Title
Similarity and Taxonomy in Categorization

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 26(26)

ISSN
1069-7977

Authors
Verbeemen, Timothy
Stroms, Gert
Verguts, Tom

Publication Date
2004

Peer reviewed

Similarity and Taxonomy in Categorization

Timothy Verbeemen (timothy.verbeemen@psy.kuleuven.ac.be)
Departement Psychologie, K.U.Leuven, Tiensestraat 102
B-3000 Leuven, Belgium

Gert Storms (gert.storms@psy.kuleuven.ac.be)
Departement Psychologie, K.U.Leuven, Tiensestraat 102
B-3000 Leuven, Belgium

Tom Verguts (tom.verguts@ugent.be)
Vakgroep Experimentele Psychologie, Ghent University, H. Dunantlaan 2
B-9000 Ghent, Belgium

Abstract

In this paper, a two by three approach to modeling categorization is presented. Similarity representations based upon a geometric, an additive tree and an additive cluster model are combined with an exemplar model and a prototype model in a single approach. The six models are applied to the categorization of pictorial known and unknown fruits and vegetables (Smits et al., 2002). For novel stimuli, the geometric exemplar model and the cluster models gave the best account, indicating a strategy where people compare stimuli with stored members on more general continua or a limited set of features. For well-known stimuli, the tree-based models gave the best account of the data, suggesting more elaborate taxonomic knowledge. More generally, the results show that different categorization models may perform better for different sets of stimuli, and that a systematic empirical comparison of such models is needed.

Introduction

A major contribution of categorization research over the last decades has been to establish the relation between similarity and categorization. Rosch and Mervis' (1975) seminal paper on the graded structure of categories showed that categories are ill-defined, and that the extent to which an instance of a category is seen as a typical member is positively related to similarity towards the category in question and inversely related to similarity towards relevant contrast categories (e.g., Verbeemen et al., 2001). Given the importance of similarity in categorization, a formal model should take a clear stance on two issues: The nature of similarity computation and the relevant objects of comparison in this calculation. First, the model must make assumptions about the nature of similarity, especially when the structure of the stimuli under investigation is not experimentally controllable. There are two main approaches to similarity, geometric and feature-based. The geometric approach (e.g., Carroll & Arabie, 1980; Shepard, 1964) represents stimuli in abstract space where similarity is inversely related to the distance between stimuli. In the feature-based approach (e.g. Shepard & Arabie, 1979; Tversky, 1977), similarity is considered a function of feature overlap, where commonalities increase and differences decrease overall similarity. Second, a model should specify the objects used in similarity calculation. In particular, information employed in making category decisions may be stored at the category level, or it may be stored at the level of individual instances of a category. The former approach is known as the prototype view (e.g., Hampton, 1979; Smith & Minda, 1998), and the latter as the exemplar view (e.g., Medin & Schaffer, 1978; Nosofsky, 1986). In this paper, we argue for a systematic evaluation of these formal models in a two by three approach that compares prototype and exemplar models on the one hand, and geometric and feature representations on the other hand.

The Generalized Context Model and a Geometric Prototype Model

In the generalized context model (GCM; Nosofsky, 1984, 1986, 1992), an exemplar model, categorization is assumed to be a function of similarity towards all relevant stored exemplars. In case (physical) dimensions are unavailable, the GCM fitting procedure starts with a multidimensional scaling procedure (MDS; Borg & Groenen, 1997) on proximity measures of all stimuli involved. The coordinates of these stimuli are then used as input for the model. In the case of two categories, *A* and *B*, the probability that stimulus *x* is classified in category *A* is given by:

$$\mathrm{P(A \mid X)} = \frac{\beta_A \eta_{XA}}{\beta_A \eta_{XA} + (1-\beta_A)\eta_{XB}} \tag{1}$$

where β_A lies between 0 and 1 and serves as a response bias parameter towards category *A*. The parameters η_{XA} and η_{XB} denote the similarity measures of stimulus *x* toward all stored exemplars of category *A* and *B*, respectively:

$$\eta_{XA} = \sum_{j \in A} \exp - \left[c \left(\sum_{k=1}^{D} w_k \left| y_{xk} - y_{jk} \right|^r \right)^{1/r} \right] \tag{2}$$

with y_{xk} and y_{jk} as the coordinates of stimulus *x* and the *j*-th stored exemplar of category *A* (or *B* for η_{XB}, respectively) on dimension *k*. The weight of the *k*-th dimension is denoted by w_k, with all weights restricted to sum to 1. The power metric, determined by the value of r, is usually given a value of either 1 or 2, corresponding to city-block and Euclidean distance, respectively.

