A New Investigation of the Nature of Abstract Categories

Loes Stukken (Loes.Stukken@student.kuleuven.be)
Steven Verheyen (Steven.Verheyen@psy.kuleuven.be)
Matthew J. Dry (Matt.Dry@psy.kuleuven.be)
Gert Storms (Gert.Storms@psy.kuleuven.be)
University of Leuven, Department of Psychology
Tiensestraat 102, B-3000, Leuven Belgium

Abstract

Concrete and abstract categories are alike in that they show graded structure. They are also said to differ in that the degree of feature overlap between a category and its exemplars can predict the graded structure of the former, but not of the latter kind of category. We show how one can improve upon this prediction by taking distinctive features into consideration and argue that concrete and abstract categories might not be that differently structured after all.

Keywords: abstract categories; graded structure; typicality; characteristic features; polymorphous concept; familiarity.

Introduction

Well-defined categories (e.g., *odd numbers*, *plane geometry figures*, Armstrong, Gleitman, & Gleitman, 1983), ad hoc categories (e.g., *ways to make friends*, *things that could fall on your head*, Barsalou, 1983), abstract categories (e.g., *beliefs*, *crimes*, Hampton, 1981), and well-established, concrete categories (e.g., *fruit*, *vehicles*, Rosch, 1973) are alike in that they all present with graded structure. That is, some of their exemplars are consistently judged to be more typical of or representative for the category than others. Graded category structure also becomes apparent in category verification where the more typical an exemplar is, the faster it will be judged a true category member. Rosch (1973) was among the first to demonstrate this relationship in concrete categories. Armstrong et al. (1983) and Hough and Pierce (1989) showed it also holds in well-defined and ad hoc categories, respectively. In addition, typical exemplars are generally among the first to be generated in response to the category label and across participants they are generated more often than less typical category members (for early demonstrations in concrete and ad hoc categories see Mervis, Catlin, & Rosch, 1976 and Barsalou, 1985, respectively). Other processing advantages of typical over atypical category members have been shown in concrete categories, but attempts to demonstrate similar effects in well-defined or ad hoc categories have generally not been undertaken. Studies pertaining to the internal structure of abstract categories are even rarer.

The presence of graded structure in a varied range of categories need not necessarily indicate that these are all represented in the same manner. The most generally accepted account of graded structure to have arisen from the study of concrete categories is probably the one in which category members are represented by characteristic features. The more of these features an exemplar shares with the category as a whole, the more typical the exemplar is of the category (Hampton, 1979; Rosch & Mervis, 1975). As penguins do not have the features <*can fly*> and <*build nests in trees*> in common with most other birds, they are considered to be among the atypical category members. The internal structure of many ad hoc and well-defined categories does not adhere to this shared feature account, however. Barsalou (1985) illustrates how participants' estimates of the number of times they had previously encountered an item as a category member and the item's ability to fulfill the goal served by its ad hoc category, provide a better account of the item's judged typicality. Those food items that are often thought of as diet products and those that are ideally suited for minimizing one's intake of calories, Barsalou found to be the worst examples of the ad hoc category *foods not to eat on a diet* for instance. Larochelle, Richard, and Soulières (2000) ascribe well-defined categories' apparent graded structure in category verification to the exemplars' familiarity and category dominance. When the influences of these variables are controlled for, reliable categorization time differences no longer come about.

With categories considered the building blocks of cognition (Pinker, 1997) the question of which categories constitute truly distinct kinds (and which ones do not) becomes one of central importance in cognitive science. In what follows, we will review evidence that pertains to the possibly different nature of concrete and abstract categories. As we already indicated, studies addressing the internal structure of abstract categories are scarce, but nevertheless seem to suggest its origin differs from that of concrete categories. Especially a study by Hampton (1981) favours this conclusion. Hampton found that the polymorphous concept - a measure of the number of characteristic features shared by category and exemplar, originating from the concrete categories literature - did not prove adequate in predicting the graded structure of a number of abstract categories. We will show how one can improve upon this prediction by taking distinctive features into consideration and argue that concrete and abstract categories might not be that differently structured after all.

