To cite this chapter:

Vanpaemel, W., Storms, G., & Ons, B. (2005). A varying abstraction model for categorization. In B. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual conference of the cognitive science society* (pp. 2277-2282). Lawrence Erlbaum.

A Varying Abstraction Model for Categorization

Wolf Vanpaemel (wolf.vanpaemel@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

Bart Ons (bart.ons@psy.kuleuven.ac.be)
Departement Psychologie, K.U.Leuven, Tiensestraat 102
B-3000 Leuven, Belgium

Abstract

A model is proposed that elegantly unifies the traditional exemplar and prototype models. These two models are extreme cases of the proposed varying abstraction model. The unifying model further makes room for many new intermediate pseudo-exemplar models. A preliminary analysis using Medin and Schaffer's (1978) 5-4 structure pointed to such an intermediate model that outperformed the prototype and exemplar models.

Introduction

Formal models of categorization can roughly be divided into two groups: exemplar models and prototype models. What are they and what are they worth?

Description

Both models share their representation and response selection assumptions. As far as representation is concerned, both models assume that the percept associated with a single exposure to a certain stimulus can be represented as a fixed point in a D-dimensional space. In case the coordinates of the positions of the stimuli are not predefined by the experimenter, they can be obtained by for instance multidimensional scaling (Borg & Groenen, 1997; Lee, 2001). The response selection assumption states that the probability that a stimulus i is categorized in category A is computed from the similarity choice model, as expressed in equation (1).

The models differ in their retrieval assumption or, simply put, in how a category is represented. Exemplar models start from the assumption that memory traces of individual exemplars are stored, without any abstraction across these stored exemplars. Prototype models on the other hand, assume that categories are stored as an abstract summary representation.[1]

Performance

Over the last 2 decades, many researchers have compared exemplar and prototype models. In numerous category learning studies with artificial stimuli (e.g., Medin, Altom & Murphy, 1984; Minda & Smith, 2001; Nosofsky, 1992; Smith & Minda, 2000, 2002), exemplar and prototype models have been contrasted and compared in relation to both between-category structure (e.g., categorization decisions) and within-category structure (e.g., typicality ratings). In the majority of these studies (e.g., Medin & Schaffer, 1978; Medin, Altom & Murphy, 1984; Nosofsky, 1992), different versions of the exemplar model outperformed the prototype model.

In contrast, many studies within the domain of natural language have proposed a prototype view of semantic concept representation (e.g., Hampton, 1993; Lakoff, 1987; Rosch, 1978). Smits, Storms, Rosseel and De Boeck (2002), however, applied formal models developed in the context of artificial category learning experiments to data from natural language concepts. They found that, also in the field of natural language, the exemplar model outperformed the prototype model.

A unifying idea

In all of the above mentioned studies, exemplar and prototype models are tacitly presented as the only two alternatives in a dichotomy. Conceptually and formally however, both models can be considered as extremes on a continuum that varies from no abstraction at all (i.e., every exemplar is represented separately) to maximum abstraction (i.e., a single prototype as a summary representation). Along this continuum, positions in-between these two extremes are held by new models, in which exemplars cluster together.

We present here a generic model that allows varying levels of abstraction. This varying abstraction model has two merits. First, it unifies the traditional exemplar and prototype models. Both models are reduced to special (extreme) cases of the generic model. Second, the varying abstraction model also allows intermediate levels of abstraction, in which exemplars are not stored separately, but merge into pseudo-exemplars. These intermediate abstraction levels define a set of new models, called pseudo-exemplar models.

A unifying model

Developing a varying abstraction model within the framework of the very successful generalized context model (GCM; Nosofsky, 1984, 1986) is rather straightforward.

[1] A far more elaborate comparison between exemplar and prototype models (as well as decision bound models) can be found in Ashby and Maddox (1993).

