the third central moment divided by the cubed standard deviation as measure of skewness, as used by Ghose (1998) and Pruzansky, Tversky, and Carroll (1982).

Table 4: Mixed Tree and MDS Model Results for the Twenty Data Sets

Data set	Entropy	p_1	p_2	Skewness tree	Skewness MDS
Soft drinks	0.990	0.850	0.149	-0.718	-0.403
Candy bars	0.987	0.898	0.101	-0.449	-0.291
Shampoos	0.950	0.774	0.226	0.019	-0.050
Beer	0.997	0.850	0.149	-0.311	-0.261
Cars	0.956	0.907	0.092	-0.449	-0.351
Audio	0.869	0.583	0.416	-0.036	0.168
Supermarkets	0.905	0.440	0.559	0.395	0.149
Recreation facilities	0.881	0.499	0.500	-0.052	-0.036
Banks	0.979	0.395	0.604	-0.028	0.020
Restaurants	0.876	0.436	0.563	0.540	0.471
Supporting facilities	0.991	0.734	0.265	-0.214	-0.226
Bars	0.999	0.249	0.750	-0.114	-0.344
Weekly magazines	0.989	0.432	0.567	-0.062	0.001
Women's magazines	0.976	0.777	0.222	0.029	0.023
TV Channels	0.999	0.800	0.200	-0.595	-0.567
Countries	0.999	0.571	0.428	0.098	-0.173
Capitals	0.972	0.693	0.306	-0.164	0.056
Cities	0.988	0.500	0.499	-0.023	-0.298
Emotions-1	0.973	0.261	0.738	-0.269	-0.088
Emotions-2	0.976	0.211	0.788	-0.069	0.072

The percentage explained variance, R^2, as defined in Equation (8) and shown in Table 3, varies substantially across the data sets, with a minimum of 0.030 for cars and a maximum of 0.606 for TV channels. Several R^2 values are rather low, which might result from the fact that a large number of observations, NI(I-1)/2, is represented by a relatively small number of parameters, K= 4I-2. The values of R^2 are paralleled by those of AIC.

In seventeen of the twenty applications, the classes are very well separated ($E_2 > 0.900$), thus indicating that nearly all subjects contributing these data sets base their judgments on either the additive tree structure or the spatial configuration. In particular, looking at the entropy and the prior probabilities, candy bars and cars seem to be almost entirely judged on the basis of features, while bars and emotions appear to be perceived predominantly according to continuous dimensions. For three sets of stimuli, namely audio, recreation facilities, and restaurants, the entropy measure is somewhat lower (around 0.85), and both proportions are around 0.50, which indicates that subjects are somewhat inclined to perceive these stimuli using both discrete features and continuous dimensions.

The estimates of the prior probabilities show that across all twenty data sets, the additive tree structure is somewhat more important than the spatial configuration. The respective average proportions, 0.593 and 0.407 for the additive tree and the spatial configuration, differ significantly at an a-level of 0.10 (paired samples t-test: $t = 1.85$; $df = 19$; $p = 0.08$). For twelve data sets the tree structure is more important; i.e., p_1 is larger than p_2, whereas for eight data sets the spatial configuration is more important; i.e., p_2 is larger than p_1. However, this result does not significantly deviate from a 50-50 split (non-parametric sign-test assuming a binomial distribution; $p = 0.50$). For most sets of stimuli, the size of both tree and spatial components is substantial. The contribution of the tree structure (spatial configuration) ranges from 0.907 (0.092) for cars to 0.249 (0.750) for bars.