Kinds of Categories

Much of our theorizing about the nature of categories arises from the study of well-established, concrete categories like *fruit* or *vehicles*. The vast amount of work currently available on this subject matter, testifies to the fruitfulness of this particular approach to semantic cognition. Medin, Lynch, and Solomon (2000) nevertheless argue that extending the scope of studies to include other kinds of categories, is a praiseworthy endeavour. They consider three major reasons for doing so: (i) it allows for a test of the generality of prevailing theories, (ii) it might reveal interesting peculiarities of certain kinds of categories, and (iii) different categories may differ in the ease with which they allow the investigation of topics of interest. Of course, leaving the familiar domains of *apples* and *oranges* and of *bicycles* and *cars* for that of abstract beliefs as *evolution* and *patriotism,* will only yield these benefits if concrete and abstract categories are in fact shown to be truly distinct kinds of categories. Medin et al. discern two (interrelated) criteria that can be used to argue for the different nature of concrete and abstract categories. These criteria pertain to the processing differences and the structural differences between both types of categories[1].

Lakoff (Lakoff & Johnson, 1980; Lakoff & Turner, 1989) introduces the idea that abstract entities are processed through reference to more concrete entities. The notion of *anger* is traditionally held up as an example. Its meaning is said to be understood by referring to water that comes to a boil. According to this idea, abstract categories are processed differently from concrete ones, in that the latter ones do not require metaphorical mapping. Structural differences support such claims in that the features that are generated in response to abstract entities are less specific than the characteristic features concrete categories present with (Wiemer-Hastings & Xu, 2005). Additional structural differences come about in feature generation tasks. Abstract categories invoke more introspective features than concrete categories do, but do not activate as much entity features. When generating relational features for abstract categories, participants emphasize agents, actions, and social aspects. When generating relational features for concrete categories, participants mainly recall objects, living things, locations and functions (Barsalou & Wiemer-Hastings, 2005; Wiemer-Hastings & Xu, 2005).

A feature generation task is also at the basis of Hampton's investigation of the structure of abstract categories (Hampton, 1981). Hampton had participants generate characteristic features for eight abstract categories. Subsequently, he had a different group of participants judge whether these features were applicable to the categories' exemplars. The degree to which category and exemplars share features had proven to be predictive of exemplars' typicality in concrete categories (Hampton, 1979), but the so-called polymorphous concept measure did not predict rated typicality in all of the studied abstract categories.

[1] Medin et al. (2000) include content-laden principles as a third criterion, but we will not discuss it here since as far as we know no principled attempts have been made to distinguish concrete and abstract categories on the basis of the causal explanatory frameworks they invoke.

Aim

Dry and Storms (2009) have recently demonstrated how the polymorphous concept (Hampton, 1979, 1981) can be generalized to include both shared and distinctive features. In all but four of the eleven concrete categories studied by Dry and Storms allowing distinctive features to influence the prediction of rated typicality improved the predictive power of the polymorphous concept model. As Hampton (1981) is up until now the only study addressing the graded structure of abstract categories, we felt it important to attempt to replicate the findings and to establish whether the poor predictions by the polymorphous concept model for some of the abstract categories could not be attributed to exclusion of distinctive feature information. Note that this would not detract from the finding that the features that make up the polymorphous concept in concrete and abstract categories differ regarding the nature of the information they communicate (Barsalou & Wiemer-Hastings, 2005; Wiemer-Hastings & Xu, 2005). To the contrary, it would establish the generality of the model across featural content.

