Psychologica Belgica
1995, 35-2/3, 85-102.

ANALYSIS OF SIMILARITY DATA AND TVERSKY'S CONTRAST MODEL

Iven VAN MECHELEN and Gert STORMS
Katholieke Universiteit Leuven

In this paper we briefly review Tversky's main criticisms of spatial models of similarity. We then recapitulate the contrast model that was proposed to remedy for the shortcomings of spatial models. Next, four data-analytic methods that are based on the contrast model are described and illustrated with analyses of a small data set; special attention is given to the interrelations of the models on which the methods are based. Finally, the four data-analytic methods are checked against Tversky's criticisms of geometric models of similarity and against recent cognitive objections to the contrast model.

Similarity is a central concept in theories of several domains of human behavior, including classification, learning, and social interaction. Furthermore, in several psychological experiments similarity or dissimilarity data are obtained in different ways. For example, subjects may be asked to judge directly the (dis)similarity of each pair of objects out of given set of objects (i.e., stimuli, persons, brands, etc.). Other experiments involve the collection of data from which the psychologist may wish to derive (dis)similarities. For example, (dis)similarities may be derived from a sorting task in which subjects are asked to partition a set of objects into groups of similar objects; (dis)similarities may also be derived from object by variable data (e.g., by calculating correlations or Euclidean distances between objects across variables).

Among psychological theories of similarity that have been developed during the last decades, a distinct subset is based on formal models. Those models, in general, contain rigorous characterizations of the structural basis that may underlie similarity judgments. Data-analytic procedures have been developed to analyze proximity data (i.e., dissimilarity or similarity data) in terms of the formal models in question. The result of such an analysis may be looked at as a parsimonious mathematical reduction of the data; alternatively, one may wish to go beyond data reduction and assess a formal proximity model with respect to its validity as a cognitive theory.

The authors thank Wim Ruts, Jean-Pierre Thibaut and André Vandierendonck for their helpful comments on a previous draft of this paper. The research reported in this paper was supported by Grant 2.0073.94 from the Belgian National Fund for Scientific Research (NFWO) to P. De Boeck, I. Van Mechelen, and D. Geeraerts.

Correspondence concerning this article should be addressed to Iven Van Mechelen, Department of Psychology, Tiensestraat 102, B-3000 Leuven (Belgium). Electronic mail may be sent to Iven.VanMechelen@psy.kuleuven.ac.be.

A major class of formal similarity models is the class of geometric or spatial models. This class includes the well-known models of multidimensional scaling (Carroll & Arabie, 1980; Kruskal, 1964a, 1964b; Shepard, 1962a, 1962b, 1980), which represent the objects as points in a low-dimensional metric space such that the distances between the object points reflect the observed proximities. The relevance of the class of spatial models for the cognitive study of similarity has recently been reviewed by Nosofsky (1992).

In a series of papers, however, the assumptions underlying spatial models of similarity have been questioned by Tversky and his colleagues (e.g., Gati & Tversky, 1982; Tversky, 1977; Tversky & Gati, 1982) (see however also Pruzansky, Tversky, & Carroll, 1982). As an alternative formal model, Tversky developed the contrast model that represents proximities in terms of sets of underlying features. In this paper we will first briefly review Tversky's critique of the assumptions underlying spatial models of similarity and recapitulate the contrast model. Next we will discuss four data-analytic methods that are based on special cases of the contrast model. Finally, we will test the four data-analytic methods against both Tversky's criticisms and against more recent cognitive psychological objections to the contrast model (Medin, Goldstone, & Gentner, 1993).

Tversky's Criticisms on Spatial Models for Similarity Data

Tversky's criticisms on spatial models can be divided into two categories: A first series of criticisms refers to the metric axioms; a second series of criticisms refers to assumptions involved in the use of the Minkowski metrics that are typically used in spatial models of similarity.

With respect to the first category of criticisms, recall that a nonnegative function δ, defined on all possible pairs of objects, is a *distance* (or *metric)* if and only if the following three axioms hold:

$$\forall a,b\colon \delta(a,b) \geq \delta(a,a) = 0 \qquad \text{(positive definite)}$$
$$\forall a,b\colon \delta(a,b) = \delta(b,a) \qquad \text{(symmetry)}$$
$$\forall a,b,c\colon \delta(a,c) \leq \delta(a,b) + \delta(b,c) \qquad \text{(triangle inequality)}$$

Tversky (1977) argues that the metric axioms do not necessarily hold for similarity judgements. As regards positive definiteness, for instance, he indicates that the probability of judging two identical objects as "same" rather than "different" is usually not constant for all objects (e.g., for more complex objects this probability may be lower). As regards symmetry, Tversky found evidence in perceptual and conceptual data that similarities can systematically and significantly deviate from symmetry; in general, a more salient (prototypical) object appears less similar to a less salient (non-prototypical) object than vice versa. As regards the triangle inequality, this implies that if a is quite similar

to *b*, and *b* is quite similar to *c*, then *a* cannot be very dissimilar from *c;* Tversky (1977) gives a few examples that cast doubt on the general tenability of this implication in psychological data.

