Yet Another Look at Thirty Categorization Results

Wolf Vanpaemel (wolf.vanpaemel@psy.kuleuven.be)
Departement Psychologie, K.U.Leuven
Tiensestraat 102, B-3000 Leuven, Belgium

Sven Pattyn (sven.pattyn@psy.kuleuven.be)
Departement Psychologie, K.U.Leuven
Tiensestraat 102, B-3000 Leuven, Belgium

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

Abstract

We re-analyzed thirty data sets reported in the literature and summarized by Smith and Minda (2000), based on Medin and Schaffer's (1978) 5-4 structure. In their meta-analysis, Smith and Minda (2000) focused on comparing the prototype and the exemplar model. In our meta-analysis, we applied the varying abstraction model, a multiple-prototype model proposed by Vanpaemel, Storms, and Ons (2005), that reduces to the prototype and the exemplar model in special cases. While we found a lot of heterogeneity in the best performing model across data sets, overall, the exemplar model turned out to account for the data best. However, a slight modification of the exemplar model improved performance in one condition, while in another condition, a modification of the prototype model recovered the data best.

Although categorization is an essential and thoroughly studied cognitive task, the debate about which model describes human category learning best hasn't settled yet. Two theories have been dominating this debate. According to the prototype theory (Nosofsky, 1987; Reed, 1972), people store an abstract summary of a category and categorize a new item by comparing the item to the summary. According to the exemplar theory (Medin & Schaffer, 1978; Nosofsky, 1986), people do not make any abstraction at all. Instead, a category is represented by stored memories of all previously encountered category exemplars. A new item is categorized by comparing it to all the category exemplars. Due to the influential studies of Medin and his colleagues, the focus has shifted from prototype theory to exemplar theory (Medin, Dewey, & Murphy, 1983; Medin & Schaffer, 1978).

The traditional representational assumptions are perfectly viable with small categories and fairly similar stimuli. However, when categories are large, remembering all exemplars seen, as the exemplar model claims, might not be possible. And when stimuli are very dissimilar, taking the average, as the prototype model claims, might be a strange thing to do. Everyday natural language categories, such as fruits and furniture, can be large, containing very dissimilar exemplars, so in this context both traditional views are problematic.

Vanpaemel et al. (2005) presented a model called the varying abstraction model that tries to find middle ground between both unrealistic representational assumptions. Therefore, the varying abstraction model introduces a set of new models which formalize the idea that people use multiple prototypes to represent a category. The model is particularly useful to uncover if and which multiple prototypes are used to represent a category. The present paper describes an elaborate application of the varying abstraction model.

In a critical review, Smith and Minda (2000) summarized and re-analyzed 30 data sets that made use of (a certain instantiation) of the so called 5-4 category structure. This category structure was introduced by Medin and Schaffer (1978) and has been particularly influential in artificial category learning studies, yielding ample evidence in support of the exemplar model (Medin, Altom, & Murphy, 1984; Medin et al., 1983; Nosofsky, Palmeri, & McKinley, 1994) (but see, Smith and Minda (2000) who questioned the evidence). The varying abstraction model puts new challengers to the exemplar model in the modeling arena. Not only the prototype model, but also all multiple-prototype models of the varying abstraction model family give a possible account of human category learning. Both the experimental and theoretical importance of the 5-4 structure and the direct availability of the data sets motivated us to analyze the 30 data sets using the varying abstraction model.

The purpose of this paper is twofold. First, we illustrate how the varying abstraction model can be applied to uncover the way people represent categories when making categorization decisions. Second, we re-analyze data from a category structure that has provided influential evidence for the exemplar model to investigate if this evidence still holds when the exemplar model is challenged by more than one model.

This paper is organized as follows. First, we review the varying abstraction model. Second, we briefly discuss the 5-4 category structure and the 30 data sets. In the third section the results of the varying abstraction model analysis of the 30 data sets are presented.

The varying abstraction model

The basic idea of the varying abstraction model is (1) to make up a partition[1] for each category and (2) to construct for every subset of the partition the prototype by averaging over all the exemplars in that subset. These prototypes are called the pseudo-exemplars and are used to represent the category. Specifying, for all categories, a partition is enough to define a model. Such a model is called a pseudo-exemplar model and can be fitted to empirically obtained data. In a pseudo-exemplar model, a stimulus is categorized based on its simi-

[1] A partition of a set S is defined as a collection of disjoint, nonempty subsets of S whose union is S.

larity to a number of pseudo-exemplars, rather than to every single exemplar in each category as is stipulated in the exemplar model, or to a single prototype for each category, as proclaimed by the prototype model. The varying abstraction model is a family of pseudo-exemplar models.