The Generalized Polymorphous Concept

The polymorphous concept model (Hampton, 1979, 1981) assumes that features are the representational units of semantic concepts and that the more of these features a particular exemplar and its category share, the more typical the exemplar is considered to be. If both the category and the exemplars are represented by binary feature vectors (v) of length k in which the presence or absence of each characteristic feature is signified by 1 and 0 respectively, then the typicality of an exemplar i with respect to category A can be formalized as:

$$PC(i, \mathrm{A}) = \sum_k v_{ik} v_{Ak} \tag{1}$$

Because of the multiplicative nature of Equation (1) only those features that are shared by i and A contribute to exemplar i's typicality. Earlier we already gave the example of the *penguin*, an animal considered to be atypical among birds as it cannot fly or does not build its nest in trees like other birds do. Penguins, however, also <*live on icy planes*> and <*have flippers*>. These are features that are distinct to the penguin and might plausibly impact on its perceived typicality as a bird. Dry and Storms (2009) have made the relevance of distinctive features for typicality judgments more salient by casting the polymorphous concept model in terms of category-exemplar similarity. From the work of Tversky (Gati & Tversky, 1984; Tversky, 1977) it has become apparent that both shared and distinctive features contribute to similarity. Following Tversky's contrast model (1977) the similarity s_{iA} between an exemplar i and a category A can be expressed as a weighted combination of the features that i and A share ($v_i \cap v_A$), the features that are distinct to i ($v_i - v_A$), and the features that are distinct to A ($v_A - v_i$):

$$s_{iA} = \left[\theta \sum_k v_{ik} v_{Ak}\right] - \left[(1-\theta)\sum_k v_{ik}(1-v_{Ak})\right] - \left[(1-\theta)\sum_k (1-v_{ik}) v_{Ak}\right], \tag{2}$$

where θ is a single parameter (ranging form 0 to 1) that weighs the contributions of shared and distinctive features (Navarro & Lee, 2004). Setting θ to low values emphasizes distinctive features, whereas setting θ to high values emphasizes common features.

Dry and Storms (2009) note that the first and third terms in Equation (2) are collinear. Consider the example feature vectors in Table 1. Exemplars i and j differ from one another in that exemplar i has one more feature in common with category A than exemplar j. Because i and A share the third feature, while j and A do not, $v_i \cap v_A$ equals 2, while $v_j \cap v_A$ equals 1. This difference is, however, also reflected in the fact that category A has one more distinctive feature in regards to exemplar j than it has in regards to exemplar i. Because of the same (category) feature that was missing in j, but not in i, $v_A - v_i$ equals 1 while $v_A - v_j$ equals 2. In Equation (2) $v_{ik}v_{Ak}$ and $(1-v_{ik})v_{Ak}$ thus act as communicating vessels. This has important implications for the category-exemplar similarity calculation in that even when θ is set to 0, s_{iA} will still be inversely affected by feature commonality if $v_A - v_i$ is included in the calculation. In order to address this Dry and Storms re-formulate Equation (2) by dropping the third term such that the expression for the generalized polymorphous concept becomes:

$$GPC(\theta, i, A) = \left[\theta \sum_k v_{ik} v_{Ak}\right] - \left[(1-\theta)\sum_k v_{ik}(1-v_{Ak})\right] \tag{3}$$

Table 1. Example feature vectors and corresponding indices of commonality and distinctiveness.

v_i	v_j	v_A	$v_i \cap v_A$	$v_A - v_i$	$v_j \cap v_A$	$v_A - v_j$
1	1	1	1	0	1	0
1	1	0	0	0	0	0
1	0	1	1	0	0	1
0	0	1	0	1	0	1

Study 1

In this section we start by describing the gathering of the materials necessary to allow a replication of Hampton (1981). First, we elaborate on the selection of abstract categories and their exemplars. Then we discuss the gathering of typicality ratings, features, and feature applicability judgments, which will allow a comparison of the predictions made by the traditional polymorphous concept (Hampton, 1979) with those by the generalized polymorphous concept (Dry & Storms, 2009).