A second series of criticisms does not apply to all metrics but to the metrics that are almost invariably used in scaling models, in particular to the family of Minkowski metrics. Recall that if o_i and o_j are points in an *R*-dimensional space with coordinates $(x_{ir}, r = 1,\ldots, R)$ and $(x_{jr}, r = 1,\ldots, R)$, the Minkowski distance between those points is defined as:

$$d(o_i,o_j) = \left[\sum_{r=1}^{R} |x_{ir} - x_{jr}|^{\gamma}\right]^{1/\gamma},$$

with $\gamma \geq 1$. The familiar Euclidean metric is obtained with $\gamma = 2$, the city-block metric with $\gamma = 1$, and the dominance metric with $\gamma = \infty$. The family of Minkowski metrics incorporates many assumptions including *interdimensional additivity* and *segmental additivity* (for exact definitions of the latter, see Gati & Tversky, 1982; Tversky & Gati, 1982). Interdimensional additivity implies that the dissimilarity of objects that vary along one dimension must be *independent* of the fixed value those objects take on another dimension. Segmental additivity means that if three objects o_i, o_j and o_k vary along a straight line, then

$$\delta(o_i,o_k) = \delta(o_i,o_j) + \delta(o_j,o_k).$$

In a series of experiments, Gati and Tversky demonstrate failures of independence and of the combination of segmental additivity and the triangle inequality, in addition to other failures of the Minkowski metrics in the representation of psychological data.

The Contrast Model

In order to remedy for the shortcomings of spatial models for similarity data, Tversky (1977) formulated the contrast model. The latter model represents the similarity between a pair of objects as a function of underlying features (i.e., *formal variables* that take a value of zero or one for each object). In particular, if *s(a,b)* is a measure of the similarity defined for all nonidentical pairs of objects *a* and *b*, and if *A* and *B* denote the feature sets of *a* and *b*, then, according to the contrast model, there exists a monotone increasing transformation *S* of *s*, a measure *f* defined on the set of all features, and nonnegative parameters θ, α, β, that are such that:

$$S(a,b) = \theta f(A \cap B) - \alpha f(A \setminus B) - \beta f(B \setminus A). \qquad (1)$$

In this formula, $A \cap B$ denotes the set of features common to *a* and *b*, $A \setminus B$ denotes the set of features that apply to *a* but not to *b*, and $B \setminus A$ denotes the set

of features that apply to b but not to a. Tversky proved that, if a similarity measure s is a function of sets of common and distinctive features, $A \cap B$, $A \setminus B$, $B \setminus A$, and if s satisfies a number of additional assumptions, then the function that expresses S in terms of the feature sets can be represented in terms of the contrast model (1). Under additional assumptions, the measure f can further be proven to be an additive function of the features. In general, $f(X)$ can be conceived as an index of the salience or importance of feature set X.

The contrast model can be shown to account for a number of phenomena that were problematic for spatial models. For example, violations of the symmetry axiom can be accounted for by accepting different values for α and β in (1). As another example, it is easy to see that the independence assumption may be violated; for, according to the contrast model, by the addition of the same feature to two objects, their similarity increases.

Data-Analytic Techniques Based on the Contrast Model

Data-analysis based on the contrast model typically starts from a proximity matrix defined on a set of objects. The data analysis may further be of two different types depending on whether the features are known in advance or not. We limit ourselves to the case where the features are not yet known and, hence, are to be induced during the analysis.

It is impossible to construct a data-analytic procedure based on the contrast model in its most general formulation. Given a set of symmetric proximities, it is always possible to construct a set of formal features and feature weights such that the proximities are exactly represented by a contrast model in terms of those features. Moreover, that set of features would not be unique (or stated in technical terms: in general the contrast model is not identifiable). However, several data-analytic methods have been proposed that are based on restricted versions of the contrast model; they can be called additive feature methods (Feger & De Boeck, 1993). Four additive feature methods will now be successively discussed: (1) hierarchical clustering, (2) additive trees, (3) additive clustering and (4) extended clustering. We will conclude this section with a discussion of the interrelations of the four methods.

The four methods will be illustrated with analyses of a common data set. The data were obtained by asking five students to judge the similarity of each pair of animals out of a set of ten animals (one order per pair, chosen at random). The animals, which were selected so as to cover a wide range of kinds, were: eagle, donkey, dog, chicken, cow, elephant, rat, pig, fish, and fly. The ratings were made on a 7-point scale (1=very dissimilar - 7=very similar) and averaged per pair across subjects. The resulting lower triangular proximity matrix is shown in Table 1.

