To cite this chapter:

Voorspoels, W., Storms, G., & Vanpaemel, W. (2011). Contrast in natural language concepts: An exemplar-based approach. In L. Carlson, C. Hölscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 531-536). Cognitive Science Society.

Contrast in natural language concepts: An exemplar-based approach

Wouter Voorspoels (wouter.voorspoels@psy.kuleuven.be)

Gert Storms(gert.storms@psy.kuleuven.be)

Wolf Vanpaemel (wolf.vanpaemel@psy.kuleuven.be)
Department of Psychology, Tiensestraat 102
B-3000 Leuven, Belgium

Abstract

We examine the influence of contrast categories on the internal graded membership structure of natural language categories. To this end we contrast two exemplar models in their account of typicality: According to the GCM, typicality is the summed similarity towards all category members. According to the SD-GCM, typicality is determined by both the summed similarity towards all category members and the summed dissimilarity towards members of contrast categories. For five animal categories, we contrast the SD-GCM and the GCM in their account of typicality. Results indicate that the internal category structure can indeed be co-determined by dissimilarity towards potential contrast categories.

Keywords: concepts; categories; typicality; contrast categories; computational models

Introduction

The platypus is a mammal. It can, however, hardly be called a "good" mammal: It has webbed feet and a beak resembling that of a duck, it is venomous like insects and reptiles, it lays eggs as do birds and fish, and it is semi-aquatic, reminding one of amphibians. A cow, on the other hand, is a good, a more representative example of a mammal. Previous research suggests that people are in general agreement as to what are representative, good examples of a certain category and which members are bad examples. The graded membership structure, or typicality gradient, can be observed in a broad range of everyday natural language categories (e.g., Hampton & Gardiner, 1983; Rosch & Mervis, 1975).

Traditionally, typicality is defined as similarity towards a category representation: A member of a category is typical to the extent that it is similar to the category representation. Different views exist on what the category representation consists of. The two most dominant computational theories of category representation propose that a category is represented by a prototype (prototype models; e.g., Hampton, 1993), or the set of previously encountered members (exemplar models; e.g., Medin & Schaffer, 1978). In general, it is found across an impressive array of conditions, both in artifical category learning experiments and natural language categories that exemplar representations provide the best description of human categorization. For present purposes we will therefore focus on exemplar models.

While the graded membership obviously reflects the internal structure of a category, natural language categories are not isolated entities, but generally reside in rich semantic domains. Categories vary along a continuum of interrelatedness (Goldstone, 1996). For example, *mammals*, *birds* and *insects* all are animal categories and clearly are more interrelated than the categories *mammals* and *doorknobs*.

The interrelatedness of natural language categories elicits the following intriguing question: Is the internal membership structure of a category determined solely by characteristics internal to the category, or do other, related categories influence the category structure? In other words, is the the platypus an atypical member of *mammals* because of its apparent dissimilarity towards other mammals, or does its similarity towards other animal categories, such as *birds* and *reptiles*, contribute to its atypicality? The present paper is concerned with exactly this question and attempts to answer it using successful exemplar models that have their roots in artificial category learning research.

Contrast category effects

The most likely candidates to exert influence on the internal structure of other categories, are contrast categories. Contrast categories are considered to be at the same level of abstraction, belonging to the same immediate superordinate as the target category. Further, they are contrastive or incompatible in the sense that one and only one word is applicable to any member of the category (Miller & Johnson-Laird, 1976). For example, *mammals* and *birds* are contrast categories, both belonging to the same immediate superordinate category and they are mutually exclusive (an animal cannot be a *bird* and a *mammal* at the same time). We use the term *contrast category effect* for manifestations of influence of contrast categories on category based tasks.

The notion of contrast category has a long history in natural language concept representation literature. For example, in their influential family resemblance model, Rosch and Mervis (1975) assume that typicality of a category member is its similarity to other category members and its dissimilarity to members of contrast categories. Despite the theoretical importance attributed to contrast categories, little effort has been invested in demonstrating the independent role of contrast categories in natural language categories, and evidence is ambiguous. In a thorough test of the independent contribution of feature overlap with the target category and feature overlap with contrast categories, using both typicality ratings and classification response times, Verbeemen, Vanoverberghe, Storms, and Ruts (2001) found no evidence for contrast effects. In sharp contrast with the findings of Verbeemen et al., using a geometric prototype model of concept repre-

sentation, Ameel and Storms (2006) found evidence that prototypes of interrelated categories tend to move away from the central tendency, in a direction opposite of the contrast categories. This finding constitutes evidence in favor of contrast effects, since the prototype, i.e., the representation of a category, is under influence of other categories. In sum, the debate on the role of contrast categories in natural language categories is far from settled, and in need of further thorough and systematic examination.