Exemplar Generation

Seven categories were selected for inclusion in this study. All of them adhered to Hampton's (1981) definition of an abstract category in that they had as referents things that are not physical, concrete objects. As part of a large exemplar generation study that included 30 categories of a varied nature, 80 first year students of the University of Leuven produced eight exemplars for the categories *art forms*, *crimes*, *diseases*, *emotions*, *media*, *sciences*, and *virtues*. A tally was kept of the number of times a particular exemplar was generated in response to a category label. This tally informed the selection of exemplars. As previous work on concrete and ad hoc categories (Barsalou, 1985; Mervis, Catlin, & Rosch 1976) has established a strong relationship between generation frequency and typicality, we hoped to obtain exemplars that differed considerably in typicality by selecting 15 exemplars that spanned the range of generation frequencies for a particular abstract category.

Typicality Judgments

A group of 30 first year students of the University of Leuven provided typicality judgments for the 15 selected exemplars of each of the seven abstract categories. They received a booklet containing seven pages with on each page the 15 exemplars of one of the seven categories. The participants were asked to indicate for every item on the page how good an example it was of the category mentioned on top of the page. They were required to provide their answer by indicating a value on a scale ranging from 1 (a very bad example) to 20 (an excellent example). If they did not know one of the presented exemplars, they were asked to indicate this by drawing a circle around it. Every participant rated the typicality of all the exemplars of every category. There were two different presentation orders for the categories and three different presentation orders for the exemplars within a category. This resulted in six different booklets that were each completed by five participants.

The reliability of the typicality judgments within each of the categories was estimated using the split-half correlation with Spearman-Brown correction. The reliabilities for six of the categories lay between .94 and .98. The reliability for *diseases* proved somewhat lower with a value of .80. To confirm the validity of the exemplar selection procedure described above, the correlation between the typicality judgments and the generation frequencies was calculated for each category. The correlations were .76 for *art forms*, .52 for *crimes*, .49 for *diseases*, .79 for *emotions*, .84 for *media*, .73 for *sciences*, and .69 for *virtues*. All these correlations proved significant at the .05 level (one-tailed t). These results thus indicate that the commonly found relationship between typicality and generation frequency also holds in abstract categories.

Feature Generation

Thirty University of Leuven students completed a task intended to elicit the features associated with each of the categories. They answered three questions for each of the category labels. (i) Which features do you feel are important for this category? (ii) Which features have to be present for something to be considered a member of this category? (iii) Which features determine whether something is a better example of this category than something else? Every participant answered these questions for each of the seven abstract categories. Each participant was presented with a different order of categories. Participants could generate as many (or few) features as they felt necessary. Across participants and questions 46 different features were generated for *art forms*, 38 for *crimes*, 50 for *diseases*, 44 for *emotions*, 35 for *media*, 46 for *sciences*, and 40 for *virtues*.

Feature Applicability

The exemplars and features that were generated in response to the category labels were combined in an exemplar-by-feature matrix. The exemplars (in alphabetical order) made up the columns of the matrix, while the features (also in alphabetical order) made up the rows of the matrix. Hence, every exemplar-feature-pair was represented by a single cell in the matrix. Five University of Leuven students indicated for each exemplar-feature-pair whether the feature applied to the exemplar or not, by entering a 1 or a 0 in the corresponding matrix cell. Participants performed the task at home and could freely choose when they worked on it. They were given the choice to work on the task row-wise or column-wise, but they were asked not to pause until a row or column was finished. The reliability of the exemplar-by-feature judgments was again assessed using the split-half correlation with Spearman-Brown correction. There are only ten different ways to divide five participants into two half groups. The ten possible estimations of the reliability averaged .76 for *art forms*, .82 for *crimes*, .81 for *diseases*, .82 for *emotions*, .74 for *media*, .72 for *sciences*, and .77 for *virtues*, indicating considerable agreement between the five judges.

Model Analyses

The empirical data gathering now allows us to evaluate the predictive power of the (generalized) polymorphous concept with regards to the graded structure of abstract categories. The five exemplar-by-feature matrices were combined into a single matrix by determining the majority judgment for each cell. The resulting binary matrix provided the exemplar feature vectors v_{ik} that enter into Equation (3). The category feature vector v_{Ak} in Equation (3) held ones for those features that were generated in response to the category label in the feature generation task, and zeros for those that were not. For each of the seven abstract categories the generalized polymorphous concept was then implemented by varying θ from 0 to 1 in increments of .0005. For every value of θ the correlation with judged typicality was computed[2]. The bold lines in Figure 1 display for each category the correlation of the generalized polymorphous concept with judged typicality across the entire range of θ.

