To cite this chapter:

Voorspoels, W., Vanpaemel, W., & Storms, G. (2008). Modeling typicality: Extending the prototype view. In B. Love, K. McRae, & V. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the cognitive science society* (pp. 757-762). Cognitive Science Society.

Modeling Typicality: Extending the Prototype View

Wouter Voorspoels (wouter.voorspoels@psy.kuleuven.be)
Department of Psychology, concat, tiensestraat 102,
3000 Leuven, Belgium

Wolf Vanpaemel (wolf.vanpaemel@psy.kuleuven.be)
Department of Psychology, concat, tiensestraat 102,
3000 Leuven, Belgium

Gert Storms (gerrit.storms@psy.kuleuven.be)
Department of Psychology, concat, tiensestraat 102,
3000 Leuven, Belgium

Abstract

In this study we re-evaluated earlier findings that natural language categories are represented by an exemplar representation rather than by a prototype representation. Using a restricted prototype model, an exemplar model and a "flying" prototype model, we predicted typicality ratings for 11 natural language categories from 2 semantic domains, animals and artifacts. We showed that exemplar models outperform prototype models when the prototype is restricted to the average exemplar, but the opposite pattern emerged when the prototype was free to move.

Keywords: natural language categories; typicality; computational models.

Introduction

One of the most important findings in the study of natural language concepts is that categories show a stable, graded category-membership structure. Members of natural language categories differ in the degree to which they belong to the category, the degree to which they are a good example of a category. For example, people generally judge a cow to be a better example of the category *mammals* than a whale. In the same way, a car intuitively seems to be a better example of the category *vehicles* than a zeppelin.

Since Rosch and Mervis' (1975) hallmark paper on family resemblances, this graded structure, generally referred to as the typicality gradient, has become an important variable in the study of natural language categories, being both explanatory and in need of explanation. Typicality has been demonstrated to be influential in a wide variety of cognitive tasks (Hampton, 1993; Malt & Smith, 1984) as well as an important predictor for priming effects (Rosch, 1977) and performance in tasks of inductive reasoning (Rips, 1975) and production (Hampton & Gardiner, 1983). Because of its importance, typicality is also an important evaluation criterion for models of concept representation. Any such model should be able to give an account of the graded category structure and correctly predict differences in the typicality of subordinate members of a category.

Two contrasting views on category representation have dominated the computational research on categories and concepts, each giving a different account of the graded internal structure of categories. On the one hand, the prototype view states that categories are represented by abstract summary representations, a prototype, and translates typicality as the similarity of a category member to this prototype. In this view, the concept *vehicle* is a summary representation of what vehicles are like on average, abstracted from specific instances of vehicles. The typicality of *car* for the category *vehicle* then is the similarity of *car* to the abstract prototype. On the other hand, the exemplar view proposes that a category is represented by previously encountered instances of the category. Here, typicality is conceptualized as the summed similarity of a category member to other members of the category. In this view, the concept vehicle consists of representations of previously encountered instances of vehicles. The typicality of *car* is then its summed similarity to all stored instances of *vehicle*.

While model-based comparisons of the prototype and the exemplar view are abundant in artificial category learning studies (e.g., Nosofsky, 1992), such comparisons have for a long time been relatively absent in the study of natural language concepts. One exception is a study by Voorspoels, Vanpaemel and Storms (in press). They contrasted a successful exemplar model – the generalized context model (GCM; Nosofsky, 1984, 1986) – with a prototype model in the prediction of typicality across a broad range of semantic concepts. The results clearly indicated that natural language categories are represented by their subordinate members rather than by a summary representation. These results were in line with earlier findings in natural language categories (e.g., Storms, De Boeck, & Ruts, 2000), using different approaches to prototype and exemplar models than Voorspoels et al. (in press), as well as the general finding in category learning tasks using artificial stimuli (e.g., Nosofsky, 1992).