In artificial category learning, empirical demonstrations of contrast category effects are more established (e.g., Davis & Love, 2010; Goldstone, Steyvers, & Rogosky, 2003; Palmeri & Nosofsky, 2001; Stewart & Brown, 2005) . For example, Palmeri and Nosofsky (2001) demonstrated that items residing at the target's category boundary most remote from other categories, were classified most accurately. In other words, clear category members were both high in similarity towards the target category and low in similarity towards other relevant categories. Goldstone et al. (2003) manipulated the interrelatedness of categories, and demonstrated a clear contrast category effect when categories are interrelated.

Interestingly, Stewart and Brown (2005) proposed and tested an exemplar model, the similarity-dissimilarity generalized context model (SD-GCM), that explicitly implements dissimilarity towards contrast categories as a component that co-determines category membership together with similarity to the target category. The SD-GCM is based on the generalized context model (GCM; Nosofsky, 1986), arguably the most influential exemplar model, but differs from the GCM in that the GCM only considers the similarity towards stored category members.

Outline

The present study examines the role of contrast categories in the representation of natural language categories. We compare two computational exemplar models developed in artificial category learning – the GCM and the SD-GCM –, in their account of the typicality gradient of five animal categories. The two models both provide a different account of the typicality gradient of a category: According to the GCM, the typicality of a member should depend solely on the summed similarity towards other category members. According to the SD-GCM on the other hand, typicality of a category depends on both similarity towards category members and dissimilarity towards contrast categories. To evaluate whether contrast categories play a role in the animal categories, we will therefore compare the two models on their ability to account for the observed typicality gradient in the animal categories.

Models

We will give a brief overview of both the standard GCM and the SD-GCM, adapted to account for typicality in natural language categories. Both models are exemplar-based, which means that they rely on stored exemplars to account for category-based decision. For natural language categories, exemplars are generally assumed to be defined at the nearest subordinate member level (e.g., Komatsu, 1992). For example, the exemplars of the category *mammals* are 'dog', 'camel', 'beaver', etc.

The generalized context model

The GCM (Nosofsky, 1986) is one of the most influential exemplar models in categorization research. Originally, the model was developed to account for categorization decisions, but it has successfully been adapted for typicality judgments (Nosofsky, 1988). Typicality of an exemplar i for category A is calculated by summing the similarity of that exemplar to all other exemplars in the category:

$$t_{iA} = \sum_{j \in A} s_{ij} \tag{1}$$

where s_{ij} is the similarity of exemplar i to exemplar j. The similarity between two exemplars is a function of the distance of the exemplars in an M-dimensional psychological space. Formally, the scaled psychological distance is given by:

$$d_{ij} = \left(\sum_{k=1}^{M} w_k \left| x_{ik} - x_{jk} \right|^r \right)^{\frac{1}{r}} \tag{2}$$

where x_{ik} and x_{jk} are the coordinates of exemplars i and j on dimension k, and w_k is the dimension weight for dimension k. We relied on Euclidean distances (r=2), which are generally accepted to be more appropriate for integral dimensions (Shepard, 1987). Similarity between i and j is derived from the weighted distance:

$$s_{ij} = exp\left(-cd_{ij}\right) \tag{3}$$

where c is the sensitivity parameter. The free parameters in the GCM consist of M-1 dimension weights and the sensitivity parameter.

The Similarity-Dissimilarity GCM

The SD-GCM (Stewart & Brown, 2005) is an extension of the GCM, making the same assumptions concerning underlying representation and similarity (i.e., Equations (2) and (3) still apply). However, whereas the GCM defines typicality of item i towards category A as the sum of the similarity towards members of that category, the SD-GCM also takes into account the dissimilarity of i towards alternative categories in the same representational space. Formally, the typicality of exemplar i to category A then is given by:

$$t_{iA} = \sum_{j \in A} s_{ij} + \sum_{l \notin A} \left(1 - s_{il}\right) \tag{4}$$

where s_{ij} and s_{il} are similarities between exemplar i and j and i and l respectively, with j being an element of category A, and l belonging to an alternate category. Typicality to a category is thus defined by the sum of the similarity towards members of the category plus the summed dissimilarity towards the other elements in the representational space. It is possible

to restrict the set of contrast categories, such that only a subset of exemplars in the representational space are included. For example, in accounting for the typicality of members of the category *mammals*, we might choose to restrict the set of contrast exemplars to *birds* and *fish*, leaving out the other potential contrast categories *insects* and *reptiles*.