In a typical categorization experiment, a participant has to classify a stimulus S_i in one of a limited number of categories. The probability of categorizing stimulus S_i in category C_J, out of M possible categories, is computed as

$$p_{iJ} = \frac{\beta_J \eta_{iJ}}{\sum_{K=1}^{M} \beta_K \eta_{iK}}. \tag{1}$$

The crucial part of this equation is η_{iK}. It denotes the similarity of stimulus S_i to category C_K. Equation (1) prescribes that a stimulus is classified in the category it is most similar to. Further, β_K are free parameters, interpreted as the response bias towards category C_K. They range from 0 to 1 and satisfy the constraint $\sum_{K=1}^{M} \beta_K = 1$.

The stimulus-to-category similarity is given by

$$\eta_{iJ} \equiv \sum_{F_q \in D_J} \eta_{iq}. \tag{2}$$

With F_q we denote a pseudo-exemplar, which is the average of a subset of exemplars. Literally, such a pseudo-exemplar is not an exemplar but formally it can be treated as an exemplar, hence its name. The set of all pseudo-exemplars of the category C_J is denoted as D_J. Further, η_{iq} denotes the similarity between stimulus S_i and pseudo-exemplar F_q. This similarity between stimulus S_i and pseudo-exemplar F_q is related to the distance d_{iq} between these items via

$$\eta_{iq} = \exp(-d_{iq}^{\alpha}). \tag{3}$$

Two special cases are popular: the one where $\alpha = 1$, resulting in an exponential decay function, and the one where $\alpha = 2$, resulting in a Gaussian decay function. When the stimuli are fairly discriminable, the exponential decay is favored over the Gaussian (Nosofsky & Johansen, 2000).

To be able to conceptualize a distance between items, it is assumed that stimuli can be represented as points in a D-dimensional psychological space. The (psychological) distance between items S_i and F_q is calculated as

$$d_{iq} = c[\sum_{k=1}^{D} w_k |x_{ik} - x_{qk}|^r]^{1/r}. \tag{4}$$

Here, x_{jk} is the coordinate of item S_j or F_j on dimension D_k. Further, w_k are free parameters that are interpreted as the proportion of attention allocated to dimension D_k. They are called the weights and satisfy, for all k, $0 < w_k < 1$ and $\sum_{k=1}^{D} w_k = 1$. The parameter c is a free scaling parameter. It reflects discriminability in psychological space and runs from 0 to ∞. Finally, r denotes the metric. The distance is called city-block when $r = 1$ and Euclidean when $r = 2$. There is some evidence that the city block distance is appropriate for stimuli with integral dimensions, while for stimuli with separable dimensions the Euclidian distance is more appropriate (Shepard, 1991).

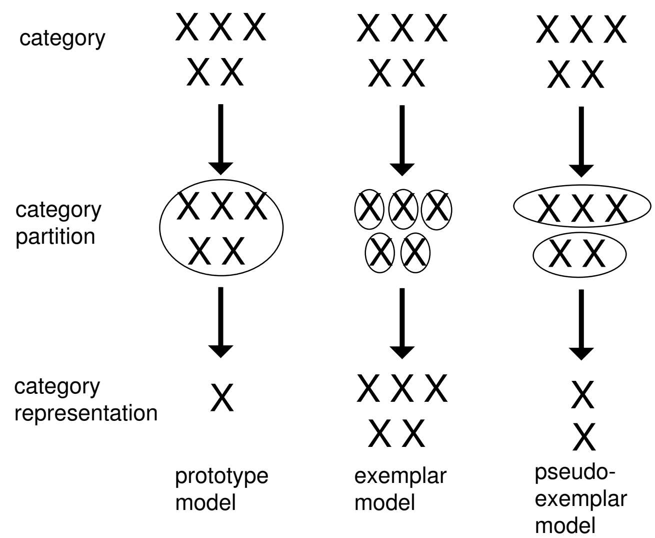

Figure 1: The two extreme partitions of a category correspond the two traditional models (i.e., the prototype model and the exemplar model). Any intermediate partition corresponds to a new pseudo-exemplar model.