Table 1
Input Proximity Matrix for the Animal Data

	eagle	donkey	dog	chicken	cow	elephant	rat	pig	fish
eagle									
donkey	1.4								
dog	1.6	1.4							
chicken	4.8	2.0	2.0						
cow	1.4	5.4	3.6	2.0					
elephant	1.4	3.4	2.8	1.4	4.2				
rat	2.0	2.4	3.4	2.0	2.2	2.6			
pig	1.4	4.4	4.0	2.2	4.6	3.2	2.4		
fish	1.6	1.0	1.2	1.6	1.0	1.0	1.4	1.4	
fly	2.0	1.0	1.0	1.8	1.0	1.0	1.6	1.2	1.6

Hierarchical Clustering

A hierarchical clustering of a set of objects consists of a hierarchy of nested partitions, indexed by some scale h (Johnson, 1967). The lowest level of the hierarchy (h=0) consists of a partition into singleton clusters and the top level (h=h_{max}) consists of the single cluster of all objects. The nestedness restriction implies that, if $h_1 < h_2$, the partition at level h_1 is a refinement of the partition at level h_2; this means that if two clusters A and B have a nonempty intersection, then either A must be a subset of B or B must be a subset of A. A hierarchical clustering can be graphically represented as a dendrogram. The dendrogram of a hierarchical clustering (obtained with the average linking method) of the data of Table 1 is represented in Figure 1 (for an application with real cognitive data, see, e.g., Vandierendonck, 1984).

A distance δ can be derived from a hierarchical clustering by defining the distance between any two objects as the lowest hierarchical level at which those objects are joined in a cluster; this distance can be converted into a similarity measure by taking $\sigma = h_{max} - \delta$. Assume now that the horizontal segments of the hierarchical clustering are conceived as features (that are present in all objects below them), and that the length of the segments represents a measure *f* of these features; for example, the segment marked with (*) can be interpreted as a feature (or a set of features) that applies to both eagle and chicken, and that is relatively important; -eventually this formal feature can be given a substantive

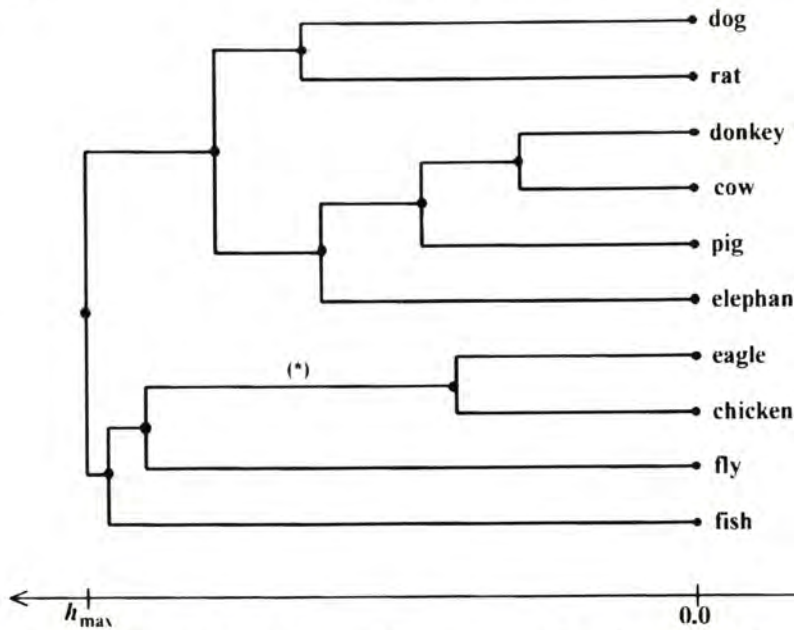

Figure 1. Hierarchical clustering representation of the data of Table 1.

interpretation (has wings, has a beak etc.)-. The similarity σ between two objects a and b derived from the hierarchical clustering can then be expressed in terms of the contrast model as the sum of the measures of the common features of a and b. Otherwise, in a hierarchical clustering the sum of the measures of all features of an object is constant across objects (and equal to h_{max}). Hence, $\sigma(a,b)$ may also be expressed as h_{max} minus 0.5 times the sum of the measures of the distinctive features of a and b.

The distance δ derived from a hierarchical clustering can be shown to be an ultrametric, which satisfies the ultrametric inequality:

$$\forall a,b,c: \delta(a,c) \leq \max[\delta(a,b), \delta(b,c)].$$

The ultrametric inequality implies the triangle inequality. It also implies that if two objects a and b belong to a cluster C, then for any object c that does not belong to C it holds that $\delta(a,c) = \delta(b,c)$; this means that, in terms of similarity with respect to objects in the outside world, the clusters derived from a hierarchical clustering are monothetic.

It can be shown that there exists a one-to-one relationship between hierarchical clusterings and ultrametrics. Hence, it will be possible to make an exact representation of observed dissimilarities d in terms of a hierarchical clustering if and only if those dissimilarities satisfy the metric axioms as well as the ultrametric inequality. If the latter is not the case, one can only look for a hierarchical clustering that approximately represents d, for example in the least

squares sense. Unfortunately, most clustering algorithms do not optimize such a criterion (for exceptions, however, see e.g., Chandon et al., 1980; De Soete 1984).