Data

To compare the models, we chose categories in the animal domain, that more or less exhaustively cover the whole domain. In this way, no potential contrast category was accidentally missed: Each of the categories can be a contrast category for all other categories, and for each category all possible contrast categories are included. Moreover, the animal categories perfectly adhere to the definition of contrast categories, belonging to the same immediate superordinate category (animals) and being mutually exclusive.

Data were taken from a recent norm study of De Deyne et al. (2008). The set contains five animal categories (*birds*, *fish*, *insects*, *mammals* and *reptiles*), in total containing 129 exemplars. Every category consists of 20 to 33 exemplars. We used goodness-of-example ratings as a measure of typicality and a derived similarity measure of categories to construct an underlying stimulus representation.

Typicality

Goodness-of-example ratings taken from De Deyne et al. (2008). The exemplars of each category were rated by 28 participants for goodness-of-example for the category they belonged to on a Likert-rating scale ranging from 1 for very bad examples to 20 for very good examples. The reliability of the judgments for each category was evaluated by means of split-half correlations corrected with the Spearman-brown formula. The estimated reliability ranged from .91 to .98 . In the present study the typicality ratings are averaged across participants, resulting in a single typicality score for each member of a category.

Pairwise similarity

To allow the SD-GCM to test for potential contrast effects, it is imperative that the exemplars of all five categories are in the same similarity representation. This requires a measure of pairwise similarity between all exemplars in the ANIMAL domain. To this end we derived a pairwise similarity measure for all pairs of 129 exemplars in the ANIMAL set. De Deyne et al. (2008) reports an exemplar by feature matrix containing all exemplars of the animal domain and all features generated for its exemplars. The animal domain matrix contains 129 exemplars and 765 features. The features were generated by 1003 participants for the exemplars of the animal domain and span a broad range of characteristics (e.g., 'has an external skeleton', 'is a carnivore', 'runs fast'). Four participants judged the applicability of each feature for each exemplar (1 referring to applicable, 0 referring to not applicable). The estimated reliability of these judgments was .83 (see De Deyne et al., 2008, p. 1042).

To arrive at a pairwise similarity measure for each pair of animals, we summed the feature-by-exemplar matrices across the 4 participants and then correlated the summed feature vectors of all possible exemplar pairs within a domain.

Results

Both models operate on an underlying geometric similarity representation, in which exemplars are represented as points in an M-dimensional space and the distance between two points in the space is inversely related to the similarity of the exemplars. After describing how the representations were derived, we will examine how many, and which, contrast categories are best used for each of the target categories. To answer the question whether a category's internal membership structure can be influenced by contrast categories, we will compare the SD-GCM with the standard GCM, both in terms of optimal data fit and generalizability of the models.

Similarity representation

To investigate whether it makes sense to derive a geometric representation for the present stimuli, the percentage of triplets in the similarity matrix that violate the triangle inequality was calculated. Its very low value (.13 %) indicates that the stimuli can in principle be represented spatially. To generate the geometric similarity representation, the pairwise similarity measure for each pair of the 129 ANIMAL exemplars was used as input in a SAS non-metric MDS-analysis (SAS, V9). We considered representations in 2 to 6 dimensions. The appropriateness of the geometric representation was evaluated using stress as a badness-of-fit measure (Kruskal & Wish, 1978) for the MDS-solutions.

Table 1: Percentage of triplets violating the triangle inequality (TIV), and stress-values for the MDS-solutions.

TIV	Stress-values				
	2D	3D	4D	5D	6D
0.13%	.18	.12	.07	.06	.04

The stress-values in Table 1 suggest the appropriateness of the MDS-solutions. Following Kruskal & Wish, 1978, solutions with stress-values exceeding .10 are not considered for further analyses, leaving representations with Dimensionality from 4 to 6 for further analysis.