In the study of Voorspoels et al.(in press), it was assumed that the prototype of a category is the mean of a category, representing the average position across all members on all relevant dimensions. While this is arguably the most widespread conception of a prototype, this assumption constitutes a strong restriction as to what a summary representation can consist of. For example, in superordinate language categories, such as fruit, it seems counterintuitive to think of a prototype being the average of members such

as litchis, bananas and pumpkins. Moreover, it has already been empirically demonstrated that, for some categories, ideal dimensions and contrast categories play a role in category membership that is more important than central tendencies (e.g., Barsalou, 1985; Lakoff, 1987; Goldstone, 1996, 2003, Ameel & Storms, 2006). In the wake of these findings, different conceptions of prototypes have been proposed, such as *ideals* and *caricatures*. Therefore, when contrasting the prototype view with other views, it seems reasonable to extend the traditional notion of prototype as a central tendency to other possible summary representations. In this study we want to contrast a prototype model with an exemplar model. Instead of only considering a central tendency prototype, we also included freely moving prototype, referred to as a "flying" prototype, to allow for different alternatives to a central tendency prototype.

In the remainder of this paper, we will discuss in more detail, how the notion of prototype can be extended. Then we will present the dataset which was used, followed by a review of the models that were used in this study, implementing the underlying concepts of exemplar representation, central tendency prototype and flying prototype. After this, we will present and discuss the results of the model fitting.

Types of prototypes

Arguably, the idea of a prototype as a central tendency of a category is the most widespread. There are however several reasons why a category representation would not be made up by a central tendency prototype, a representation that is the result of averaging across category members. Here we will discuss two important findings which lead to the notions of *caricatures* and *ideals*. We will discuss these ideas in turn. We will use the term prototype without further specification to refer to a summary representation, without specifying whether it concerns a central tendency, caricature or ideal prototype.

It has been shown that, at least for some categories and in some circumstances, the representations of those categories display exaggerations, deviations from the mean on certain dimensions. The prototype of a category can in these cases either be a *caricature* or an *ideal*. Categories are represented by a *caricature* to the extent their representation is determined by not being something else. Goldstone (1996) characterizes caricatures as follows: "Caricatures […] assume dimension values that depart from the central tendency *in the opposite direction from the central tendency of other concepts* […]" (Goldstone, 1996, p. 617). He also empirically demonstrated that when concepts are interrelated, caricatures show an advantage to central tendency prototypes in categorizing tasks (Goldstone, 1996, 2003).

Another type of "deviant" prototypes, *ideals* are, like caricatures, characterized by extreme values on certain dimensions, either true of only a few category members or true of none at all (Barsalou, 1985), but nonetheless present in the summary representation. Ideal prototypes differ from caricatures to the extent that ideals are not dependent on the interrelatedness of categories. Barsalou (1985) demonstrated convincingly that ideal representations predict typicality better than central tendency prototypes for goal-derived categories. More recently, Lynch, Coley and Medin (2000) demonstrated that for tree experts, ideal representations such as extreme height and weediness predicted goodness of example ratings better than similarity to the average tree. Distinguishing between ideals and caricatures can prove to be difficult, since both are characterized by extreme dimensional values.

Related to these findings regarding non-central prototypes, Ameel and Storms (2006) predicted typicality within several natural language categories using a central tendency prototype and a freely moving prediction point, searching for the optimal prediction. They found that the optimal predicting prototype was not attained by the central tendency prototype but rather by a prototype that moved away from the central tendency of the category, in the opposite direction of the other categories (making it a caricature).

The model used by Ameel and Storms (2006), further referred to as flying prototype model, will also be used in this study, to allow the prototype to deviate from a central tendency prototype, and thus implementing the extended notion of a prototype. The main aim of this study was to evaluate whether earlier findings favoring the GCM in the prediction of typicality hold true when using the flying prototype model instead of a more restricted prototype model.

Data

For the present study we used goodness-of-example ratings and a derived similarity measure of 11 categories, taken from a recent norm study of De Deyne, Verheyen, Ameel, Vanpaemel, Dry, Voorspoels, and Storms (in press). The set contains five animal categories (*birds*, *fish*, *insects*, *mammals* and *reptiles*) and six artifact categories (*clothing*, *kitchen utensils*, *musical instruments*, *tools*, *vehicles*, *weapons*), thus enclosing two semantic domains – animals, containing 129 exemplars, and artifacts, containing 166 exemplars. Every category consists of 20 to 33 exemplars.
The categories used in this study are different from the ones used in Voorspoels et al. (in press) because the flying prototype has been found to be succesfull at predicting typicality in representational spaces that contain possible contrast categories (Ameel & Storms, 2006). The data we used in this study allowed us to derive similarity matrices including all exemplars of a semantic domain, leaving the possibility of contrast effects open.