The results replicate those of Hampton (1981) in that the original polymorphous concept proves a good predictor of the graded structure of some of the abstract categories, but not of all. When θ equals 1, indicating that only those features that the exemplars and the category share are taken into account, the correlation reaches .88 for *media* and .72 for *art forms*, *crimes*, and *emotions*. The correlations for *diseases* (.42) and *virtues* (.04) do not reach significance at the .05 level, however (one-tailed *t*). The .48 correlation for *sciences*, albeit low, does. Importantly, Figure 1 also shows that the prediction of typicality in the latter categories can be improved considerably by taking distinctive features into account. Allowing θ to differ from 1 yields an increase in the correlation with typicality to .73 for *diseases* (θ = .61), .49 for *virtues* (θ = .37), and .69 for *sciences* (θ = .52). Unlike the correlations for θ = 1, these correlations are all significant. Some evidence for a contribution of distinctive features is also found for the categories *art forms* (θ = .90) and *emotions* (θ = .90) where the correlation increases with .02. Obviously, the increase in correlation observed for these categories is negligibly small when compared to the .31, .45, and .21 increase observed for *diseases*, *virtues*, and *sciences*. The latter improvements and corresponding θ values suggest that the inability of the original polymorphous concept model to predict the graded structure of some abstract categories might arise from its exclusion of distinctive features.

Study 2

In Study 1 we have shown that the generalized polymorphous concept allows good to excellent predictions of the graded structure of abstract categories by utilizing both shared and distinctive feature information. However, one could argue that the evidence found for an influence of distinctive features follows from the entanglement of typicality and familiarity. Several researchers have brought the existence of a relationship between typicality and familiarity under the attention for concrete categories (e.g., Malt & Smith, 1982; McCloskey, 1980). Larochelle, Richard, and Soulières (2000) make a similar claim for well-defined categories. Hampton (1981) speculates that familiarity of exemplars might influence the graded structure of some abstract categories to a large extent. If a relationship were to exist between an item's judged familiarity and the number of distinctive features that characterize it (Is an *elephant* familiar because it is one of the few mammals that has marked features such as a trunk and big, floppy ears?) and familiarity and typicality were to cohere in abstract categories, the influence of distinctive

[2] Using the individual exemplar-by-feature matrices as input to the model analyses and averaging across the resulting correlations yielded results that were virtually identical to the ones displayed in Figure 1.

features should considerably weaken or disappear when the shared variance between typicality and familiarity is partialed out. To test this we obtained familiarity ratings for each of the studied exemplars, used regression analyses to partial out the shared variance between the typicality and familiarity ratings, and subjected the resulting residuals to the model analyses employed in Study 1.

Familiarity Judgments

Familiarity judgments were obtained in much the same manner as the typicality judgments were. A group of 30 first year students of the University of Leuven received a booklet containing seven pages with on each page the 15 exemplars of one of the seven categories. The participants were asked to indicate for every item on the page how familiar they were with it. They were required to provide their answer by indicating a value on a scale ranging from 1 (never seen, heard, or used) to 7 (seen, heard, or used very often). As was the case in the typicality judgment task, there were two different presentation orders for the categories and three different presentation orders for the exemplars in a category. The judgments proved very reliable with Spearman-Brown corrected split-half correlations ranging from .92 for *emotions* to .98 for *media*. Only the reliability of *crimes* proved a little lower with an estimation of .81.

Model Analyses

Regression analyses were employed to partial out the shared variance between the typicality and familiarity ratings. The model analyses introduced earlier were then repeated but the generalized polymorphous concept predictions were now correlated with the residuals resulting from the regression analyses instead of with typicality. The dotted lines in Figure 1 display for each category the correlation of the generalized polymorphous concept with these residuals.