Finally, since a pseudo-exemplar is defined as the average of a subset of exemplars, it is natural to define the coordinates of pseudo-exemplar F_q as

$$x_{qk} = \frac{1}{R_q} \sum_{S_i \in B_q} x_{ik}, \tag{5}$$

where B_q denotes a subset in a partition of a category and R_q denotes the number of items in subset B_q. The coordinates of the stimuli can be predefined by the experimenter, or can be identified in a psychological space using multidimensional scaling on pairwise similarity data for all the stimuli (Borg & Groenen, 1997; Lee, 2001).

A set of N elements has two "extreme" partitions: one when there is only 1 subset (of N elements), and one when there are N subsets (of 1 element each). Figure 1 shows that these extreme partitions pick out the traditional models and that all other partitions pick out new, intermediate models. In such a model, a category is represented by multiple prototypes.

In sum, the varying abstraction model reduces to the traditional models when the two extreme partitions are chosen for each category and introduces multiple-prototype models when non-extreme partitions are picked.

The 30 5-4 data sets

In their seminal paper, Medin and Schaffer (1978) defined two categories that have fueled the categorization research ever since. The first category consists of five elements, while the second category has four elements, hence it is commonly referred to as the 5-4 structure. Apart from these nine training stimuli, there are seven transfer stimuli. All 16 stimuli vary on four binary dimensions. Their logical structure is shown is Table 1.

Table 1: Medin and Schaffer's (1978) 5-4 structure

	$D1$	$D2$	$D3$	$D4$
category A stimuli				
$A1$	1	1	1	0
$A2$	1	0	1	0
$A3$	1	0	1	1
$A4$	1	1	0	1
$A5$	0	1	1	1
category B stimuli				
$B1$	1	1	0	0
$B2$	0	1	1	0
$B3$	0	0	0	1
$B4$	0	0	0	0
transfer stimuli				
$T1$	1	0	0	1
$T2$	1	0	0	0
$T3$	1	1	1	1
$T4$	0	0	1	0
$T5$	0	1	0	1
$T6$	0	0	1	1
$T7$	0	1	0	0

Smith and Minda (2000) summarized 30 data sets that made use of the 5-4 structure, collected by different researcher (see their appendix A). The 30 data sets not only differ in their exact instantiation of the 5-4 structure, but also in the specific instructions given to the participants, in the moment of measuring the transfer performance, etc. These differences might facilitate the use of an exemplar or prototype strategy, so it is useful to divide the data sets in subgroups.

There are six data sets from a study where participants were trained on the individual dimension values that were prototypical of each category. These six sets are 11, 12, 14, 15, 17 and 18. In a comment on Smith and Minda's (2000) article, Nosofsky (2000) argues that the category structure tested in these sets does not conform to the 5-4 structure. For three data sets, participants were instructed to use a prototype or rule-based strategy: the sets 4, 5 and 25. Sets 27, 28 and 29 are produced under deadline conditions. The sets 20, 21, 22 and 23 are from a study where transfer performance was tested at different points in time. The number of the data sets follows the chronological order, i.e. data set 20 was collected at the earliest stage, data set 23 at the latest stage. Data set 19 averages these data. According to Smith and Minda (2000), data set 23 does not reflect early learning and thus does not disfavor an exemplar strategy. Further, data set 6 sampled exemplars from an infinite pool without replacement. Finally, data set 3 can be omitted since it is exactly the same as data set 2. In short, for Smith and Minda (2000), the subgroup of data sets that "most heavily favored exemplar processing" is 1, 2, 7, 8, 9, 10, 13, 16, 23, 24, 26 and 30 (p. 13).

In his comment, Nosofsky (2000) reduces this set further. First, he argues that data set 30 should be in the deadline condition subgroup too. Second, he leaves out data sets 8 and 9 because he argues that they are no appropriate instantiations of the 5-4 structure. It is not clear in which subgroup he would put data sets 19, 20, 21, 22 and 23. All in all, according to Nosofsky (2000), the "exemplar subgroup" is 1, 2, 6, 7, 10, 13, 16, 24 and 26.

According to Smith and Minda, data set 23 did not reflect early learning, while data set 22 did, so it potentially disfavored an exemplar strategy. However, after analysing data set 22, it became clear that such is not the case, hence, we consider only data sets 20 and 21 as reflecting the early stages of learning.[2] Further, data sets 6 and 7 come from the same study as data sets 8 and 9 and are dubious in the same respect as these two sets. To be safe, we excluded these two sets as well. Taking all remarks together, one can safely consider data sets 1, 2, 10, 13, 16, 22, 23, 24 and 26 as good instantiations of the 5-4 structure that favor an exemplar strategy.