Derived similarity measure

In order to derive appropriate underlying representations of the domains included in the study, a similarity measure for each exemplar pair within a domain was needed. The data set presented in De Deyne et al. (in press) does not include similarity matrices for a whole semantic domain, but it does

include generated exemplar features and their rated applicability to each exemplar for each domain, with which we were able to derive similarity measures for all exemplar pairs within a domain.

For both domains, De Deyne et al. (in press) created an exemplar by feature matrix containing all exemplars of a domain and all features generated for the exemplars of a domain. The animal domain matrix contains 129 exemplars and 765 features, the artifact domain matrix contains 166 exemplars and 1295 features. Both matrices were filled in by 4 participants, judging the applicability of each feature for each exemplar (1 referring to applicable, 0 referring to not applicable). The reliability of these ratings was evaluated by applying the Spearman-Brown formula to the split-half correlations, resulting in an estimated reliability of .83 for the animal matrix and .81 for the artifact matrix (De Deyne et al., in press).

To derive the similarity measure for the present study, we started from the summed (across participants) feature-by-exemplar matrix of each domain and correlated the feature vectors of all exemplar pairs within a domain. These correlations can be considered similarity measures for the exemplar pairs, and were used to construct similarity matrices for each domain.

Goodness-of-example

The exemplars of each category were rated by 28 participants for goodness-of-example on a likert-rating scale ranging from 1 for very bad examples to 20 for very good examples. The reliability of the judgments was evaluated by means of split-half correlations corrected with the Spearman-brown formula and ranged from .91 to .98 (De Deyne et al., in press). In this study we analysed the typicality ratings, averaged across participants.

Model review

All three models evaluated in this study, base their prediction on an underlying spatial representation, in which the exemplars of a category are represented as fixed points in an M-dimensional space. The similarity between two exemplars in this space is inversely related to the distance of the two exemplars.

The typicality or goodness-of-example of an exemplar in a certain category can be conceptualized as the similarity of the exemplar to the category. The three models presented below, differ with respect to what is understood by the category, implying differences in typicality predictions.

The generalized context model

The GCM (Nosofsky, 1984, 1986) essentially assumes that categorization decisions are based on similarity comparisons with individually stored category exemplars.

Similarity The similarity between two exemplars is derived from the distance of the exemplars in the M-dimensional psychological space, adjusted by dimension weights, and a sensitivity parameter, which magnifies or demagnifies the psychological space. Formally, the scaled psychological distance between two exemplars i and j is given by:

$$d_{ij} = c\left(\sum_{k=1}^{M} w_k \left| x_{ik} - x_{jk} \right|^r \right)^{1/r}, \quad (1)$$

where x_{ik} and x_{jk} are the coordinates of exemplars i and j on dimension k, w_k a parameter reflecting the attention weight for dimension k, M is the number of dimensions, and c is the sensitivity parameter. The distance reduces to Euclidean distances when $r = 2$ and city-block distances when $r = 1$. Since Euclidean distances are generally accepted to be more appropriate for integral dimensions (Shepard, 1964), we fixed r at 2 in this study.

Similarity of a stimulus i to another stimulus j, is related to psychological distance as follows:

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

where d_{ij} is the scaled psychological distance between exemplar i and j.

Typicality prediction Following Nosofsky (1988), typicality of an exemplar is calculated by summing the similarity of that exemplar to all other exemplars in the category. Formally, the typicality of an exemplar i for category A is then given by:

$$T_{iA} = \sum_{j=1}^{n} \eta_{ij}, \quad (3)$$

where η_{ij} is the similarity of exemplar i to exemplar j, with j belonging to category A. The free parameters in the model consist of M-1 dimension weights and a scaling parameter c.

Central tendency prototype model

In the central tendency-based prototype, typicality of an exemplar in a category is the similarity of that exemplar to the prototype of the category instead of the summed similarity to all members of the category. As is typically assumed in prototype models, the category prototype (denoted by P_A) in this model is the centroid of the category's exemplars in the M-dimensional space (i.e., the point defined by averaging the coordinates of the exemplars on all underlying dimensions). The prototype has the status of an exemplar, consequently equation (1) and equation (2) apply.