A prototype model can be constructed with the GCM as a start (Nosofsky, 1986, 1987, 1992; Smits et al., 2002). With the prototype defined as the central tendency of a category (Malt & Johnson, 1992; Malt & Smith, 1984; Rosch & Mervis, 1975), the object created by taking, on each dimension, the average coordinate over all members of the category, is a good way to define a prototype. The similarity function changes to:

$$\eta_{XA} = \exp - \left[c \left(\sum_{k=1}^{D} w_k \left| y_{xk} - \overline{y_{.k}} \right|^r \right)^{1/r} \right] \tag{3}$$

where $\overline{y_{.k}}$ denotes the mean value of all stored members of category A on dimension k. We will refer to (3) in combination with (1) as the Geometric Prototype Model (GPT).

A number of studies have already been conducted that compared prototype and exemplar models (e.g., Nosofsky, 1992; Nosofsky & Zaki, 2002; Smith & Minda, 1998, 2000; Smits et al., 2002). In many, the GCM performed better than prototype models. In the next section we elaborate on the major alternative to geometric similarity models, the contrast model (Tversky, 1977).

The Contrast Model and Categorization

In the contrast model, similarity between two stimuli is defined as a function of the features that these stimuli possess:

$$Sim(a,b) = \theta g(A \cap B) - \alpha f(A - B) - \beta f(B - A) \tag{4}$$

where $g(A \cap B)$ is a function of the features shared by objects a and b (the *common features*), and $f(A-B)$ and $f(B-A)$ are functions of the features that belong to one stimulus but not the other (the *distinctive features*). Different models have been proposed, mostly focused on either the common feature component or the distinctive feature components.

Pruzansky, Tversky and Carroll (1982) reanalyzed 20 data sets taken from various published studies, divided into two groups depending on the hypothesized structure of the stimuli used: conceptual (e.g., vegetables) and perceptual (e.g., polygons) stimuli. For 10 out of 11 studies of conceptual stimuli, analyses of proximity data proved better when performed by ADDTREE, a distinctive features approach to similarity. Seven out of nine studies of perceptual stimuli showed a clear advantage for low-dimensional MDS solutions.

A number of studies have been conducted that compared geometric and featural exemplar models. Lee and Navarro (2001) used additive clustering to extract common features from similarity data and provided excellent accounts of an artificial learning experiment with ALCOVE (Kruschke, 1992). Takane and Shibayama (1992) analyzed identification data of digits taken from Keren and Baggen (1981) and they too obtained excellent results for a featural version of the similarity-choice model (Luce, 1962) based on ADDTREE (Corter, 1982; Sattath & Tversky, 1977). Whereas clustering provides a very flexible way of representing similarity, allowing for overlapping clusters, additive trees are more restrictive in that they impose a hierarchy. There are, however, reasons to apply tree models, especially in the case of conceptual knowledge. A tree model produces, in general, a higher amount of features for the total set than additive clustering. But the amount of shared features is lower in general, and most weight is given to idiosyncratic features. This may be appropriate for well-known stimuli (McCrae & Cree, 2002), as people can be expected to have a fair amount of background knowledge about these stimuli, but it is unclear whether this is plausible in the case of novel stimuli.

The implementation of the feature structure in the GCM yields the *featural exemplar model* with similarities as:

$$\eta_{XA} = \sum_{j \in A} \exp - \left[c \left(\sum_{k=1}^{F} w_k \left(y_{xk}(1 - y_{jk}) + y_{jk}(1 - y_{xk}) \right) \right) \right] \tag{5}$$

where y_{jk} = 1 if stimulus j has feature k and y_{jk} = 0 otherwise. Therefore, the term y_{xk} (1-y_{jk}) is 1 if and only if the target stimulus x possesses the feature and the "reference" stimulus j does not, and vice versa for y_{jk} (1-y_{xk}). Each feature has a weight w_k that corresponds to the length of the segments in the tree. We will refer to this model as GCM-F (generalized context model – featural).