Additive Trees

A graph consists of a set of points, called vertices, and links between those vertices, called edges. In a weighted graph, each edge has a nonnegative weight. A connected graph without cycles is called a tree; hence, in a tree there is a unique path between any pair of vertices. A rooted tree is a tree in which one nonterminal vertex is given the status of root. A rooted weighted tree representation (obtained with the ADDTREE algorithm) of the similarity data of Table 1 is represented in Figure 2.

Note that the root is represented at the left hand of the figure. The terminal vertices of the tree represent the objects of the data set and the length of a horizontal segment represents the weight of the corresponding edge.

A distance δ can be derived from a weighted tree by defining the distance between any pair of objects as the sum of the weights of the edges that constitute the unique path between those objects; this distance can be converted into a similarity measure by taking $\sigma = \delta_{max} - \delta$ (δ_{max} being the maximum value of δ). Assume now that the edges of the weighted tree are conceived as features (that are present in all objects below them) and that the weight of the edges represents a measure *f* of these features; for example, the segment marked with (*) can be interpreted as a feature (or a set of features) that applies to all four-footed

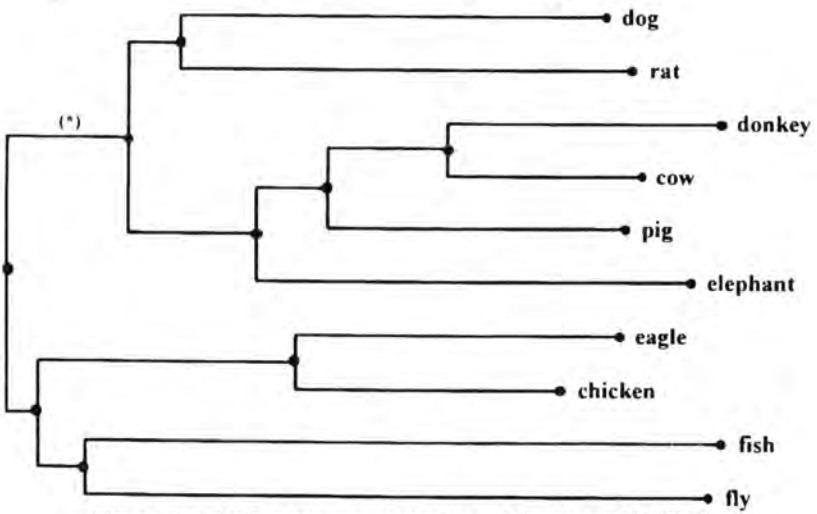

Figure 2. Additive tree representation of the data of Table 1.

animals -optionally, this feature can also be interpreted as a prototype-based category, with objects that are closer to the category edge being more prototypical category members (e.g., dog and cow are more prototypical than donkey and elephant)-. The similarity σ between two objects a and b derived from the weighted tree can be expressed in terms of the contrast model as δ_{max} minus the sum of the measures of their distinctive features.

Note that in the calculation of the tree-distances the root is not involved; hence, the root may be placed at any internal vertex without affecting the tree-distances. Moving the root, in general, implies that a number of features will be replaced by their settheoretical complement (e.g., (*) could be replaced by a feature that applies to all non-four-footed animals). In cognitive terms, this comes down to a change of the *marked* pole of the features in question. Note also that it follows from the preceding that a hierarchical clustering is a special case of an additive tree, namely one for which there exists an internal vertex (-the root of the hierarchical clustering-) which is such that all terminal vertices are equidistant from it.

The distance δ derived from a weighted tree can be shown to satisfy the following inequality, called the four-points condition or additive inequality:

$$\forall a,b,c,d\text{: } [\delta(a,b) + \delta(c,d)] \leq \max\{[\delta(a,c) + \delta(b,d)], [\delta(a,d) + \delta(b,c)]\}.$$

The additive inequality is implied by the ultrametric inequality and implies the triangle inequality. Its substantive interpretation is as follows: By deleting a single edge in a tree, two clusters (i.e., connected subgraphs) are obtained; the additive inequality implies that the average within-cluster distance is smaller than or equal to the average between-cluster distance.

It can be shown that there exists a one-to-one relationship between additive trees and metrics that satisfy the additive inequality. Hence, it will be possible to make an exact representation of observed dissimilarities in terms of an additive tree if and only if those dissimilarities satisfy the metric axioms as well as the additive inequality. If the latter is not the case, one can only look for an additive tree that approximately represents the observed dissimilarities (see, e.g., De Soete, 1983; Guénoche, 1987; Sattath & Tversky, 1977).

Additive Clustering

Shepard and Arabie (1979) proposed a model that assumes that similarities are based on common features only; those features are allowed to be overlapping (non-nested) sets. The model defines the similarity σ between two objects a and b as follows:

$$\sigma(a,b) = w_0 + \sum_{k=1}^{K} w_k p_{ak} p_{bk},$$

where K denotes the total number of features, p_{ak} (resp. p_{bk}) equals 1 if object a (resp. b) has the k-th feature and equals 0 otherwise, and where w_k denotes the measure of the k-th feature (w_0 being an additive constant, i.e., the measure of a feature that applies to all objects). An additive clustering representation (obtained with the MAPCLUS algorithm) of the data of Table 1 is presented in Figure 3.