Formally, the typicality of exemplar i for category A is given by:

$$T_{iA} = \eta_{iP_A}, \quad (4)$$

where η_{iPa} is the similarity of exemplar i to the prototype of category A. The free parameters of the model are identical to those of the exemplar model.

The "flying" prototype model

The third model evaluated in this paper, referred to as the flying prototype model, allows a summary representation to deviate from the central tendency of a category, thus implementing the extended prototype approach as described earlier. The flying prototype model can still be seen as a prototype model in that typicality is modeled as similarity to the prototype of a category. The prototype in this model however is free to "fly" within the M-dimensional psychological space, in order to find an optimal prediction of typicality. In this way the prototype can adopt other than average values, e.g. extreme values as in ideal or caricature representations.

The model used in this study is mathematically identical to the model used by Ameel and Storms (2006) to find optimal prediction points. Typicality of item *i* in category *A* is formalized as the similarity of the item to the category (see equation (4)). Here however, the similarity of item *i* to category *A* is given by:

$$\eta_{iP_A} = -\sqrt{\sum_{k=1}^{M}(x_{ik} - p_k)^2} \quad , \qquad (5)$$

where x_{ik} is the coordinate of exemplar *i* on dimension *k* and p_k is the coordinate of the prototype P_A on dimension *k*. Contrary to the central tendency-based prototype, P_A is not restricted to being the average of the category, but the coordinates of P_A are the free parameters in the model, enabling the prototype to move freely in the M-dimensional space to optimize the prediction. The number of free parameters thus equals the number of dimensions of the underlying psychological space. Note that, contrary to the GCM and the central tendency-based prototype model, the flying prototype model does not make use of attention weights. Given a certain psychological space, this results in an equal number of free parameters for the three models presented here.

Results

In order to obtain the underlying spatial representations, we used the similarity matrices derived for both semantic domains as input for a SAS MDS-analysis (SAS, V9). Solutions in 2 to 10 dimensions were considered for further analysis, as determining the number of underlying psychologically relevant dimensions for semantic concepts is not always obvious. As can be seen in Table 1, the stress values of both domains show a clear monotonically decreasing pattern, indicating the MDS-routine did not get trapped in local minima.

Table 1. MDS-stress values for the 2 domains.

	2D	3D	4D	5D	6D	7D	8D	9D	D10
Animals	.18	.12	.07	.06	.04	.04	.03	.03	.03
Artifacts	.16	.10	.07	.06	.05	.05	.04	.04	.03

For each category, in all considered dimensionalities, the three models described earlier were fitted to the observed typicality ratings (averaged across participants). The optimal parameter values were searched by maximizing the correlation between calculated and observed typicality. To ensure the maximizing routine did not get trapped in local maxima, we repeated this search several times. Generally the variation in obtained correlations was rather small, indicating optimal solutions were found. The resulting correlations are displayed in figure 1, displaying three dominant patterns.

First, it is clear from Figure 1 that the central tendency prototype does a worse job at predicting the typicality gradient of the natural language categories than does the flying prototype. This finding is in line with what Ameel and Storms (2006) have found. Note that in this study we used a central tendency prototype model with freely varying dimensions weights. This means that in terms of number of free parameters, it has the same amount of free parameters as the flying prototype model.

Second, figure 1 also shows that the GCM (solid lines) does a better job at predicting typicality than the central prototype model (dashed line). Although differences are not always large, the GCM consistently outperforms the central prototype model, except for the category *reptiles* in 4 dimensions and *kitchen utensils* in 2 and 3 dimensions. This finding is consistent with the findings of Voorspoels et al. (in press), who contrasted these two models across a different set of natural language categories.

Finally, comparing the GCM with the flying prototype model, it is obvious that the latter model convincingly does the best job in predicting typicality for all categories included in the study, across both semantic domains. From three dimensions onwards, the flying prototype model consistently outperforms both the GCM and the central prototype model. These results suggest that there are summary representations that, in these circumstances, predict typicality better than the the exemplar representation provided by the GCM. In other words, contrary to earlier findings (Storms et al., 2000, Voorspoels et al., in press), exemplar models do not seem to predict typicality better than summary representations in all circumstances.