As can be seen from the figure, the procedure employed in Study 2 did not impact heavily upon the results. If anything, the evidence for an influence of distinctive features on abstract categories' graded structure became more apparent with *crimes* being the only category for which $\theta = 1$ resulted in the optimal correlation (r = .60). The other θ values that yielded optimal predictions were .78 for *art forms* (r = .67), .57 for *diseases* (r = .62), .87 for *emotions* (r = .78), .82 for *media* (r = .70), .61 for *sciences* (r = .74), and .27 for *virtues* (r = .60). These results are most likely due to the weak relationship between the familiarity and the typicality judgments for the majority of the categories. Only for *art forms* (r = .53) and *media* (r = .60) did the correlation reach significance at the .05 level (two-tailed t) and as a result affected the shape of the curve in Figure 1. It would appear that familiarity does not impact strongly upon typicality and can therefore not be held responsible for the influence of distinctive feature information on abstract categories' graded structure.

Summary and Discussion

Abstract categories can be considered different from concrete categories. There is evidence to suggest that they are processed differently (Lakoff & Johnson, 1980; Lakoff & Turner, 1989) and there are said to be structural differences between them (Barsalou & Wiemer-Hastings, 2005; Wiemer-Hastings & Xu, 2005). It is questionable, however, whether they also differ with regard to the origin of their graded structure. Hampton (1981) found that the polymorphous concept - a measure of the number of characteristic features shared by category and exemplars -

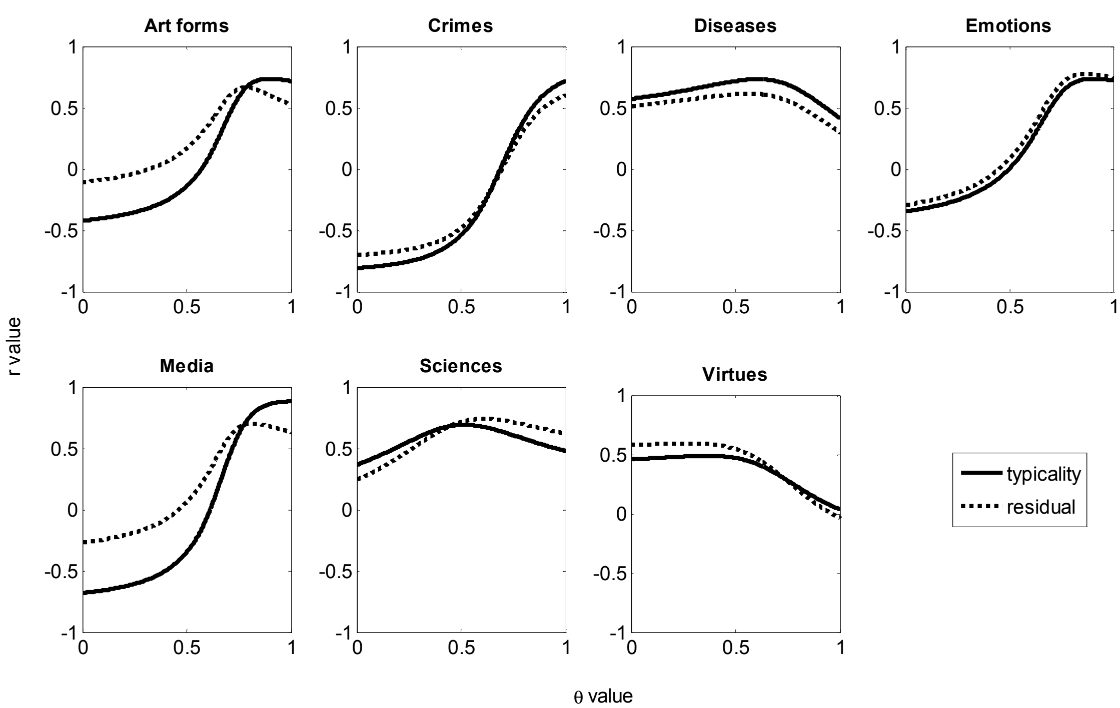

Figure 1: Correlation between graded structure and generalized polymorphous concept across θ values.