The similarity σ of the additive clustering model can, in principle, also be defined for identical pairs of objects, and self-similarities $\sigma(a,a)$ can take different values for different objects a; in practical applications

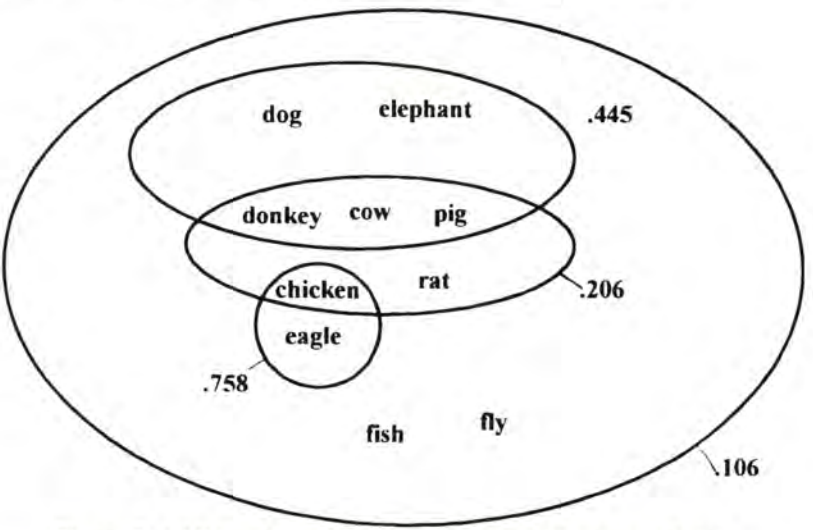

Figure 3. Additive clustering representation of the data of Table 1.

of additive clustering, however, self-similarities are ignored. The similarity σ is symmetric. The additive clustering model further allows that three objects a, b and c are such that a is very similar to b, b very similar to c but a dissimilar to c.

Without constraints on the number of features, any set of observed symmetric similarities can be exactly represented by an additive clustering model. To represent a given data set one can, however, specify a (limited) number of features in advance and look for the set of features and weights that optimally fits the data under that restriction (e.g., in the least squares sense). Algorithms to do this have been developed by Shepard and Arabie (1979), Arabie and Carroll (1980), DeSarbo (1982), Hojo (1983), Sarle (1985), Mirkin (1987), and Chaturvedi and Carroll (1994). Optimal solutions with different numbers of features can then be compared in terms of percentage of variance accounted for to choose the "best" solution.

Extended Trees

Corter and Tversky (1986) proposed a model that represents dissimilarities in terms of distinctive features that are not necessarily nested. The model can be graphically represented in terms of an additive tree that includes marked segments; such marked segments appear in more than one place in the tree and represent nonnested sets of features.

An extended tree (obtained with the EXTREE algorithm) that represents the data of Table 1 is shown in Figure 4. It includes four marked segments. For example, segment C could be interpreted as the conjunction of small, fast and wild; segment H could be interpreted as the conjunction of large and ponderous.

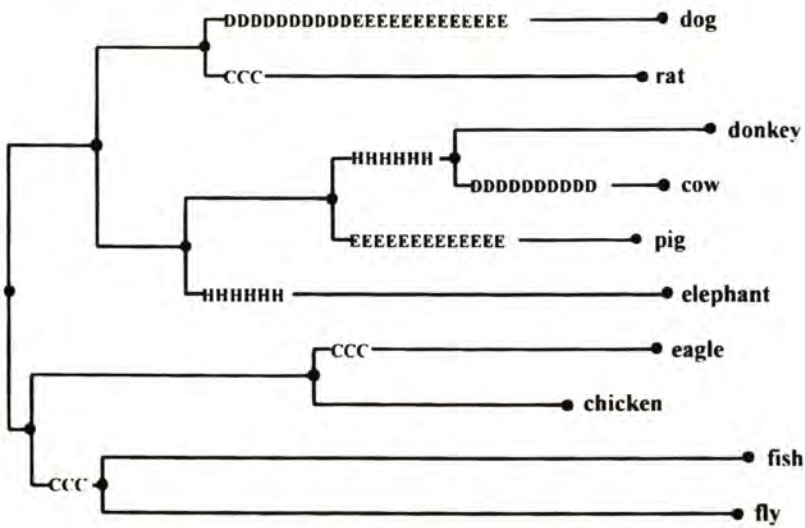

Figure 4. Extended tree representation of the data of Table 1.

A distance δ can be derived from the extended tree model by calculating the path-length distance between a pair of objects; unlike in simple trees, however, marked segments that occur twice on the path between the objects and, hence, represent common features, do not enter into the computation of that distance. δ satisfies the axioms of positive definiteness, symmetry and the triangle inequality. Extended trees can represent any set of distances generated by the so-called distinctive features model; the latter expresses the distance between two objects as the sum of the measures of their distinctive features:

$$D(a,b) = \sum_{x \in A \Delta B} f_x \,, \tag{2}$$

with $A\Delta B = (A \setminus B) \cup (B \setminus A)$ and with f_x denoting the weight of feature x.