Identifying contrast categories

The notion of contrast categories refers to categories that are mutually exclusive and organized under the same superordinate term. Thus, for the animal categories, each category can have all other animal categories as potential contrast categories. Before comparing the SD-GCM to the GCM, we will therefore examine for each concept which are the appropriate contrast categories, and how many contrast categories should be included in the SD-GCM. For example, for the category

mammals the contrast category can be *birds*, *fish*, *insects* or *mammals*, or any combination of these four categories.

To identify the appropriate set of contrast categories, we optimized the correlation of the observed typicality ratings with the SD-GCM-based typicality ratings for the category members, varying the contrast categories included in the model. We first optimized the SD-GCM's typicality predictions using only one contrast category. For *mammals*, the best contrast category was *birds*. We then allowed two contrast categories, *birds* being one of the two, the second category being the second best contrast category of step one.[1]. This procedure was repeated until all four contrast categories were included. Analyses[2] were performed for all five target categories and the results are presented in Table 2.

Table 2: Optimal correlations between observed typicality ratings and the SD-GCM-based typicality ratings.

category	n of contrast categories			
	1	2	3	4
mammals	.80	.81	.87	.83
birds	.88	.88	.89	.89
fish	.84	.83	.83	.83
insects	.74	.49	.68	.66
reptiles	.75	.65	.64	.65

Table 2 reveals the importance of selecting an appropriate contrast category set. For *mammals*, best results were found when including three contrast categories: *birds*, *insects* and *reptiles*. For *birds*, including all other animal categories in the SD-GCM prediction produced the best result. For the three remaining categories, best results were found when including only one contrast category. For *fish*, the appropriate contrast category was *mammals*. For *insects* and *reptiles*, the appropriate contrast category was *fish*.

In sum, the best results were found including a particular set of contrast categories. For the following analyses, the we will always include the appropriate set of contrast categories as found in these analyses.

Comparing the SD-GCM to the GCM

To evaluate the performance of the SD-GCM in its account of typicality of the five animal categories, we will compare the model to the GCM, i.e. an exemplar model that does not include contrast information in its account of typicality. We first consider the goodness of fit of both models. In these analyses we establish whether both models are sufficiently able to capture the typicality gradient by optimizing the correlation across exemplars between observed and model-based typicality.

[1]Note that, unlike in e.g., regression analyses, adding a contrast category does not necessarily produce better results. This depends on the structure of the underlying similarity representation.

[2]For these analyses, the dimensionality of the underlying representation was fixed at 4

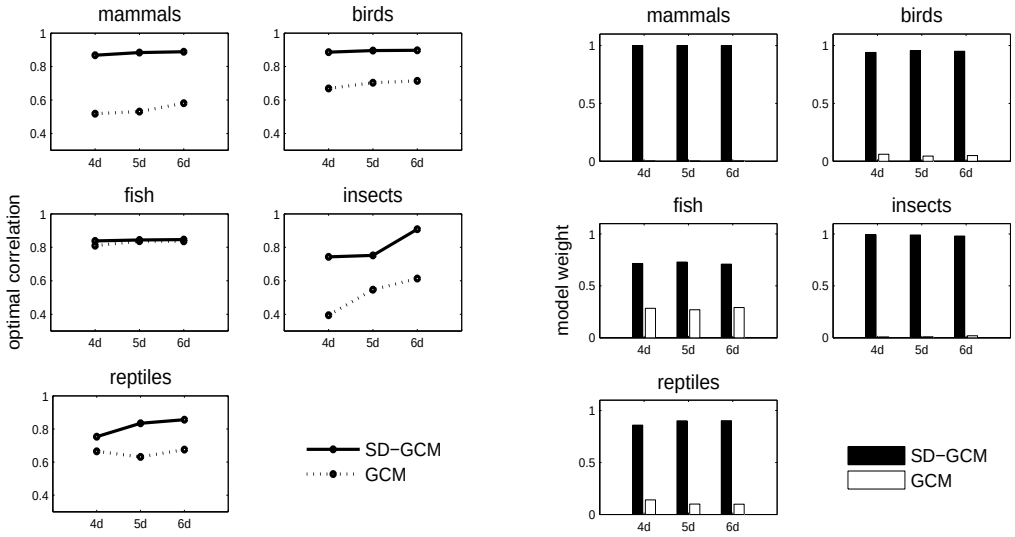
Goodness of fit. Results of the model fit analyses are presented in Figure 1(a). For each of the five animal categories, we calculated the optimal correlation between observed typicality ratings of a category's members and the model-based typicality scores of the members, derived from the GCM and the SD-GCM. For the SD-GCM, we included the most appropriate set of contrast categories for each target category, following the findings of the previous analyses. The optimal correlations are presented as a function of the dimensionality of the underlying stimulus representation.