The remaining question then is whether these summary representations are interpretable. Looking at the optimal prototype's position in 2 dimensional representations[1], our results seem to confirm the findings of Ameel and Storms (2006). While our flying prototypes for some categories did tend to move farther away from the central tendency prototype than in Ameel and Storms (2006), for the majority of the categories under study, the flying prototype moved away from the central tendency, in the opposite direction of the central tendencies of the other categories in the representation of the semantic domain. This observation corresponds to the notion of a caricature, and together with the better performance in predicting typicality, seems to suggest that rather than abandoning the notion of a summary representation all together, extending the notion of a prototype might be more fruitful.

[1] While it is not impossible to do the same in higher dimensionalities, this is not an easy task.

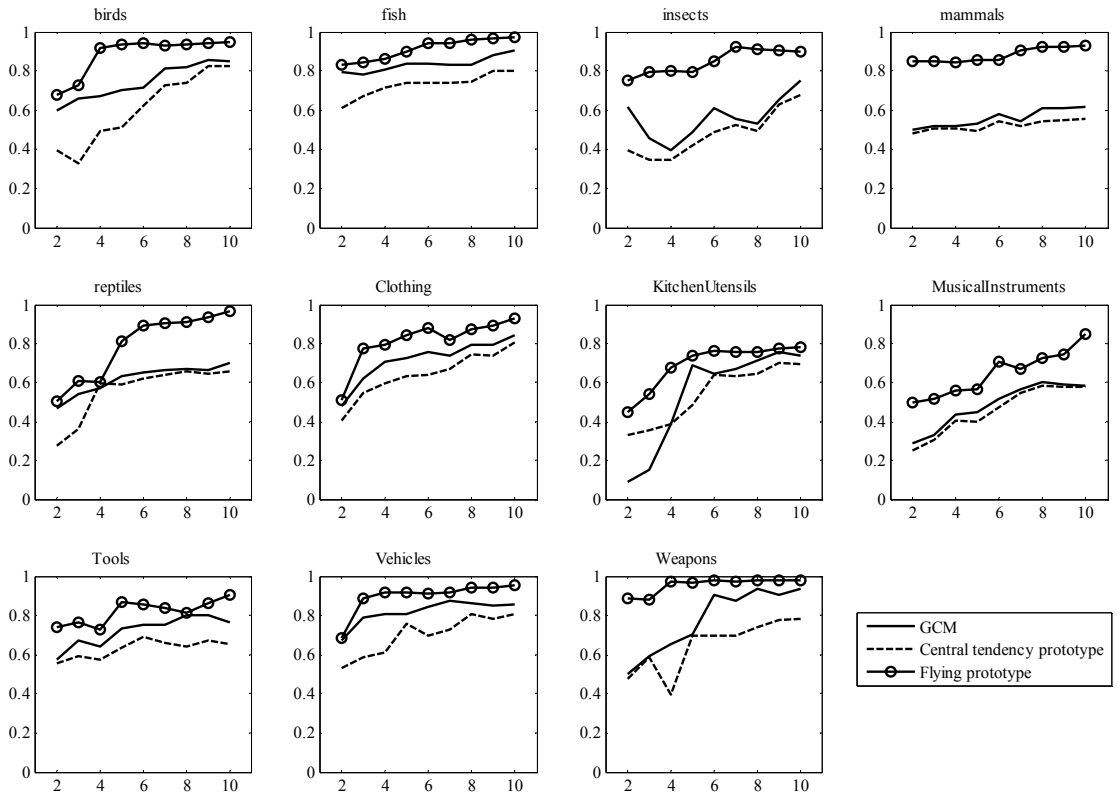


Figure 1. Correlations between observed and predicted typicality for the different categories in function of dimensionality. Note that the exemplars of a category are embedded in spatial representations of the domain they belong to (either *animals* or *artifacts*).

Discussion

In this study we examined whether recent results favoring an exemplar model over a prototype model in the prediction of typicality in natural language categories (Voorspoels et al., in press), were due to restricting the prototype model to one possible notion of a summary representation, the prototype as central tendency. To this end, we contrasted the GCM (Nosofsky, 1984, 1986) with a restricted prototype model and a prototype model that allows different notions of a summary representation, such as ideal prototypes and caricature prototypes. The three models were contrasted in the prediction of typicality across 11 natural language categories, belonging to one of two semantic domains – *animals* and *artifacts*.