It can be shown that any set of symmetric dissimilarities can always be exactly represented (in a trivial way) in terms of the distinctive features model (2), and, hence, in terms of an extended tree. Corter and Tversky (1986), however, devised an algorithm that looks for an "imperfect" extended tree representation of a given data set that consists of a well-fitting additive tree to which a small number of nonnested (marked) features is added.

Interrelations of the Four Methods

If the hierarchical clustering model is interpreted as a common features model, the four above mentioned methods can be ordered in the two-by-two classification presented in Table 2, which was originally proposed by Corter and Tversky (1986). The classifiction is based on the type of similarity function used in the model and the type of feature structure that is assumed. Similarities can be either a function of common features, or a function of distinctive features. The feature structure of the model can be nested or non-nested. As explained above, hierarchical clustering and additive clustering both are common features models, with additive clustering, unlike hierarchical clustering, allowing for overlapping features. Additive and extended trees both are distinctive features models, with extended trees, unlike simple additive trees, allowing for overlapping features.

Sattath and Tversky (1987) have further investigated the relation between common and distinctive features models. Excluding self-(dis)similarities, they found that if a similarity measure s can be exactly represented in terms of a common features model, then there exists a constant K such that K - s can be exactly represented in terms of a distinctive features model and vice versa; in general, the two models can be shown to have the same number of free parameters, although they may involve different sets of features.

Table 2

Corter and Tversky's Classification of Additive Feature Models

	Nested	Non-nested
Common Features	Hierarchical Clustering	Additive Clustering
Distinctive Features	Additive Trees	Extended Trees

From a data-analytic perspective, it may be informative to compare the numbers of free *continuous* parameters involved in the four additive feature methods. If n denotes the total number of objects, the number of free continuous parameters equals n-1 for hierarchical clustering. For additive trees the number is $2n$-3; the difference with the n-1 of hierarchical clustering reflects the removal of the restriction of having an internal vertex (root) that is equidistant from all terminal vertices (objects). For the algorithms based on the additive clustering model the number of free parameters depends on the prespecified number k of common features; taking into account the feature that applies to all objects, the total number of free parameters is k +1. For extended trees the maximum number of free parameters, which also depends on the number of marked segments m, is $2n$-3+m. In comparison with the other methods, the advantage of the additive clustering algorithms clearly is that the user has control over the number of free parameters that are estimated. On the other hand, when fitting additive trees, and especially extended trees, one may run the risk of overfitting the data. This may have been the case in the extended tree analysis of our example.

A graph-theoretic method for the analysis of proximity data in which the number of continuous parameters is under the control of the user has been developed by Klauer and Carroll (1989). The latter authors developed an algorithm that calculates a graph with a user-specified number of edges which is such that the minimum path length distances optimally fit the observed dissimilarities (in the least squares sense). The graphs in question, however, are not trees and cannot be readily linked to Tversky's contrast model.

Discussion

In this section we will first check the four data-analytic methods based on the contrast model against Tversky's criticisms of geometric models of similarity. Second, we will check the same methods against recent cognitive objections to the contrast model.

Check Against Tversky's Criticisms

A first category of Tversky's criticisms referred to the metric axioms. Although the four above mentioned data-analytic methods are based on the contrast model, they do not effectively deal with those criticisms: Except for additive clustering, all methods represent (dis)similarities in terms of a metric. Additive clustering itself involves a model-derived similarity for which the metric axioms can only be checked indirectly: As regards positive definiteness,

the additive clustering model can, in principle, take different values for identical object pairs; in practice, however, algorithms to fit the additive clustering model ignore self-similarities. As regards symmetry, the model-derived similarity is symmetric. As regards the triangle inequality, as indicated above, the additive clustering model may allow for triplets of objects *a*, *b*, and *c* that are such that *a* is very dissimilar to *c*, whereas *a* is very similar to *b* and *b* is very similar to *c*.

With respect to the symmetry axiom, it may further be noted that several data-analytic methods have been proposed that involve asymmetric dissimilarities. In particular, Hutchinson (1989) has proposed a weighted directed graph model from which an asymmetric minimum directed path-length dissimilarity can be derived. Klauer and Carroll (1991) have devised an algorithm to look for the weighted directed graph with a prespecified number of edges that optimally fits observed proximity data (in the least squares sense). On the other hand, DeSarbo (1982) has proposed the GENNCLUS model, which generalizes the model of additive clustering, and which subsumes several asymmetric model variants. Yet, neither the weighted directed graph model nor GENNCLUS can be linked in a straightforward way to the contrast model.