did not prove adequate in predicting the graded structure of a number of abstract categories, while it did do a good job predicting the graded structure of concrete categories (Hampton, 1979). However supportive Hampton's results might seem for the proposition that abstract categories are a structurally different kind of categories than concrete ones, the shared feature model does not fare equally well in all concrete categories either (Hampton, 1979 reports Kendall's τ values that range from .61 to .78). In addition, Dry and Storms (2009) have generalized the polymorphous concept to include both shared and distinctive features. In seven of the eleven concrete categories studied by Dry and Storms allowing distinctive features to influence the prediction of rated typicality improved the predictive power of the polymorphous concept. The resulting predictions nevertheless still varied considerably, with Pearson correlation coefficients ranging from .47 to .91. Here we applied the generalized polymorphous concept to seven abstract categories and found similar results. In five of the categories the optimal correlation with typicality was obtained through a weighted combination of shared and distinctive features. As was the case for the concrete categories, the resulting correlations varied considerably, with values ranging from .49 to .88. This suggests that the graded structures of concrete and abstract categories do not necessarily have a different origin, but opens up the question of what might account for the differences found amid concrete and abstract categories.

Acknowledgments

This research was supported in part by grants G.0513.08 of the Research Foundation – Flanders and OT/05/27 of the Leuven University Research Council awarded to Gert Storms. The two first authors contributed equally to this manuscript. They are listed alphabetically. Steven Verheyen is a research assistant at the Research Foundation - Flanders.

References

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). What some concepts might not be. *Cognition, 13*, 263-308.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition, 11*, 211-227.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 629-654.

Barsalou, L. W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher & R. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thought.* New York: Cambridge University Press.

Dry, M. J., & Storms, G. (2009). Features of graded category structure: Generalizing the family resemblance and polymorphous concept models. *Manuscript submitted for publication.*

Gati, I., & Tversky, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgements. *Cognitive Psychology, 16*, 341-370.

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

Hampton, J. A. (1981). An investigation of the nature of abstract concepts. *Memory & Cognition, 9*, 149-156.

Hough, M. S., & Pierce, R. S. (1989). Exemplar verification for common and ad hoc categories in aphasia. In T. Prescott (Ed.), *Clinical Aphasiology*. Austin, TX: PRO-ED.

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science, 4*, 195-208.

Lakoff, G., & Turner, M. (1989). *More than cool reason: A field guide to poetic metaphor.* Chicago, IL: University of Chicago Press.

Larochelle, S., Richard, S., & Soulières, I. (2000). What some effects might not be: The time to verify membership in "well-defined" categories. *The Quarterly Journal of Experimental Psychology, 53A*, 929-961.

Malt, B. C., & Smith, E. E. (1982). The role of familiarity in determining typicality. *Memory & Cognition, 10*, 69-75.

McCloskey, M. (1980). The stimulus familiarity problem in semantic memory research. *Journal of Verbal Learning and Verbal Behavior, 19*, 485-502.

Medin, D. L., Lynch, E. B., & Solomon, K. O. (2000). Are there kinds of concepts? *Annual Review of Psychology, 51*, 121-147.

Mervis, C. B., Catlin, J., & Rosch, E. (1976). Relationships among goodness-of-example, category norms, and word frequency. *Bulletin of the Psychonomic Society, 7*, 283-284.

Navarro, D. J., & Lee, M. D. (2004). Common and distinctive features in stimulus similarity: A modified version of the contrast model. *Psychonomic Bulletin & Review, 11*, 961-974.

Pinker, S. (1997). *How the mind works.* New York, NY: Penguin Books.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive Development and the Acquisition of Language*. New York, NY: Academic Press.

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

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

Wiemer-Hastings, K., & Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive Science, 29*, 719-736.