GCM framework

Probability The traditional GCM states that, in a categorization task with two categories A and B, the probability of responding A given a stimulus i equals

$$P(A,i) = \frac{\beta_A \eta_{iA}}{\beta_A \eta_{iA} + (1-\beta_A)\eta_{iB}}, \quad (1)$$

where β_A is the response bias towards category A and η_{iJ} is the similarity of the stimulus i to the category J.

Similarity A necessary condition to be able to calculate the similarity of the stimulus i to the category J is a clear specification of what makes up a category. It is exactly at this specification that exemplar and prototype models diverge, as will become clear immediately. They do agree however on the definition of the similarity of the stimulus i to another stimulus j. It is assumed to be related to (psychological) distance d via

$$\eta_{ij} = \exp(-d_{ij}^{\alpha}). \quad (2)$$

Two special cases are popular: the one where $\alpha = 1$, resulting in the exponential decay function, and the one where $\alpha = 2$, resulting in the Gaussian function.

Distance The distance between the two stimuli i and j in turn is calculated from the coordinates of the two stimuli. There are several versions of how to compute a distance from the coordinates. The most common expression for distance is

$$d_{ij} = c[\sum_{k=1}^{D} w_k |x_{ik} - x_{jk}|^r]^{1/r}, \quad (3)$$

where x_{ik} is the coordinate of stimulus i on dimension k. This is a so-called weighted distance: w_k denotes the proportion of attention allocated to dimension k and so $\sum_{k=1}^{D} w_k = 1$. The parameter c is a scaling parameter. This distance is called a city-block distance when $r = 1$ and Euclidean when $r = 2$.

Exemplar and prototype models

Both prototype and exemplar models share the above assumptions and definitions. They differ however in their exact understanding of what makes up 'the category J'. Therefore, they differ in the way the similarity of the stimulus i to the category J is calculated.

Exemplar model In the exemplar model, a category is assumed to be represented by memory traces of all the encountered exemplars of the category. Hence, the similarity of the stimulus i to the category J is calculated by summing the similarity of the stimulus i to all N stored exemplars of J, that is:

$$\eta_{iJ} \equiv \sum_{j \in J}^{(N)} \eta_{ij}, \quad (4)$$

where N is the number of exemplars of J. Using equations (2) and (3), we find that

$$\begin{aligned} \eta_{iJ} &= \sum_{j \in J}^{(N)} \exp(-d_{ij}^{\alpha}) \\ &= \sum_{j \in J}^{(N)} \exp(-(c[\sum_{k=1}^{D} w_k |x_{ik} - x_{jk}|^r]^{1/r})^{\alpha}). \end{aligned} \quad (5)$$

Prototype model A prototype model can easily be formulated within the framework of the GCM. In this model, a category is assumed to be represented as an abstract summary representation of all the encountered exemplars of the category. This abstract summary of the category is called the category prototype and is denoted as P_J. Hence, the similarity of the stimulus i to the category J equals the similarity of the stimulus i to the category prototype, that is:

$$\eta_{iJ} \equiv \eta_{iP_J}. \quad (6)$$

The coordinates of P_J are simply the averaged coordinates of all the exemplars within the category J on each of the underlying coordinate axes:

$$x_{P_J k} = \frac{1}{N} \sum_{j \in J}^{(N)} x_{jk}. \quad (7)$$

Using equations (2), (3) and (7), we find that

$$\begin{aligned} \eta_{iJ} &= \exp(-d_{iP_J}^{\alpha}) \\ &= \exp(-(c[\sum_{k=1}^{D} w_k |x_{ik} - \frac{1}{N} \sum_{j \in J}^{(N)} x_{jk}|^r]^{1/r})^{\alpha}). \end{aligned} \quad (8)$$

The unifying model

Principle In the above presentation of the exemplar and prototype models within the GCM framework, the extreme positions of these models are easily seen: in the exemplar model, the distance of i towards all N exemplars is calculated, while in the prototype model, the distance of i towards just one single exemplar (i.e., the category centroid) is calculated. The varying abstraction model now assumes that the number of items to which i is compared can vary, in principle, from 1 to N. This means that the varying abstraction model not only incorporates the two traditional models, but also invokes new intermediate models.