The *featural prototype model* will be illustrated using Figure 1, for an additive tree solution for birds and mammals. Distances between objects are defined by the sum of the horizontal segments on the shortest path between two stimuli (vertical segments are added for visual ease only). Each segment represents a feature that applies to all of its children with more general features closer to the left ("root") and more specific features located towards the right (endpoints) of the tree. The model is again formally similar to the featural GCM with the prototype treated as a pseudo-exemplar. The distance function equals:

$$\eta_{XA} = \exp - \left[c \left(\sum_{k=1}^{F} w_k \left(y_{xk}(1 - y_{pk}) + freq_{pk} y_{pk}(1 - y_{xk}) \right) \right) \right] \tag{6}$$

where y_{pk} is 1 if the prototype of A has the feature, and 0 otherwise. The frequency weighting term corresponds to the relative frequency or proportion with which the feature

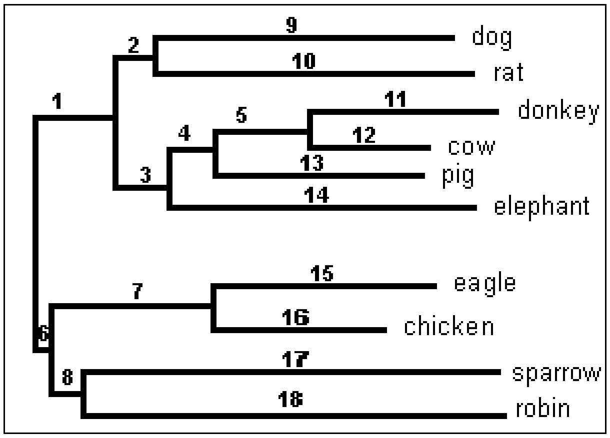

Figure 1: Example of a rooted additive tree.

occurs within A, i.e. the proportion of stored members of category A that possess that particular feature. This corresponds to the idea that the impact of the features in the prototype, which is seen as a pseudo-exemplar, depends on the prevalence of those features in the category[1]. We will refer to this model as FPT (featural prototype model).

The application of the featural models to an additive clustering solution, i.e. a common features model, is straightforward, as the feature structure is defined. This is not the case for additive trees as they produce distinctive features: a feature that adds to the difference between two stimuli may belong to one or the other. This is not a problem for exemplar models as distances between objects remain unchanged, but it will be required for a prototype model. To define a particular structure, one needs to define a root. If the root in Figure 1 is placed anywhere else it would imply that some members of one category possess some of the most general features of the other category but share none of the features belonging to members of their proper category. This is implausible as it would imply that some stimuli are seen as members of a category on a purely idiosyncratic basis and not because they share any features with that category, even though these stimuli would possess the most general features of a related contrast category. Therefore, in the remainder of this article, we will assume that the root is placed on the segment or path that best approximates this linearly separable structure (Medin & Schwanenflugel, 1981) for *stored* (well known) categories[2]. (This implies that one has to decide, *a priori*, which objects are considered to be stored members of a given category, as is the case for all other models. The choice of a root that best approximates separability serves to define the feature structure and not category membership for stored items, which is already determined.)

Analysis of Smits et al.'s data

Smits et al. (2002) analyzed a stimulus set consisting of pictures of 79 well-known items, retained after an exemplar *generation* task for the categories *fruit* and *vegetables*, and 30 fruits or vegetables, mostly exotic, that were completely unknown to participants. Ten participants completed a feature applicability task for all stimuli, for the 17 most frequently generated features for *fruit* and *vegetables,* generated by a different group of thirty participants. (Taking the most frequently generated features ensures that the analysis is not clouded by potentially unreliable features that are important to only a few subjects.) A similarity matrix was then obtained by correlating the feature applicability vectors for all 109 stimuli, after summing over participants. A different group of thirty participants classified the well-known stimuli as belonging to either *fruit* or *vegetables*. A group of twenty different participants did the same for the novel stimuli. Smits et al. then predicted category decisions based on the geometric versions of the GCM and the GPT and found a clear advantage of the GCM over the prototype model. Since their data from the categorization task had a fair amount of variance in the categorization proportions even for the well-known stimuli, it is possible to fit the respective models to old and novel stimuli separately. Therefore, we will analyze the data in a 2 × 3 × 2 framework, where the last factor is added to assess the fit of the models for old and novel stimuli separately.