Results of the analysis of the 30 data sets

There are 15 possible partitions for a set of four elements and 52 for a set of five, hence there are 780 different pseudo-exemplar models. All these 780 pseudo-exemplar models were fit to all 30 data sets using maximum likelihood estimation (Myung, 2003). There are two categories, so it was assumed that the categorization responses follow a binomial probability distribution with success probability p_{iJ}, as expressed in equation (1). For the 5-4 data, every pseudo-exemplar model has five free parameters: three weights, one bias parameter β_A, and one scaling parameter c. Both r and α were set to 1, as in Smith and Minda (2000).

For ease of reference, all pseudo-exemplar models discussed in this section are listed in Table 2. The notation used is the following: model 597 is characterized by the sequences 12332 and 1231, indexing subset membership of exemplars A1 A2 A3 A4 A5 and B1 B2 B3 B4 respectively. As such, model 597 is defined by the partition {{A1};{A2,A5};{A3,A4}} for category A and the partition {{B1,B4};{B2};{B3}} for category B. Hence, in this specific pseudo-exemplar model, category A and B are each represented by three pseudo-exemplars.

Table 2: Pseudo-exemplar models discussed in this paper. Mn means model number.

mn	$A1$	$A2$	$A3$	$A4$	$A5$	$B1$	$B2$	$B3$	$B4$
1	1	1	1	1	1	1	1	1	1
166	1	2	1	2	2	1	1	1	1
597	1	2	3	3	2	1	2	3	1
750	1	2	3	4	3	1	2	3	4
765	1	2	3	4	4	1	2	3	4
780	1	2	3	4	5	1	2	3	4

Results for individual data sets

There was no single pseudo-exemplar model that outperformed all other pseudo-exemplar models for all 30 data sets. There were 25 different winning pseudo-exemplar models.

[2] In fact, it will turn out that data set 22 is one of the two data sets where the exemplar model outperformed all the other pseudo-exemplar models.

Model 597 was the best in two cases (sets 4 and 6), model 750 in three cases (sets 1, 16 and 19) and model 780 (the exemplar model) in two cases (sets 22 and 23). All other data sets had a unique best fitting pseudo-exemplar model (except sets 2 and 3, which are identical). The prototype model was never the best.

In both traditional models, all categories are assumed to have the same representation: every category is represented by one prototype, or by all exemplars. In contrast, the varying abstraction model incorporates models in which only one category has a prototype or an exemplar representation.

There were two data sets where the best model had an exemplar representation for category A (sets 22 and 23). In contrast, there were nine data sets where a model with an exemplar representation for category B did best (1, 11, 14, 16, 19, 24, 30, and of course 22 and 23 again). As far as a prototype representation is concerned, only for data set 17, a model where category A had a prototype representation did best, while there were two data sets where a model with a prototype representation for category B did best (sets 15 and 29).

Because of the heterogeneity among the winning pseudo-exemplar models, there is no obvious conclusion to draw from the individual results from the 30 data sets (except perhaps the heterogeneity itself). We tried to interpret the results across all 30 data sets by calculating, for all 780 pseudo-exemplar models, its averaged position over the 30 data sets.

Results averaged across data sets

Averaging the positions of the pseudo-exemplar models across all 30 data sets led to a most remarkable finding. Although the exemplar model was only the best pseudo-exemplar model for two data sets, overall, the exemplar model outperformed all other 779 models. Its average position was 82.6. The prototype model reached place 155, with an average position of 251.7. Also when we limited ourselves to the appropriate data sets (i.e., excluding sets 6, 7, 8, 9, 11, 12, 14, 15, 17 and 18), the exemplar model was the winning pseudo-exemplar model (with an average position of 29.4). The prototype model reached place 305, with an average position of 344.8.

To understand this unexpected finding we looked at the averaged performance of the pseudo-exemplar models for the different subgroups of data sets. The first subgroup was the "exemplar subgroup" 1, 2, 10, 13, 16, 22, 23, 24 and 26. Also for this subgroup, the exemplar model was overall the best (average position 6.22). The prototype model ranked at place 396 (average position 400.4).