Formalization One can consider a category J as a set of N elements j_n.[2] The basic idea is to make up a partition for each category. A set partition, or simply a partition, of a set S is defined as a collection of disjoint, nonempty subsets of S whose union is S. Each subset in such a partition is called an equivalence class or a block. For every block one can easily construct the

[2] To simplify the notation, the index n is dropped most of the time.

prototype of this block by averaging over all the exemplars in that block. Such a block prototype is called a pseudo-exemplar.[3]

In general, a partition of J consists of Q different blocks F_q, where q ranges from 1 to Q. Q itself ranges from 1 (when F equals J, the set of all exemplars) to N (when every F_q is a singleton). The number of elements in a block F_q is denoted as R_q. The block prototype in a block F_q is denoted as e_q.[4] The set of all the block prototypes of a certain partition of J, is denoted as E and has a set size of Q.

How is a category defined according to the varying abstraction model? In the model, it is assumed that a category is represented by *a number* of abstract summary representations of *some* of the encountered exemplars of the category. These abstract summaries of the category are called the category pseudo-exemplars and are denoted as e_q. Hence, the similarity of the stimulus i to the category J is calculated by summing the similarity of the stimulus i to all Q pseudo-exemplars of J, that is:

$$\eta_{iJ} \equiv \sum_{e \in E}^{(Q)} \eta_{ie}. \tag{9}$$

The coordinates of e_q are simply the averaged coordinates of all the R_q exemplars within the block F_q on each of the underlying coordinate axes:

$$x_{ek} = \frac{1}{R_q} \sum_{j \in F_q}^{(R_q)} x_{jk}. \tag{10}$$

Making use of equations (2), (3) and (10), we finally find that

$$\begin{aligned} \eta_{iJ} &= \sum_{e \in E}^{(Q)} \exp(-d_{ie}^{\alpha}) \\ &= \sum_{e \in E}^{(Q)} \exp(-(c[\sum_{k=1}^{D} w_k |x_{ik} - x_{ek}|^r]^{1/r})^{\alpha}) \\ &= \sum_{e \in E}^{(Q)} \exp(-(c[\sum_{k=1}^{D} w_k |x_{ik} - \frac{1}{R_q} \sum_{j \in F_q}^{(R_q)} x_{jk}|^r]^{1/r})^{\alpha}). \end{aligned} \tag{11}$$

Model family One specific partition of each category picks out one specific model. Such a model is called a pseudo-exemplar model. The two extreme partitions (i.e. $Q = N$ for each category and $Q = 1$ for each category) correspond to the two extreme pseudo-exemplar models (i.e. the exemplar model and the prototype model, respectively). The other, intermediate partitions correspond to intermediate pseudo-exemplar models. Therefore, the varying abstraction model is not just a model but it is a family of models.

[3]Alternative names for pseudo-exemplar are super-exemplar or sub-prototype. They are all used interchangeably.

[4]To simplify the notation, the index q is dropped most of the time.

A key role is played by Q: it denotes the number of blocks and hence the number of block prototypes (or pseudo-exemplars) that are used to represent a category. On a conceptual level, Q indexes the level of abstraction (higher Q means lesser abstraction). Since Q is allowed to vary from 1 to N for each category and Q counts the number of prototypes used to represent a category, the varying abstraction model is a formalization of the idea that people use multiple prototypes.

The varying abstraction model makes clear that the exemplar and prototype model are extreme pseudo-exemplar models and are therefore of the same nature as the intermediate pseudo-exemplar models. However, for ease of exposition, the term pseudo-exemplar model will only be used to refer to the intermediate models, not to the extreme ones.

The traditional models Formally, the expressions (4) and (6) can be considered as special cases of the general expression (9). The same holds for expressions (5) and (8): they are special cases of expression (11). The varying abstraction model reduces to the traditional models when a specific partition is chosen for each category.