Generating Similarity Representations

In order to obtain dimensions, similarities between 109 old and novel fruits and vegetables were reanalyzed with ALSCAL (Takane, Young & De Leeuw, 1977), using the BIC criterion (Schwarz, 1978)[3] to determine the optimal

[1] As the similarity structure was derived from the presented stimuli, we assume that a *presented* exemplar has all of its features to the full extent. Because a prototype is a construction after encountering exemplars, we assume that the activation and impact of its features is dependent on the frequency of those features in its own category (Kellogg, 1980). Because the prototype is treated as a pseudo-exemplar, we assume that its features can be no more than fully active (in the case of a feature that applies to all of its members), resulting in a factor of 1. We assumed that the features of *stored* exemplars in the *exemplar model* were not weighted dependant on the frequency of occurrence in *other* exemplars. This was confirmed post hoc: fit values were much worse when weighted for frequency.

[2] ADDTREE starts by grouping together the closest pair of objects, and then creates a dummy object with the average of the distance of the original objects to all other objects. This procedure is repeated until there are three objects left, and the root is placed so as to minimize the variance to the last three objects. Here we minimize the variance on the path that provides the best linear separation for *well-known* stimuli in a similar way.

[3] BIC = -2 ln(L) + k ln(n), where L is the likelihood value (the probability of the data given a certain model), k is the number of free parameters, and n is the number of data points. *Lower* means better, and only the difference in free parameters needs to be taken into account. The first term decreases with increasing model fit, the second is a penalty term that increases with the number of free parameters and data size. As such, the measure is a trade-off between model fit and model complexity. The statistic is most appropriate when the information provided by the data is relatively large as compared to any prior information, as is the case in all the

dimensionality (Lee, 2001b). A three-dimensional solution was chosen that explains 96 percent of the variance.

Following the same procedure for additive clustering, an analysis with ADCLUSGROW (Lee, 2001a) resulted in 32 clusters that explained 96 percent of the variance.

Finally, the same similarity matrix was reanalyzed using ADDTREE/P (Corter, 1982). The explained variance was 84 percent. The algorithm does not readily lend itself to the BIC-guided approach and usually fits to maximum complexity, in this case with 209 arcs (features). To the extent that this may cause the fitting of error, it may cause a drawback for the categorization models as the error would be "plugged" into the model, clouding the explanatory power of the underlying feature structure. At first sight this, and the lower fit value, would indicate that these models are less appropriate.

(It is important to note that these analyses were based on correlation patterns and not on the rough feature vectors, so the similarity algorithms, especially in the featural approach, are in no way restricted to have as much or less features than the original feature vectors.)

Fitting the Similarity Representations to the Categorization Models

The geometric models were fitted with the Euclidean distance metric (r = 2), as this resulted in clearly better fit values. The GCM and GPT were fitted as discussed previously with four free parameters: the bias parameter β, the sensitivity parameter c, and two dimension weights, as weights are restricted to add to 1. Feature weights for the tree- and cluster-based models were taken from the original solutions, however, to keep the number of parameters feasible for estimation, hence there are two free parameters, β and c. Stored members are the same in all models and are based on the earlier exemplar generation task. (Note that the tree-based models were based on the original ADDTREE solution after placing the root so as to provide the best linear separation, for known stimuli, between fruit and vegetables. Compared to the actual *generation* task for well-known stimuli, only one item, rhubarb, was generated as a fruit but closer to vegetables according to the ADDTREE solution.[4])

Results and Discussion All models were fitted by maximizing the binomial likelihood. Correlations between predicted and observed category decisions ranged from .85 to .93 with the best performing models $\geq$ .92, indicating a fair but not perfect amount of explained variance. The models were evaluated using BIC. Results are summarized in Table 1.

Analyses of the 30 novel stimuli separately are presented in the first panel of Table 1. The lowest (best) BIC value was obtained for the GCM, but the difference with the cluster-based exemplar and prototype models is small. All ADDTREE-based models performed clearly worse. The current results do not clearly favor exemplar or prototype models for novel stimuli, but it appears that the geometric approach to prototypes provides less explanatory power as compared to the clustering approach.

Table 1: -ln(L)[5] and BIC (only the difference in parameters taken into account) for the category *fruit* for all models.