Surprisingly, for the three data sets produced under prototype/rule instructions, the result was much alike the result for the exemplar subgroup. Again, the exemplar model outperformed all other pseudo-exemplar models (average position 12.3). The prototype model reached place 357 (average position 363.3). This observation coincides with Nosofsky's (2000) finding that performance on stimulus A2 versus stimulus A1 under prototype/rule instructions is qualitatively identical to performance in the exemplar condition (see his figure 2, his "standard tasks" correspond more or less to our "exemplar subgroup", his "instructed tasks" correspond to our "prototype/rule subgroup")

For the two data sets collected at early stages of learning, the exemplar model did, somewhat surprisingly, very well. It reached place 5 with an average position of 19.0. The prototype model reached place 640 only, with an average position of 528.0.

The only condition where the exemplar model performed poor was the deadline condition. For this subgroup, the exemplar model reached place 81 (average position 138.0). The prototype model outperformed the exemplar model, reaching the second place (average position 13.00). Including data set 30 in the deadline condition subgroup, as Nosofsky (2000) suggested, gave a better result for the exemplar model (it climbed to place 59, with an average position of 106.3) and a worse result for the prototype model (it fell to place four, with an average position of 24.0), but the observation that the exemplar model performed poor did not change.

In short, the exemplar model did not only perform well for the "exemplar condition" data sets, as one could expect, but also, rather unexpectedly, in some "prototype conditions" data sets, i.c. the data sets with prototype/rule instructions and the data sets reflecting early learning. This explains why the exemplar model was the best model across all data sets.

For two conditions, the varying abstraction model uncovered one or more pseudo-exemplar models outperforming the exemplar model and prototype model. In the early learning condition, the winning pseudo-exemplar model (model number 765, average position 8.0) was only a slight modification of the exemplar model (see table 2). The opposite pattern was found for the data sets produced under deadline conditions. Now, the pseudo-exemplar model outperforming all other pseudo-exemplar models (model number 166, average position 4.3) was a modification of the prototype model (see table 2). The inclusion of data set 30 did not affect the observation that model 166 did best for these data sets (average position 6.5).

It is instructive to go beyond the winning model only and broaden the view by looking at the other models with a good averaged performance. There were four models in the overall top 20 where category A had an exemplar representation and 12 models in the overall top 20 with an exemplar representation for category B. Only one model with a category A prototype representation entered the overall top 20, no model with a prototype representation for category B did that well. This pattern was exactly mirrored in the "exemplar subgroup", except for the fact the one model with a prototype representation for category A disappeared from the top 20.

In the two conditions where the exemplar model did surprisingly well (i.e., the prototype/rule instructions and the early stages of learning), a high number of pseudo-exemplar models with an exemplar representation for category B were in the top 20: seven and nine respectively. Three models of the top 20 had an exemplar representation for category A in the prototype/rule subgroup, two such models were present in the early learning subgroup. This time, a prototype representation for category A never entered the top 20, but one model with a prototype representation for category B was in the top 20 for the prototype/rule subgroup.

It is apparent that this scheme breaks down for the 20 best models that account for categorization under deadline conditions. The prototype representation for category A was present in three top 20 models, and for category B in 2 top 20 models. The exemplar representation for category A was never present in the top 20 and for category B in two top 20 models only. These findings are in sharp contrast with the other subgroups.

General Discussion

Three conclusions regarding category learning behavior on the 5-4 structure can be formulated based on our analyzes.

First, overall, the exemplar model gave the best account of the 30 5-4 data sets, even when challenged by 779 other models. When we look at the representations for category A and category B separately, the analysis of the individual sets and the analysis of the averaged performance suggests the same conclusion: especially for category B, there was an advantage to have an exemplar representation. This exemplar advantage was also present, but to a lesser extent, for category A. There are two obvious reasons to explain this difference. First, there are 52 pseudo-exemplar models with an exemplar representation for category B, while there are only 15 pseudo-exemplar models with an exemplar representation for category A. Second, as noted by Smith and Minda (2000), category B is more ambiguous than category A, which might induce the need for an exemplar representation.

A second conclusion is that there was a lot of heterogeneity in the set of best pseudo-exemplar models. While the exemplar model was overall the best model, it was only the best in two of the 30 data sets. Only one model, model number 750, gave the best account of more than two data sets. This heterogeneity among the best fitting models reflects the fact that the data sets themselves are very different in nature. The only aspect that is shared by all sets is the category structure used. The instantiation of this structure and the exact experimental conditions are different for many data sets. The exact instructions given to the participants, the moment of measuring transfer and the amount of time allowed are only a few of the most obvious factors influencing a categorization strategy. Interestingly, even when only the two traditional models are contrasted, the analysis of the 30 data sets does not univocally select one of the two models. In 9 cases (the data sets 6, 8, 9, 12, 14, 17, 27, 28 and 29), the prototype model outperformed the exemplar model, so it not surprising that there is no best fitting pseudo-exemplar model for all 30 data sets.