The exemplar model follows when each category J with N elements is partitioned in N subsets of one element each. The block prototypes therefore equal the exemplars. More formally:

- $Q = N$
- $F_q = \{j_q\}$ so $R_q = 1$ for every q
- $E = J$.

The prototype model follows when each category J with N elements is partitioned in only one subset of N elements. There is only one block prototype which equals the category prototype P_J. More formally:

- $Q = 1$
- $F_q = J$ so $R_q = N$ for every q
- $E = \{P_J\}$.

The pseudo-exemplar models Every partition defines or corresponds to a certain model. How many non-extreme (i.e. $1 < Q < N$) partitions (and hence models) should one consider when fitting the model to a data set?

In the ideal case, one might explore all possible partitions of the N stimuli in a category J. However, this strategy is not always feasible. The number of possible partitions of a set of N elements is given by the Bell number of N (denoted as B_N). This number increases very rapidly.[5] In a categorization task with two categories A and B with N_A and N_B exemplars respectively, this implies fitting $B_{N_A} \times B_{N_B}$ different models. When more than, say, eight, stimuli per category are used, the number of possibilities to consider becomes intractable.

[5]For instance, B_6 equals 203 and B_{10} equals 115975.

A straightforward way to keep the number of partitions within practically feasible boundaries is to select at every level of abstraction only one partition. Instead of blindly considering all partitions, we could limit ourselves to only one partition for each category for every value of Q. This partition can for example be selected through clustering. This reduces the number of different models to be fit to $N_A \times N_B$. An application of this approach in the context of natural language can be found in Verbeemen, Storms and Verguts (in press).

When practically possible, fitting all possible models is a fruitful strategy. It has the clear advantage over the clustering approach that no assumptions have to be made about spatial representation or about which exemplars should or should not be merged together. What drives clustering is inferred, not imposed.

The main advantage of the pseudo-exemplar models is that they allow for sensitive modeling. They allow for adaptation to category complexity, category experience and individual skills.

Evaluating the model

A first data set of a category learning experiment with artificial stimuli has been analyzed using the varying abstraction model and corresponding Matlab algorithms. This preliminary analysis indicated that a pseudo-exemplar model outperformed the traditional prototype and exemplar models.

Matlab fitting algorithms

Determining to what extent the model can account for category-related behavior is done by fitting all the models of the varying abstraction model family to a data set.

A typical categorization experiment consists of a training phase and a test phase. In the test phase, N stimuli are presented to S subjects. Each subject classifies every stimulus as either A or B. In the following, $P(A, i \mid \theta)$ is the varying abstraction model's estimate of the probability of responding A to i, given the parameters θ. The expression for this probability is obtained by combining expressions (1) and (11).

We have developed two Matlab algorithms to fit prototype, exemplar and all pseudo-exemplar models and compare the performance of the different models. Both least squares and maximum likelihood methods are used to estimate the model's unknown parameters.[6] These parameters are the response bias β_A, the scaling parameter c and $D-1$ attention weights w_k. All these parameters are, for the sake of brevity, summarized in the parameter vector θ.

The least squares algorithm looks for the θ that *most accurately* describes the observed responses. The algorithm seeks those parameter values that minimize the sum of squared errors, that is

$$SSE(\theta) = \sum_{n=1}^{N} (p_n - P(A, n \mid \theta))^2, \qquad (12)$$

where p_n is proportion of A responses for stimulus n.

The maximum likelihood algorithm looks for the θ that *most likely* have produced the observed responses. Therefore, it should seek those parameter values that maximize this likelihood. Assuming a binomial distribution, this likelihood equals

$$Lik(\theta) = \prod_{n=1}^{N} \binom{S}{y_n} P(A, n \mid \theta)^{y_n} P(B, n \mid \theta)^{S-y_n}, \quad (13)$$

where y_n denotes number of subjects choosing category A for stimulus n. For computational efficiency, it is better to look for the θ that maximizes the natural logarithm of this likelihood. Hence the function to be maximized in the maximal likelihood algorithm is the loglikelihood[7]:

$$Loglik = \sum_{n=1}^{N} [y_n \ln P(A, n \mid \theta) + (S - y_n) \ln P(B, n \mid \theta)]. \qquad (14)$$