	MDS		**ADCLUSGROW**		**ADDTREE/P**	
	GCM	GPT	GCM-F	FPT	GCM-F	FPT
1.*New*						
-ln(L)	73.95	83.27	81.38	82.12	90.83	93.01
BIC	160.69	179.33	162.76	164.24	181.66	186.02
2.*Well-known*						
-ln(L)	351.75	405.29	364.87	388.51	334.75	330.18
BIC	719.04	826.12	729.74	777.02	669.50	660.36

Analyses of the well-known stimuli resulted in a very different pattern. BIC values for the analyses of the 79 well-known stimuli separately are presented in the second panel of Table 1. The BIC values for the geometric and the cluster-based models were clearly higher (worse) than for the tree models. Clearly, the data from the well-known subset is best accounted for by the tree-based models that assume more elaborate taxonomic knowledge. The difference between the exemplar and prototype model is rather small and should be interpreted with caution. This result appears to contradict the earlier fit values where the MDS and additive clustering solutions provided a substantially better fit to the *similarity* data. In fact, a better fit to similarity data need not imply a better fit of the categorization model: those aspects of stimuli that are activated in a similarity task may very well be different from what is activated in a categorization task, especially after a concept has become well-elaborated. In other words, the less flexible and hierarchic structure of trees may not have captured all aspects of similarity, but the aspects it did capture may be more relevant for categorization of well-known concepts. Indeed, every aspect of similarity that is not used in categorization can be considered error in the model.

In fact, the most interesting pattern that emerges from these data is the fact that categorization of novel stimuli is best explained by those models that are based on the flexible representations that best explain similarity. These models have either a limited number of dimensions or a limited number of features, with little idiosyncratic features in the

analyses presented here. It is also fit to compare nonnested models. (For an extensive discussion, see Kass & Raftery, 1995.)

4 In the actual classification task (not the *generation* task), the proportion of classifications for *rhubarb* as a fruit was only .33.

5 This value is the most "democratic" measure as it only incorporates model fit, (incorrectly) disregarding the penalty term for free parameters. The measure is equal to the sum of minus the log likelihoods of the individual data points and is therefore sensitive to the size of the data set; hence differences in fit *between* the two datasets are *not* directly interpretable (the same is true for the BIC measure).

latter case. Categorization of well-known stimuli, on the other hand, is best explained by the models that use a representation that is less close to the similarity data but that impose a more elaborate taxonomy and more idiosyncratic features.

An interesting interpretative property of additive trees in this respect is the fact that, at each node of the tree, a feature that applies to all of its children is linked to a limited number of alternatives. Second, branches in the tree tend to have a higher weight as one goes down in the tree. This implies that the number and weight of commonalities decreases with the number of nodes between stimuli. It also means that those features that add most weight to the difference are less likely to be related as the number of nodes increases and vice versa. A similar argument was made by Markman & Gentner (1993) who presented stimuli with different ontological distances and found a similar pattern when subjects listed commonalities and alignable and nonalignable differences. A possible explanation for the good results of tree-based models could be that, as a concept becomes more elaborated, people tend to gravitate to an alignable structure that might dominate other, presumably less alignable, aspects of similarity.

Conclusion

The goal of the present paper was two-fold. First we presented a general framework, in which different models (i.e., exemplar and prototype models, embedded in either dimensional or featural similarity representations) could be systematically formulated, compared and tested. Given the framework, one can investigate precisely in what situations which model aspects perform best. Second, the framework was applied to categorization data of well-known and novel stimuli in the context of familiar natural language concepts. The results indicate that, depending on the amount of knowledge and mastery of the stimuli, different representational structures and different decision processes may operate.

One may wonder how these results relate to the findings from the category learning literature (e.g., Nosofsky, 1992; Smith & Minda, 2000; Stanton, Nosofsky & Zaki, 2002). In most of these studies, exemplar models embedded in multidimensional representations have been shown to account very well for the categorization data. However, in these studies, artificial categories are used almost invariably, with stimuli that vary along a limited number of salient dimensions. Formal models, such as the ones described in our paper, have seldom been applied to natural language concepts, which are far more complex than the stimuli used in the artificial category literature, and of which our participants arguably have a much richer and more elaborate knowledge than even the best trained participants have of artificial stimuli. (For other attempts to apply formal models to natural language concepts, see Bailey & Hahn, 2001; Smits et al., 2002; Storms, De Boeck, & Ruts, 2000, 2001; Verbeemen et al., 2001.) However, in spite of participants' extensive knowledge of such concepts, determining the relevant underlying dimensions or features for categorization with natural language concepts is perhaps the most crucial problem in modeling natural language categories (see, e.g., Murphy & Medin, 1985). The two by three framework that was presented here may serve as a valuable tool in this endeavor.