A second category of Tversky's criticisms referred to various phenomena that are problematic for Minkowski metrics. Although the models of our four data-analytic methods are not but special cases of the contrast model, they can deal with several of those phenomena. For example, before we discussed the assumption implied by Minkowski metrics that the dissimilarity of objects which vary along one dimension but are constant on a second dimension must be independent of the specific value taken on the second dimension. This assumption is clearly violated in the additive clustering model. As a second example, Tversky and Gati (1982) discuss in detail the so-called *corner inequality* that violates the conjunction of two implications of Minkowski metrics (viz., the triangle inequality and segmental additivity). From Tversky and Gati's (1982) argument, it can be derived that the extended tree model allows for violations of the corner inequality.

Check Against Recent Cognitive Criticisms of the Contrast Model

The above mentioned data-analytic methods can be looked at from two different viewpoints. First, one may look at them as methods of cluster analysis that enable the user to induce categories from proximity data. Second, one may look at them from the point of view of the contrast model as methods that enable the user to induce the features and the associated measure that underlie observed proximities. From a cognitive psychological perspective, the

assumptions underlying the methods looked at from the two different viewpoints have recently been criticized.

First, several authors including Fried and Holyoak (1984), Murphy and Medin (1985), Keil (1989), and Rips and Collins (1993) have questioned the assumption that similarity provides the basis of categorization; these authors reported several empirical results that point to dissociations between category membership and similarity judgements. A critical reassessment of these results has been formulated by Goldstone (1994).

Within the framework of the present article, we will mainly focus on a second group of criticisms that pertains to the data-analytic methods viewed as methods to model proximities on the basis of specific instantiations of the contrast model. A first subset of those criticisms, raised by authors like Murphy and Medin (1985) reads that similarity is too unconstrained and arbitrary, in that the similarity of two objects can vary from very dissimilar to very similar, depending on the features (and feature weights) that are taken into account. A theory of similarity that does not identify constraints on the features that enter into the computation of similarity must therefore be considered at least very incomplete (see also Thibaut & Schyns 1995). It must be admitted that the contrast model indeed is agnostic with respect to this issue, in that it simply presupposes that similarity is a function of some formal feature sets without further specifying the origin or nature of those feature sets. Within this context, one must however note that the four data-analytic methods described above do not require the user to specify a set of potentially relevant features in advance; instead, relevant features are precisely induced during the data analysis. This may be considered a particular advantage in view of the fact that the set of possible features is infinitely large.

A second subset of criticisms, which is related to the first, reads that similarity judgements may vary depending on the context in which they are made (e.g., Barsalou, 1982). This view, however, is not at odds with the contrast model. In his seminal article, Tversky (1977) emphasized that feature weights as well as the overall weights of common and distinctive features may vary depending on the context of the similarity judgements. In particular, Tversky argued that the weight of features has two components: intensity and diagnosticity, the latter depending on classifications of the objects that may change according to the context. With respect to the data-analytic methods based on the contrast model, one may further note that for two of them extensions have been worked out that can be used to study the context-dependence of similarity judgements in a parsimonious way (Carroll & Arabie, 1983; Carroll, Clark, & DeSarbo, 1984).

A third subset of criticisms concerns the nature of the predicates that are involved in similarity judgements. Gentner and Ratterman (1991) have argued that during development similarity judgments are increasingly based on

correspondences in relational structure rather than on matches or mismatches in features or attributes. As regards the position of the contrast model with respect to this issue, it must be emphasized that the features involved in the model are formal binary variables. The latter variables can be associated with various characteristics of objects that at first sight are not linked to binary variables. As an example, one may refer to Gati and Tversky's (1982) discussion of how formal binary variables can be used to represent attributes that are quantitative or qualitative dimensions. Similarly, various relational characteristics can also be represented in terms of formal binary variables.

A fourth subset of criticisms is based on a series of studies that show that, in a number of cases, people do not combine features in an additive way when making similarity judgements (Goldstone, Medin, & Gentner, 1991). It must first be noted that the contrast model as such does not require the measure of the feature sets, f, to be additive. Furthermore, it must also be emphasized that violations of feature additivity can only be determined, *given a fixed set of features*. That is, a set of features that are nonadditively combined in similarity judgements can always be replaced by an alternative set of (possibly configural) features that are additively combined.

A final set of criticisms reads that similarity assessment is a dynamic context-specific search and alignment process rather than a computation over some feature space (Medin, Goldstone, & Gentner, 1993) (for a somewhat related critique, see Shanon, 1988). When, during such a process, a stimulus B is compared with a target stimulus A, properties that are closely associated with A may have a higher chance to become activated in B. What is more, Medin, Goldstone and Gentner presented evidence that an ambiguous stimulus B may be perceived differently depending on the target stimulus with which it is compared. As regards the position of the contrast model to this issue, it must be admitted that, as a *structural* model, it does not make any claims about possible cognitive (search or alignment) *processes* through which similarity judgements come about. Moreover, insofar the perception of objects B varies depending on the target object A with which they are compared, similarity judgements clearly fall outside the scope of the contrast model. From a formal point of view, such similarities can no longer be considered two-way *one-mode* data. On the contrary, a similarity judgement with respect to the pair (A,B) should be formally treated as a judgement about the pair (A,B_A), with the second object, B_A, being involved in only one data element. Hence, in that case the similarities constitute a (highly incomplete) set of two-way two-mode data. It is impossible to handle such data with inductive data-analytic methods. Otherwise, one may also wonder how a highly pair-specific similarity could play a role in the basic cognitive operations in which similarity is usually assumed to be involved, like stimulus and response generalization during learning.