The SD-GCM captures the observed typicality gradient of the categories very well, with optimal correlations between observed and model-based typicality scores rising well above .8 for stimulus representations of Dimensionality 6 (.88 averaged across categories). For *insects* in Dimensionalities 4 and 5, and *reptiles* in Dimensionality 4, the optimal correlation is just under .8. In contrast, the standard GCM has more difficulty in providing a good account of the typicality ratings. While for *birds* and *fish* the optimal correlations are good, rising above .7 and .8 respectively, the model has difficulty in the categories *insects* and *reptiles*, optimal performance dropping below .7 and in *mammals*, optimal performance dropping below .6.

In sum, the SD-GCM seems to be able to account for the typicality gradient of the animal categories. Moreover, the model provides equally good or better optimal fits to the observed typicality ratings than does the standard GCM. While providing an acceptable optimal fit is a necessary condition for a model to be taken seriously, deciding which model captured the data best cannot rely solely on goodness of fit, since inherent complexity differences between the models will bias the conclusion. Rather, model comparison should depend on a measure of generalizability (Myung, 2000), which balances the complexity in the models with their ability to fit the empirical data. [3] We relied on marginal likelihood as a measure of generalizability, which considers the averaged rather than optimal fit (Kass & Raftery, 1995).

Generalizability. Results of the generalizability analyses are presented in terms of model weights. The model weight reflects the relative evidence in favor of a particular model, given a set of models (Lee, 2004). The evidence for a model is quantified through marginal likelihood. To estimate the marginal likelihood, we rely on standard non-informative priors for the parameters. For the dimensional weights in both the GCM and SD-GCM, a uniform prior in the interval [0,1] is used, adding the restriction that the dimensional weights have to sum to 1. For the sensitivity parameter the choice of prior is less straightforward. We therefore ran the analyses using different priors: A Uniform distribution and an approximation to the Gamma(0,0) distribution (Lunn, Thomas, Best, & Spiegelhalter, 2000). All results were qualitatively nearly

[3]While the models that are compared in the present study are very similar in terms of number of free parameters and the nature of the parameters, differences in complexity can still arise from differences in functional form of the models



(a) goodness of fit

(b) generalizability

Figure 1: Optimal correlations between observed and model-based typicality scores (panel a) and model weights (panel b) for the GCM and the SD-GCM.

identical, regardless of the exact choice of the prior. We report the results of the Gamma prior with upper bound set at 10. The SD-GCM makes use of the same contrast sets for each target category as in the goodness of fit analyses.

Figure 1(b) confirms the results of previous analyses: For all categories tested, the SD-GCM clearly has better model weights, which indicates that the SD-GCM provides a more likely account of the data than does the GCM, also when considering potential differences in complexity between the models. This strongly suggests that indeed dissimilarity to exemplars from (a subset of) contrasting categories considerably improves the account of the typicality gradient of these categories.

General Discussion

In the present study we have tackled the controversial question whether contrast categories co-determine the internal graded structure of natural language categories, using computational models that have their roots in artificial category learning tradition. For five animal categories, we contrasted two exemplar models, the standard GCM and the SD-GCM in their account of typicality. According the standard GCM, typicality of an exemplar to a category is the summed similarity to all category members: The more a member is similar to many other members, the more typical it is of the category. According to the SD-GCM on the other hand, typicality of an exemplar derives from both the summed similarity towards all category members and the summed dissimilarity towards members of contrast categories: The more a member is similar to other members and *in the meantime* is dissimilar to the members of (a set of appropriate) contrast categories, the more that member will be typical of the category.

We found strong evidence for all categories tested, that dissimilarity towards the members of contrast categories improved the exemplar based account of typicality of the target categories. Interestingly, the SD-GCM performed best when we included a subset of the available contrast categories. For three categories – *fish*, *insects* and *reptiles* –, one contrast category was found to be optimal. For *mammals*, three contrast categories optimized the account of typicality and for *birds* all other animal categories were added as contrast categories. In sum, to return to our starting example, the present study provides clear-cut evidence that the platypus is considered to be an atypical member of the *mammal* category, not only because it is not very similar to the members of *mammals*, but also because it is similar to members of *birds*, *insects* and *reptiles*.