When fitting the varying abstraction model to a data set, the algorithm seeks, among all possible partitions of each category, the parameter values that minimize SSE or maximize $Loglik$. The partition yielding the smallest minimal SSE or largest maximal $Loglik$ of all the possible partitions corresponds to the pseudo-exemplar model that best accounts for the categorization process.[8]

Category learning experiment

The proposed model's performance was tested in a categorization experiment using the well-known 5-4 structure (Medin & Schaffer, 1978; Smith & Minda, 2000).

Subjects Twenty-four first year university students participated for course credit.

Stimuli Stimuli were constructed according to the 5-4 category structure from Medin and Schaffer (1978). The stimuli were artificial sheep, varying on four dimensions: eyes (open or closed), fleece (four or five curls), feet (black or white) and tail (rounded or starred). During the training phase, subjects only encountered the five stimuli of category A and the four stimuli of category B. During the transfer phase, all 16 stimuli were presented.

Procedure To motivate the subjects, the categorization task was presented as a sheepdog game. The participants were asked to drive the sheep to the correct meadow. A-sheep should be driven to the left and B-sheep to the right. In each trial, a sheep appeared on

[6] A very useful tutorial on maximum likelihood estimation is Myung (2003).

[7] The additive constant $\sum_{n=1}^{N} [\ln S! - \ln(S - y_n)! - \ln y_n!]$ does not depend on θ and therefore is dropped.

[8] Both measures only take descriptive accuracy into account. To avoid overfitting, model complexity should be taken into account as well. Although all the models of the varying abstraction model family have the same number of parameters with the same range, the functional form of these parameters differ (i.e. the way the parameters are combined). There are tools available combining goodness-of-fit with model complexity, such as Bayes factors and minimum description length (Myung & Pitt, 1997; Pitt, Myung & Zhang, 2002).

the screen, staring at a certain direction. The participant had to evaluate the correctness of the initial staring direction, by pressing button 1 for a correct and button 2 for a wrong direction. Each sheep appeared twice, once in the correct direction and once in the incorrect direction. Hence there were 18 trials in the training phase and 32 trials in the transfer phase. The order of appearance of the sheep was randomized. In the training phase, every trial was followed by "good" or "false". Feedback was omitted in the transfer phase.

Results Both categories are small, so all possible partitions/models could be examined. Since there are 16 possible partitions for a set of four elements and 52 for a set of five, 780 different models were fit to the data. In principle, four different families of models could be fitted, since both r and α can take the values 1 or 2. In the preliminary analysis presented here, both r and α were set to 1. The summary fits are presented in Table 1.

Table 1: Summary fits of the prototype (PM), exemplar (EM) and the best pseudo-exemplar (PE) model under the least squares estimation method.

	PM	EM	PE
minimal SSE	0.13	0.09	0.03
r^2	0.93	0.95	0.99
parameter β_A	0.58	0.52	0.09
parameter c	7.53	6.88	11.31
parameter w_1	0.24	0.20	0.17
parameter w_2	0.00	0.10	0.14
parameter w_3	0.37	0.47	0.24
parameter w_4	0.39	0.24	0.44

Discussion This preliminary analysis with small categories revealed that, as expected, the exemplar model outperformed the prototype model. More importantly, it also revealed that a pseudo-exemplar model outperformed both traditional models.[9] The best pseudo-exemplar model corresponds to the following partition of the categories, using the labels as described in Medin and Schaffer (1978): category A is divided in three clusters ($A1$ and $A4$; $A2$; $A3$ and $A5$) and category B remains one simple cluster. It is not surprising at all that $A2$ is singled out since $A2$ is the stimulus that is the least similar to the prototype of category A.

General Discussion

The model proposed in this paper elegantly unifies two traditional formal models for categorization. Further, it gives rise to a set of new models, called pseudo-exemplar models. All these models, traditional and new, are formalized along the continuum of abstraction.