We perform three analyses of covariance (see Stevens 1996, Chapter 9) to determine whether the proportion of tree versus MDS classes, the entropy measure, and the R^2 fit statistic (see Tables 3 and 4) are affected by the type of 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 analyses of covariance yield one significant effect: the type of stimuli affects the proportion of the tree class ($F = 4.44$; d.f. $= 5$; $p = 0.02$). The tree class is larger for fast moving consumer goods (average proportion of 0.843) and durables (0.746), and smaller for services (0.459) and emotions (0.237).[4] This result is in line with results of previous studies (Pruzansky, Tversky, and Carroll 1982; Johnson et al. 1992) that reported effects of the type of stimuli on the relative fit of additive tree structure models and MDS methods. In particular, tree structure models have been shown to outperform MDS methods in fitting conceptual stimuli rather than perceptual stimuli (Pruzansky, Tversky, and Carroll 1982) and concrete

5. When we tested the relative fit of three competitive mixture models, no significant effect of the type of stimuli was found. However, the direction of the results were similar to those presented in this paragraph, e.g., for fast moving consumer goods and durables in four out of six cases the two tree structures model was indicated by AIC as most appropriate, whereas for services and emotions this was the case for two out of eight data sets (see Table 2). The fact that this effect was not significant might caused be by the relatively small number of data sets (20), which reduced the power of this particular chi-square test having a substantial number of degrees of freedom (10).

stimuli such as brands rather than of more abstract stimuli such as product categories (Johnson et al. 1992). Bivariate analyses show a negative correlation between the number of subjects and the explained variance (correlation = -0.44, $p = 0.05$). This effect can be explained by the fact that for larger numbers of subjects, the model parameters represent the judgments of individual subjects less well, which results in lower R^2.

The skewness of the dissimilarity data is lower in the tree class than in the MDS class for twelve of the twenty data sets (Table 4). This finding corresponds to those of previous studies, for example Ghose (1998), which have shown that skewness discriminates between trees and spatial configurations, the former having a more negative skewness. However, a non-parametric sign-test shows that the result of our study does not deviate significantly from a 50-50 split (assuming a binomial distribution; $p = 0.50$).

6. Conclusion and Discussion

We propose a mixture model of an additive tree structure and a spatial configuration for the analysis of dissimilarity judgments. The mixture model accounts for heterogeneity among 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 modeled as distances in an additive tree and the dissimilarity judgments of the other class are modeled as distances in a Euclidean space. The model thus accommodates structural heterogeneity among classes, which is an advance over the parametric heterogeneity accommodated in previous mixture models. Through the analysis of synthetic data sets, we show that the model adequately recovers known mixed tree and spatial structures that may underlie dissimilarity data. The results of the mixture model are illustrated in an application to previously published data from a sensory experiment with colas. Analysis with the mixed tree and MDS model yields a much better and richer representation of the dissimilarities than the results obtained from using aggregate level models. Hence, the application demonstrates that our model yields highly interpretable, useful solutions, whereas pure tree or MDS models, ignoring structural heterogeneity, may lead to erroneous conclusions.

In the application of the mixed model to twenty empirical data sets, we find substantial evidence of heterogeneity across subjects for each data set. When examining the mixed tree structure and spatial configuration model, in general, both the classes turned out to be reasonably large and they were well separated. Hence, there seem to be clear differences among individual subjects with respect to whether a feature-based or a dimension-based representation fits their dissimilarity judgments better. These results support the need to model heterogeneity in stimulus representation, either by assuming parametric heterogeneity or structural heterogeneity. MDS, tree, or 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.

There turned out to be no significant relationship between on the one hand the relative fit of the two-tree structures, two spatial configurations, and tree-space mixture models 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. Furthermore, we found that these study design factors had little to no effect on the relative importance of the tree versus the space in the tree-space mixture. These findings may be considered reassuring, because in applying our model one may assume that the outcomes are relatively robust against the study design.