Acknowledgments

The first author is a research assistant of the Fund for Scientific Research – Flanders. This project was in part sponsored by grant OT/01/15 of the University of Leuven research council to Gert Storms.

References

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language*, **44**, 568-591.

Borg, I., & Groenen, P. (1997). *Modern Multidimensional Scaling: Theory and Applications.* Springer-Verlag, New York.

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

Corter, J. E. (1982). ADDTREE/P: A PASCAL program for fitting additive trees based on Sattath & Tversky's ADDTREE algorithm. *Behavior Research Methods & Instrumentation*, **14**, 353-354.

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, **18**, 441-461.

Kass, R. E. , & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, **90**, 773-795.

Kellogg, R. T. (1980). Simple feature frequency versus feature validity models of formation of prototypes. *Perceptual and Motor Skills*, **51**, 295-306.

Keren, G., & Baggen, S. (1981). Recognition of alphanumeric characters. *Perception and Psychophysics*, **23**, 234-246.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.

Lee, M.D. (2001a). On the complexity of additive clustering models. *Journal of Mathematical Psychology,* **45**, 131-148.

Lee, M. D. (2001b). Determining the dimensionality of multidimensional scaling representations for cognitive modeling. *Journal of Mathematical Psychology*, **45**, 149-166.

Lee, M. D., & Navarro, D. J. (2002). Extending the ALCOVE model of category learning to featural stimulus domains. *Psychonomic Bulletin and Review*, **9**, 43-58.

Luce, R. D. (1962). An observable property equivalent to a choice model for discrimination experiments. *Psychometrika*, **27**, 163-167.

Malt, B. C., & Johnson, E. C. (1992). Do artifact concepts have cores? *Journal of Memory and Language*, **31**, 195-217.

Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, **23**, 250-269.

Markman, A. B., & Gentner, D. (1993). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language*, **32**, 517-535.

McRae, K., & Cree, G. S. (2002). Factors underlying category-specific semantic deficits. In E. M. E. Forde & G. W. Humphreys (Eds.), *Category-specificity in brain and mind* (pp. 211-249). East Sussex, England: Psychology Press.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, **7**, 355-368.

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

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **10**, 104-114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **13**, 87-108.

Nosofsky, R. M. (1992). Exemplars, prototypes and similarity rules. In A F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning theory to connectionist theory: Essays in honour of William K. Estes*, Vol. 1. Lawrence Erlbaum, Hillsdale, NJ.

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **28**, 924-940.

Pruzansky, S., Tversky, A., & Carroll, J. D. (1982). Spatial versus tree representations of proximity data. *Psychometrika*, **47**, 3-24.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, **7**, 573-605.

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

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461-464.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, **1**, 54-87.

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

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **24**, 1411-1436.

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **26**, 3-27.

Smits, T., Storms, G., Rosseel, Y., & De Boeck, P. (2002). Fruits and vegetables categorized: An application of the generalized context model. *Psychonomic Bulletin and Review*, **9**, 836-844.

Stanton, R. D., Nosofsky, R. M., & Zaki, S. R. (2002). Comparisons between exemplar similarity and mixed prototype models using a linearly separable category structure. *Memory and Cognition*, **30**, 934-944.

Storms, G., De Boeck, P., & Ruts, W. (2000). Prototype and exemplar-based information in natural language categories. *Journal of Memory and Language*, **42**, 51-73.

Storms, G., De Boeck, P., & Ruts, W. (2001). Categorization of unknown stimuli in well-known natural language concepts: a case study. *Psychonomic Bulletin and Review*, **8**, 377-384.

Takane, Y., & Shibayama, T. (1992). Structures in stimulus identification data. In F. G. Ashby (Ed), *Multidimensional models of perception and cognition*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Takane, Y., Young, F. W., & De Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, **42**, 7-67.

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

Verbeemen, T., Vanoverberghe, V., Storms, G., & Ruts, W. (2001). The role of contrast categories in natural language concepts. *Journal of Memory and Language*, **44**, 618-643.