[9]In fact, there were several pseudo-exemplar models that outperformed the traditional models. Only the best one is reported here.

Model performance

Analysis of a categorization experiment indicated that a pseudo-exemplar model outperformed the traditional models. It is important to note that this finding does not lead to the conclusion that the varying abstraction model outperformed the exemplar and prototype models. In fact, we would be able to make that conclusion even before having a look at any data set at all. For, the varying abstraction model includes both traditional models, so its performance is at least as good as the performance of the traditional models. Of course, this comparison would not be a fair one. We do conclude however that the varying abstraction model can single out a new pseudo-exemplar model (or several new pseudo-exemplar models), that clearly outperforms the traditional ones. The comparison between this specific pseudo-exemplar model and the traditional models is a fair one, or at least as fair as the comparison between the exemplar model and the prototype model.

Future Directions

Due to the pseudo-exemplar models, the varying abstraction model allows for sensitive modeling. This sensitivity makes the varying abstraction model highly useful for investigating in full detail Smith and Minda's (1998) findings that category representation may change during the learning process. In fact, they found that with large categories, a prototype model yields better accounts of the initial phases of categorization, while exemplar models yield better accounts in later stages of the learning process. One can expect that, using the varying abstraction model, better fits will be obtained for models with more pseudo-exemplars as learning proceeds.

Apart from doing more elaborate modeling of categorization decisions, the varying abstraction model can easily be extended to account for typicality ratings. Nosofsky (1988) contrasted prototype and exemplar accounts of rated typicalities. Further research will investigate whether intermediate abstraction levels account better for rated typicalities than extreme abstraction levels do.

Another extension could be the account of response times for exemplars in a speeded categorization task, for instance along the lines of Nosofsky and Palmeri's (1997) exemplar-based random walk model.

A theoretical issue is the connection between the proposed varying abstraction model and related models. The idea of considering the prototype and exemplar models as extremes along a continuum has been adopted in other models. The rational model (Anderson, 1991), SUSTAIN (Love, Medin & Gureckis, 2004) and the mixture model (Rosseel, 2002) share with the varying abstraction model the idea that a category is represented by multiple prototypes, but show clear differences with the varying abstraction model too. The varying abstraction model is also closely related to a general-rule-plus-exception model such as RULEX (Nosofsky, Palmeri & McKinley, 1994). Investigating exactly how these models and the varying abstraction model differ in principle, formalization and performance is important work for the future.

Acknowledgments

The research reported here is part of an interdisciplinary research project sponsored by the Research Council of the University of Leuven (IDO/02/004), given to the second author. The authors wish to thank Michael D. Lee and Daniel Navarro.

References

Anderson, J. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*, 372–400.

Borg, I., & Groenen, P. J. F. (1997). *Modern multidimensional scaling*. New York: Springer.

Hampton, J. A. (1993). Prototype models of concept representations. In I. Van Mechelen, J. A. Hampton, R. S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis*. London: Academic Press.

Lakoff, G. (1987). *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

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

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, *111*, 309–332.

Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 333–352.

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

Minda, J. P., & Smith, J. D. (2001). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *27*, 775–799.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, *47*, 90–100.

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review*, *4*, 79–95.

Nosofsky, R. M. (1984). Choice, similarity, and the context model 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. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 700–708.

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 honor of William K. Estes, Vol. 1*. Lawrence Erlbaum, Hillsdale, NJ.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266-300.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S.C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*, 53–79.

Pitt, M. A., Myung, I., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, *109*, 472–491.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization*. Erlbaum, Hillsdale, NJ.

Rosseel, Y. (2002). Mixture models of categorization. *Journal of Mathematical Psychology*, *46*, 178–210.

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.

Smith, J. D., & Minda, J. P. (2002). Distinguishing prototype-based and exemplar-based processes in dot-pattern category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *28*, 800–811.

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.

Verbeemen, T., Storms, G., & Verguts, T. (in press). Varying abstraction in categorization: a K-means approach. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*.