References

Arabie, P., & Carroll, J. D. (1980). MAPCLUS: A mathematical programming approach to fitting the ADCLUS model. *Psychometrika, 45*, 211-235.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition, 11*, 211-227.

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. *Annual Review of Psychology, 31*, 607-649.

Carroll, J. D., & Arabie, P. (1983). An individual differences generalization of the ADCLUS model and the MAPCLUS algorithm. *Psychometrika, 48*, 157-169.

Carroll, J. D., Clark, L. A., & DeSarbo, W. S. (1984). The representation of three-way proximity data by single and multiple tree structure models. *Journal of Classification, 1*, 25-74.

Chandon, J. L., Lemaire, J., & Pouget, J. (1980). Construction de l'ultramétrique la plus proche d'une dissimilarité au sens des moindres carrés. *Revue Française d'Automatique, d'Informatique et de Recherche Opérationnelle, 14*, 2, 157-170.

Chaturvedi, A., & Carroll, J. D. (1994). An alternating combinatorial optimization approach to fitting the INDCLUS and generalized INDCLUS models. *Journal of Classification, 11*, 155-170.

Corter, J. E., & Tversky, A. (1986). Extended similarity trees. *Psychometrika, 51*, 429-451.

DeSarbo, W. S. (1982). GENNCLUS: New models for general nonhierarchichal clustering analysis. *Psychometrika, 47*, 449-475.

De Soete, G. (1983). A least squares algorithm for fitting additive trees to proximity data. *Psychometrika, 48*, 621-626.

De Soete, G. (1984). A least squares algorithm for fitting an ultrametric tree to a dissimilarity matrix. *Pattern Recognition Letters, 2*, 133-137.

Feger, H., & De Boeck, P. (1993). Categories and concepts: Introduction to data analysis. In I. Van Mechelen, J. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 203-223). London: Academic Press.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 234-257.

Gati, I., & Tversky, A. (1982). Representations of qualitative and quantitative dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 325-340.

Guénoche, A. (1987). Cinq algorithmes d'approximation d'une dissimilarité par des arbres à distances additives. *Mathématiques et Sciences Humaines, 89*, 21-40.

Gentner, D., & Ratterman, J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language & thought: Interrelations in development* (pp. 225-277). Cambridge: Cambridge University Press.

Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition, 52*, 125-157.

Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology, 23*, 222-262.

Hojo, H. (1983). A maximum likelihood method for additive clustering and its applications. *Japanese Psychological Research, 25*, 4, 191-201.

Hutchinson, J. W. (1989). NETSCAL: A network scaling algorithm for nonsymmetric proximity data. *Psychometrika, 54*, 25-51.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika, 32*, 241-254.

Keil, F. (1989). *Concepts, kinds and development.* Cambridge, MA: Bradford Books/ MIT Press

Klauer, K. C., & Carroll, J. D. (1989). A mathematical programming approach to fitting general graphs. *Journal of Classification, 6*, 247-270.

Klauer, K. C., & Carroll, J. D. (1991). A comparison of two approaches to fitting directed graphs to nonsymmetric proximity measures. *Journal of Classification, 8*, 251-268.

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29*, 1-27.

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika, 29*, 115-129.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254-278.

Mirkin, B. G. (1987). Additive clustering and qualitative factor analysis methods for similarity matrices. *Journal of Classification, 4*, 7-31.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289-316.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology, 43*, 25-53.

Pruzansky, S., Tversky, A., & Carroll, J. D. (1982). Spatial versus tree representations of proximity data. *Psychometrika, 47*, 3-19.

Rips, L. J., & Collins, A. (1993). Categories and resemblance. *Journal of Experimental Psychology: General, 122*, 468-486.

Sarle, W. S. (1985). The OVERCLUS procedure. In SAS Users Group International, *Supplemental Library User's Guide* (Version 5, pp. 401-421). Cary: SAS Institute Inc.

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika, 42*, 319-345.

Sattath, S., & Tversky, A. (1987). On the relation between common and distinctive feature models. *Psychological Review, 94*, 16-22.

Shanon, B. (1988). On the similarity of features. *New Ideas in Psychology, 6*, 307-321.

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika, 27*, 125-140.

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika, 27*, 219-246.

Shepard, R. N., & Arabie, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review, 86*, 87-123.

Thibaut, J.-P. & Schyns, P., (1995). The development of feature spaces for similarity and categorization. *Psychologica Belgica*, this issue.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.

Tversky, A., & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review, 2*, 123-154.

Vandierendonck, A. (1984). Concept learning, memory and transfer of dot pattern categories. *Acta Psychologica, 55*, 71-88.

Received April, 1995