Results confirmed earlier findings that the exemplar model consistently outperforms the restricted prototype model in the prediction of typicality in natural language categories. However, the "flying" prototype model, allowing ideal and caricature prototypes, in turn clearly outperformed the GCM for all categories in both semantic domains. In general, the prototype producing the best fits to the typicality data were caricature representations. This finding suggest that, at least under the circumstances implied in this study, there exist prototype representations that give a better account for the typicality gradient in natural language categories than exemplar representations.

An important remark is in place here. As we have already indicated, we based our prediction, as Ameel and Storms (2006) did, on spatial representations of a domain rather than isolated categories, contrary to Voorspoels et al. (in press). Since these representations contain possible contrast categories, part of the strength of the flying prototype model in predicting typicality might be due to contrast information inherent in the underlying representation. We selected these circumstances to test the flying prototype model against other, more widely known models, to give it an optimal chance at being competitive. Moreover, to interpret the summary representations resulting from the model, it is useful to see the category embedded in a larger semantic domain. Future research should be able to determine whether the results presented here hold true for contrast unaffected representations, i.e.

for isolated categories. Note that extending the representational space does not change the robust finding that an exemplar-based approach outperforms a prototype-based approach, in which the prototype is restricted to a central tendency. This is an important observation, since it shows that larger representational spaces as such do not necessarily favor prototype-like models.

Since part of the strength of the flying prototype model in this study might be due to contrast information, it would be interesting to adapt the GCM so that it too uses information from contrast categories within the domain. This could be implemented by translating typicality as a function of similarity towards members of the same category *and* members of other, contrasting categories. It remains an open question whether expanding the GCM in this way would tip the scales back in favor of exemplar models.

Acknowledgments

The research reported in this article is part of an interdisciplinary research project (IDO/02/004) and research project OT/05/27 sponsored by the Research Council of the university of Leuven (IDO/02/004). Both of these grants were given to the third author.
Correspondence concerning this article should be addressed to Wouter Voorspoels, K.U. Leuven, Department of Psychology, ConCat, Tiensestraat 102, B-3000 Leuven, Belgium. Email: wouter.voorspoels@psy.kuleuven.be.
Website: ppw.kuleuven.be/concat.

References

Ameel, E., & Storms, G. (2006). From prototypes to caricatures: Geometrical models for concept typicality. *Journal of Memory and Language, 55*, 402-421.

De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M. J., Voorspoels, W., & Storms, G. (in press). Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavioral Research Method*s.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory and cognition, 24*, 608-628.

Goldstone, R. L. (2003). Conceptual interrelatedness and caricatures. *Memory & cognition, 31*, 169-180.

Hampton, J. A. (1993). Prototype models of concept representation. In I. Van Mechelen, J. A. Hampton, R. S. Michalski, and P. Theunis (eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 67-95). London: Academic Press.

Hampton, J. A., & Gardiner, M. M. (1983). Measures of internal category structure: A correlational analysis of normative data. *British Journal of Psychology, 74*, 491-516.

Komatsu, L., K. (1992). Recent views of conceptual structure. *Psychological Bulletin, 3*, 500-526.

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

Lynch, E. B., Coley, J. D., & Medin, D. L. (2000). Tall is typical: Central tendency, ideal dimensions, and graded structure among tree experts and novices. *Memory & Cognition, 28*, 41-50.

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

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

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

Nosofsky, R., N. (1988). Exemplar-Based Accounts of Relations Between Classification, Recognition, and Typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 700-708.

Nosofsky, R. N. (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, pp. 149-167). Hillsdale, NJ: Erlbaum.

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

Rosch, E. (1978). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104*, 192-233.

SAS Institute Inc, SAS STAT ® User's guide 8, Cary, NC: SAS Institute Inc., 1999. 3884 pp.

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

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.

Voorspoels, W., Vanpaemel W., Storms, G. (in press). Exemplars and prototypes in natural language categories. *Psychonomic Bulletin and Review*.