Consistent with previous studies (Johnson and Fornell 1987; Johnson and Hudson 1996; Johnson et al. 1992; Pruzansky, Tversky, and Carroll 1982), we found the importance of the tree structure relative to the spatial configuration to depend on the type of stimuli. In particular,the class proportions show that the additive tree class is larger than the MDS class for fast moving consumer goods and durables, whereas the MDS class is larger than the tree class for services and emotions. Previous studies showed that tree structure models outperform MDS methods in fitting dissimilarity judgments between conceptual stimuli and concrete stimuli such as brands, whereas the opposite holds for perceptual stimuli and more abstract stimuli. Our results appear to support these earlier findings, because services and emotions can be considered more abstract than fast moving consumer goods and durables.

The skewness of the dissimilarity judgments is somewhat lower for the tree class than for the MDS class, although the difference is not significant. This result corresponds to findings of Ghose (1998), who has shown that skewness discriminates between trees and spaces, the former having a more negative skewness.

Further research is needed several directions. First, as evidence of the appropriateness of spatial versus tree representations for certain types of stimuli accumulates, a meta-analysis of published results should provide a more definitive conclusion on the topic. Second, 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, our study revealed that individual differences may be more important in that respect. Therefore, future research should address subject-related factors, such as cognitive complexity (Bieri 1955), style of processing (Childers, Houston, and Heckler 1985), and the need and ability to achieve cognitive structuring (Bar-Tal, Kishon-Rabin, and Tabak 1997), as possible determinants of the adequacy of tree and spatial structures to fit dissimilarity judgments at the individual- or class-level. Because our analyses are based on previously collected data, information on those characteristics was not available. Third, research on 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. Those authors showed that the relative fit of alternative tree structures (ultrametric and additive trees) and spatial configurations (Euclidean and city-block distances) may not be very strongly related to the psychological processes presumed to underlie the data. Hence, care should be taken in considering the results of tree structure models, MDS models, or mixed models as direct evidence of the underlying psychological processes. If the main interest is to reveal the processes underlying dissimilarity judgments, alternative approaches should be used in conjunction with statistical modeling procedures such as the one described in this paper. One could examine the underlying psychological processes and the judgment task through studies extending, for example, the work of Bijmolt et al. (1998) using a process-tracing perspective, that is through the analysis of verbal protocols of dissimilarity judgments. In addition, carefully designed experiments can shed further light on the issue. Such studies may focus on (a) the nature of the attributes used by respondents while comparing stimuli and (b) the characteristics of the respondents, the stimuli, and the judgment task, that affect subjects' perceptual representation of stimuli.

References

AITKIN, M., and RUBIN, D.B. (1985). "Estimation and Hypothesis Testing in Finite Mixture Distributions," *Journal of the Royal Statistical Society Series B, 47*, 67-75.

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.

BAR-TAL, Y., KISHON-RABIN, L., and TABAK, N. (1997), "The Effect of Need and Ability to Achieve Cognitive Structuring on Cognitive Structuring," *Journal of Personality and Social Psychology, 73*, 1158-1176.

BARTHÉLEMY, J.P., and GUÉNOCHE, A. (1991), *Trees and Proximity Representations,* Chichester: Wiley.

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

BIJMOLT, T.H.A., DESARBO, W.S., and WEDEL, M., (1998), "A Multidimensional Scaling Model Accommodating Differential Stimulus Familiarity," *Multivariate Behavioral Research, 33*, 41-63.

BIJMOLT, T.H.A., and WEDEL, M. (1995), "The Effects of Alternative Methods of Collecting Similarity Data for Multidimensional Scaling," *International Journal of Research in Marketing, 12*, 363-371.

BIJMOLT, T.H.A., and WEDEL, M. (1999), "A Comparison of Multidimensional Scaling Methods for Perceptual Mapping," *Journal of Marketing Research, 36*, 277-285.

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. (1998), "Multidimensional Scaling", in: *Handbook of Perception and Cognition, Volume 3: Measurement, Judgment and Decision Making*, M.H. Birnbaum (ed.), San Diego: Academic Press, 179-250.

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 CHANG, J.J. (1973), "A method for Fitting a Class of Hierarchical Tree Structure Models to Dissimilarities Data and its Application to some 'Body Parts' Data of Miller's", *Proceedings of the 81st Annual Convention of the American Psychological Association, 8 (part 2)*, 1097-1098.