While the present results are in line with a number of findings in artificial category learning experiments (e.g., Davis & Love, 2010; Palmeri & Nosofsky, 2001), they directly contradict those of Verbeemen et al. (2001), who found no evidence for contrast category effects in typicality judgments. The difference in approach most likely can explain this difference. Verbeemen et al. (2001) used a family resemblance approach and regression analyses to evaluate the role of contrast categories. In the present study, we applied computational models that operate on geometric similarity representations and that are potentially better at accounting for typicality. This is reflected in the overall better quality of the optimal correlations in the present study as compared to Verbeemen et al. However, it must also be noted that the set of categories tested in the present study was rather small, and restricted to a single

semantic domain. Although our approach has proven to be sensitive to contrast effects, including more categories, from different semantic domains, in the analyses is imperative to make more general statements regarding the role of contrast information.

The choice for the domain of animals was motivated by the clear hierarchical structure, mutual exclusivity of the categories, and the complete coverage of the domain, in that all possible contrast categories could be added (if we restrict ourselves to the animal domain, following the definition of contrast categories). As such, the animal domain was a perfect arena to test for contrast effects. The present methodology can, however, be applied to other semantic domains. Obvious candidates are categories that are traditionally considered contrast categories, e.g., fruits and vegetables or carnivores and herbivores. Another interesting candidate domain is artifacts, yet a number of difficulties arise. For one, some members of the artifact domain can be classified in multiple categories (e.g., a knife can be a kitchen utensil, but also a weapon, or even a tool). Moreover, it is less clear how far the artifact domain extends: Are vehicles artifacts? Or clothes? Given that including appropriate contrast categories is essential, the extension of a domain requires careful consideration.

In a broader perspective, the present findings highlight the caution that needs to be undertaken when transposing models from artificial category learning to natural language category research. While in the former field a number of excellent models has been developed that carefully implement different views on category representation, these models might not always capture all aspects that are relevant in natural language categories. For example, the GCM is one of the most influential and most successful categorization models, but in its account of typicality (Nosofsky, 1988; Voorspoels, Vanpaemel, & Storms, 2008) it does not take into account the richness of the semantic domain in which a natural language category is embedded.

Acknowledgments

The Research in this article is part of research project G.0281.06, Belgian National Science Foundation Flanders. Thanks to Tom Waits for inspiration and musical company.

References

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

Davis, T., & Love, B. C. (2010). Memory for category information is idealized through contrast with competing options. *Psychological Science*, *21*, 234-242.

De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M. J., Voorspoels, W., et al. (2008). Exemplar by feature applicability matrices and other dutch normative data for semantic concepts. *Behavior Research Methods*, *40*, 1030-1048.

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

Goldstone, R., Steyvers, M., & Rogosky, B. (2003). Conceptual interrelatedness and caricatures. *Memory & Cognition 2003, 31 (2), 169180, 31(2)*, 169-180.

Hampton, J. (1993). Categories and concepts: Theoretical views and inductive data analysis. In I. Van Mechelen, J. Hampton, R. Michalski, & P. Theuns (Eds.), (pp. 67–95). Academic Press.

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

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

Komatsu, L. K. (1992). Recent views of conceptual structure. *Psychonomic Bulletin*, *112*, 500-526.

Kruskal, J., & Wish, M. (1978). *Multidimensional scaling* (E. M. Uslaner, Ed.). SAGE publications, inc.

Lee, M. (2004). A Bayesian analysis of retention functions. *Journal of Mathematical Psychology*, *48*, 310–321.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS: A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325-337.

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

Miller, G., & Johnson-Laird, P. (1976). *Language and perception*. Harvard University Press.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, *44*, 190-204.

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

Nosofsky, R. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 54–65.

Palmeri, T. J., & Nosofsky, R. (2001). Central tendencies, extreme points, and prototype enhancement effects in ill-defined perceptual categorization. *The Quarterly Journal of Experimental Psychology*, *54A*, 197-235.

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

Shepard, R. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

Stewart, N., & Brown, G. D. A. (2005). Similarity and dissimilarity as evidence in perceptual categorization. *Journal of Mathematical Psychology*, *49*, 403-409.

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.

Voorspoels, W., Vanpaemel, W., & Storms, G. (2008). Exemplars and prototypes in natural language concepts: a typicality-based evaluation. *Psychonomic Bulletin & Review*, *15*, 630–637.