Finally, in the early learning condition, the varying abstraction model uncovered a slight modification to the exemplar model that recovered the data best. For the deadline condition data, the exemplar model was clearly not the appropriate model. The varying abstraction model uncovered a prototype-resembling model that accounted for the data best. The winning pseudo-exemplar models uncovered have, in both cases, a strong intuitive appeal. If people use an intermediate representation, one would expect that this representation closely resembles the exemplar representation in cases where the exemplar model performs well, and a prototype resembling representation in cases where the prototype model performs well, while the opposite pattern would be suspicious. These examples illustrate how the varying abstraction model can be used as a tool to find the middle ground between the two extreme and often unrealistic representational assumptions by identifying models with an intermediate representation.

A few remarks should be made with respect to these conclusions. Fitting and comparing all the pseudo-exemplar models of the varying abstraction model family is an extension of comparing the fit of the exemplar and prototype models only. Therefore, all concerns that have been raised against this endeavor hold in this case too. The two most severe concerns regard generality and model complexity.

Smith and Minda (2000) suggested that the 5-4 category structure is not theoretically neutral, in the sense that the 5-4 category structure "may encourage exemplar-memorization processes because of its poor coherence, its difficulty, and its small, memorizeable exemplar sets" (p. 3). Therefore, the data and results obtained based on the 5-4 structure may only generalize narrowly. Maybe the exemplar strategy is a specialized categorization strategy, suitable for, for example, small and difficult categories, but not a general one. If it is the case that category representation is sensitive to category structure, the varying abstraction model might prove to be a useful tool when researchers set out to explore the "space of category structure" more intensely. Relevant dimensions in this space could be, among others, category size and category complexity (Feldman, 2003).

Second, all pseudo-exemplar models have been evaluated on their ability to account for the data only. Recently, mathematical psychologists have raised the issue that selecting computational models by looking at their goodness-of-fit alone is problematic (Pitt & Myung, 2002). Model complexity, which is the inherent flexibility of a model, should be taken into account as well. Pitt, Myung, and Zhang (2002) found that the exemplar model is about $e^{60} \approx 1.8$ times as complex as the prototype model.[3] This difference in model complexity, however moderate, suggests that more trustworthy results could be achieved when complexity is taken into account. Comparing the performance of all pseudo-exemplar models using a combined measure for goodness-of-fit and complexity is important work for the future.

Conclusion

Our study illustrated how the varying abstraction model can be applied to gain additional insight in human categorization behavior when analyzing data from a thoroughly studied category structure. The model goes beyond the strict prototype-exemplar dichotomy by uncovering plausible intermediate pseudo-exemplar models that outperform the traditional models. The analyses gave overall support to the exemplar model, but at the same time indicated a modification of the exemplar model that accounts best for the data reflecting early learning and a modification of the prototype model that accounts best for the data produced under deadline conditions.

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 third author.

[3] Pitt et al. (2002) did not include a bias parameter in their analysis. This omission is not expected to influence the finding that the exemplar model is slightly more complex than the prototype model.

References

Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling: Theory and applications*. New York: Springer-Verlag.

Feldman, J. (2003). The simplicity principle in human concept learning. *Current Directions in Psychological Science*, *12*, 227–232.

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

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. L., Dewey, G. I., & Murphy, T. D. (1983). Relationships between item and category learning: Evidence that abstraction is not automatic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 607–625.

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

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

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. (2000). Exemplar representation without generalization? Comment on Smith and Minda's (2000) "Thirty categorization results in search of a model". *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1735–1743.

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*, 375–402.

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. J. (2002). When a good fit can be bad. *Trends in Cognitive Sciences*, *6*, 421–425.

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

Reed, S. K. (1972). Pattern recognition and categorization. *Cogntive Psychology*, *3*, 392–407.

Shepard, R. N. (1991). Integrality versus separability of stimulus dimensions: From an early convergence of evidence to a proposed theoretical basis. In J. Pomerantz & G. Lockhead (Eds.), *The perception of structure: Essays in honor of Wendell R. Garner.* Washington, D.C.: APA.

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.

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). Mahwah, NJ: Lawrence Erlbaum.