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, NJ: 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.), Amsterdam: Elsevier, 193-257.

DESARBO, W.S., MANRAI, A.K., and MANRAI, L.A. (1994), "Latent Class Multidimensional Scaling: A Review of Recent Developments in the Marketing and Psychometric Literature," in: *Advanced Methods of Marketing Research*, R.P. Bagozzi (ed.), Cambridge, MA: Blackwell, 190-222.

DE SOETE, G., MEULMAN, J., and HEISER, W. (1992), "A Mixture Distribution Approach to Points of View of Analysis," Paper presented at the Annual Meeting of the Classification Society of North America, Michigan State University, East Lansing, Michigan.

DE SOETE, G., and CARROLL, J.D. (1996), "Tree and other Network Models for Representing Proximity Data," in: *Clustering and Classification*, P. Arabie, L.J. Hubert, and G. De Soete (eds.), River Edge, NJ: World Scientific, 157-197.

GARNER, W.R. (1978), "Aspects of a Stimulus: Features, Dimensions, and Configurations," in: *Cognition and Categorization*, E. Rosch and B.B. Lloyd (eds.), Mahwah, NJ: Erlbaum, 99-133.

GASCUEL, O., and LEVY, D. (1996), "A Reduction Algorithm for Approximating a (Non-metric) Dissimilarity by a Tree Distance," *Journal of Classification, 13*, 129-155.

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.

HUBERT, L., and ARABIE, P. (1994), "The Analysis of Proximity Matrices Through Sums of Matrices Having (Anti-)Robinson Forms," *British Journal of Mathematical and Statistical Psychology, 47*, 1-40.

HUBERT, L., ARABIE, P., and MEULMAN, J. (1997), "Linear and Circular Unidimensional Scaling for Symmetric Proximity Matrices," *British Journal of Mathematical and Statistical Psychology, 50*, 253-284.

HUBERT, L., ARABIE, P., and MEULMAN, J. (1998), "Graph-Theoretic Representations for Proximity Matrices Through Strongly-Anti-Robinson or Circular Strongly-Anti-Robinson Matrices," *Psychometrika, 63*, 341-358.

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.

MCLACHLAN, G.J., and BASFORD, K.E. (1988), *Mixture Models: Inference and Application to Clustering*, New York: Marcel Dekker.

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

RAMSAY, J.O. (1982), "Some Statistical Approaches to Multidimensional Scaling Data," *Journal of the Royal Statistical Society, 145 (A)*, 285-312.

RAMSAY, J.O. (1991), *MULTISCALE Manual.* Montreal, Canada: Department of Psychology, McGill University.

ROUX, M. (1987), "Techniques of Approximation for Building Two Tree Structures," in: *Proceedings of the Franco-Japanese Scientific Seminar: Recent Developments in Clustering and Data-Analysis,* Diday, E. (Ed.) Tokyo, 127-146, Boston: Academic Press.

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.

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

SRB, A.M., OWEN, R.D., and EDGAR, R.S. (1965). *General Genetics*, San Francisco: Freeman.

STEVENS, J. (1996), *Applied Multivariate Statistics for the Social Sciences*, third edition, Mahwah, NJ: Erlbaum.

TITTERINGTON, D.M., SMITH, A.F.M., and MAKOV, U.E. (1985), *Statistical Analysis of Finite Mixture Distributions*, New York: Wiley.

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.), Mahwah, NJ: Erlbaum, 79-98.

WEDEL, M., and DESARBO, W.S. (1998), "Mixtures of (Constrained) Ultrametric Trees," *Psychometrika, 63*, 419-443.

WEDEL, M., and KAMAKURA, W.A. (2000), *Market Segmentation: Conceptual and Methodological Foundation, second edition,* Dordrecht: Kluwer.