Prototype Models of Categorization: Basic Formulation, Predictions, and Limitations

John Paul Minda
Department of Psychology
The University of Western Ontario

J. David Smith
Department of Psychology and Center for Cognitive Science
University at Buffalo, the State University of New York

Abstract

The prototype model has had a long history in cognitive psychology and prototype theory posed an early challenge to the classical view of concepts. Prototype models assume that categories are represented by a summary representation of a category (i.e., a prototype), that might represent information about the most common features, the average feature values, or even the ideal features of a category. Prototype models assume that classification decisions are made on the basis of how similar an object is to a category prototype. This chapter presents a formal description of the model, the motivation and theoretical history of the model, as well as several simulations that illustrate the model's properties. In general, the prototype model is well-suited to explain the learning of many visual categories (e.g. dot patterns) and categories with a strong family-resemblance structure.

Categories are fundamental to cognition, and the ability to learn and use categories is present in all humans and animals. An important theoretical account of categorization is the prototype view (Homa, Cross, Cornell, & Shwartz, 1973; Homa & Cultice, 1984; Minda & Smith, 2001, 2002; Posner & Keele, 1968; J. D. Smith & Minda, 2001; J. D. Smith, Redford, & Haas, 2008; J. D. Smith & Minda, 1998, 2000). The prototype view assumes that a category of things in the world (objects, animals, shapes, etc.) can be represented in the mind by a prototype. A prototype is a cognitive representation that captures the regularities and commonalities among category members and can help a perceiver distinguish category members from non-members. The prototype of a category is often described as the central tendency of the category, as a list of frequently occurring features, or even as an ideal category member. Furthermore, the prototype is similar to category members within the

This work was completed with the assistance of a grant from the National Science and Engineering Research Council of Canada (Minda) and grant HD-38051 from the National Institute of Child Health and Human Development (Smith)

category and less similar (or very dissimilar) to members of other categories. According to the prototype view, objects are classified by first comparing them to the prototypes that are stored in memory, evaluating the similarity evidence from those comparisons, and then classifying the item in accord with the most similar prototype.

The prototype view can be realized as a computational model (i.e. the prototype model) that enables a researcher to make specific predictions about the category membership of novel exemplars within a prototype-based framework. The prototype model has been influential in categorization research for several decades as a complementary and balancing perspective to exemplar theory. In this chapter, we present a detailed description of the prototype model (Minda & Smith, 2001; J. D. Smith & Minda, 1998, 2000), we review the historical development of the prototype model, and we present several key predictions of the prototype model.

Description of the Model

In this section, we provide a basic formulation of how the prototype model calculates similarity and makes a classification decision (Nosofsky, 1992; Minda & Smith, 2001). The formulation of the basic prototype model is closely related to the *Generalized Context Model* of Nosofsky (Nosofsky, 1986, 1987) which is covered in Chapter 2 of this volume. Of course, the key difference is that to-be-categorized items are compared to prototypes, rather than multiple, specific exemplar traces as in the *Context Model.* The prototype model makes a classification decision in two steps: comparison and decision. In the comparison phase, a to-be-classified item is compared to the stored prototypes (usually calculated as the modal or average feature values) and the psychological distance between them is converted to a measure of similarity. In the decision phase, the model calculates the probability of the item's category membership based on the similarity of the item to one prototype divided by the similarity of the item to all the prototypes.

The model can be formulated with three equations. First, the distance between the item i and the prototype P is calculated by comparing the two stimuli along each weighted dimension k (see Equation 1).

$$d_{iP} = \left[\sum_{k=1}^{N} w_k |x_{ik} - P_k|^r \right]^{1/r} \tag{1}$$

In this case, a dimension usually corresponds to some variable feature (e.g., if a set of stimuli appear as either green or blue, colour would be a dimension)[1]. The value of r is used to reflect two common ways to calculate distance. When $r = 1$ the model uses a city-block distance metric which is appropriate for separable-dimension stimuli. When $r = 2$ the model uses a Euclidean distance metric which is appropriate for integral-dimension stimuli. All of the simulations in this chapter use stimuli with separable dimensions and so r can be set to 1. Each dimension can be weighted to reflect how much attention or importance it is given by the model. In the present case, each attentional weight (w) varies between 0.0

[1] Most of the work with this model or related models like the GCM assumes that the dimensions exist in a psychological space that is representative of physical space. The dimensions of this psychological space can be derived from similarity scaling studies, or by making a simplifying assumption that each perceptual component will be interpreted as a dimension.

(no attention) and 1.0 (exclusive attention). Attentional weights are normally constrained to sum to 1.0 across all the dimensions. The results of these weighted comparisons are summed across the dimensions to get the distance between the item and the prototype.

This distance (d_iP) between the item and the prototype is then converted into a measure of similarity (η_{iP}), following Shepard (1987), by taking:

$$\eta_{iP} = e^{-cd_{iP}} \tag{2}$$

which gives a measure of similarity of an item i to a prototype P. It is the exponent in Equation 2 that allows for the exponential-decay of similarity (meaning that trait dissimilarities tend to decrease psychological similarity very steeply at first, and then more gradually later on) and allows for the close correspondence between the prototype model and the *Generalized Context Model* of Nosofsky (Nosofsky, 1992). The exponent is distance d_iP multiplied by the scaling or sensitivity parameter c. This parameter is a freely-estimated parameter that can take on values from 1 to ∞ and reflects the steepness of the decay of similarity around the prototype. Low values of c indicate a gradual, more linear decay. High values of c indicate a steep, exponential decay. Generally, higher values of the sensitivity parameter will result in stronger category endorsements for typical items and lower values of c will result in classification probabilities that are closer to chance.

The process of item-to-prototype comparison is repeated for all the prototypes (in this case P_A and P_B, but typically one to four in experimental settings). Once the item has been compared to the prototypes the probability of a Category A response is calculated for each stimulus. Prototype A similarity (η_{iP_A}) is divided by the sum of Prototype A and Prototype B similarity to generate the model's predicted probability of a Category A response ($P(R_A)$) for stimulus (S_i) as shown in the probabilistic choice rule in Equation 3.

$$P(R_A|S_i) = \frac{\eta_{iP_A}}{\eta_{iP_A} + \eta_{iP_B}} \tag{3}$$

This is the standard version on the model, and the one that was used by Nosofsky and colleagues to argue in favor of exemplar theory and that was used by Smith and Minda to argue against exemplar theory and in favor of prototype theory (Nosofsky, 1992; J. D. Smith, Murray, & Minda, 1997; J. D. Smith & Minda, 1998, 2000). The basic prototype model makes precise, prototype-based predictions about stimuli, and can be used to estimate the effectiveness of the prototype view in comparison to other computational accounts. Fitting the model involves parameter estimation and is described in the "Implementation" section. In later work, Smith and Minda considered an alternative model that was prototype based, but included an exemplar-memorization process as well (Minda & Smith, 2001, 2002; J. D. Smith & Minda, 1998, 2000). This chapter is primarily concerned with prototype-based processing and readers may wish to consult these other papers for work on the mixture model.

Motivation

The modern version of the prototype model can trace its history back to several key developments in cognitive psychology. The first of these was the influential dot-pattern research of Posner and Keele, and Homa and colleagues (Homa et al., 1973; Homa &

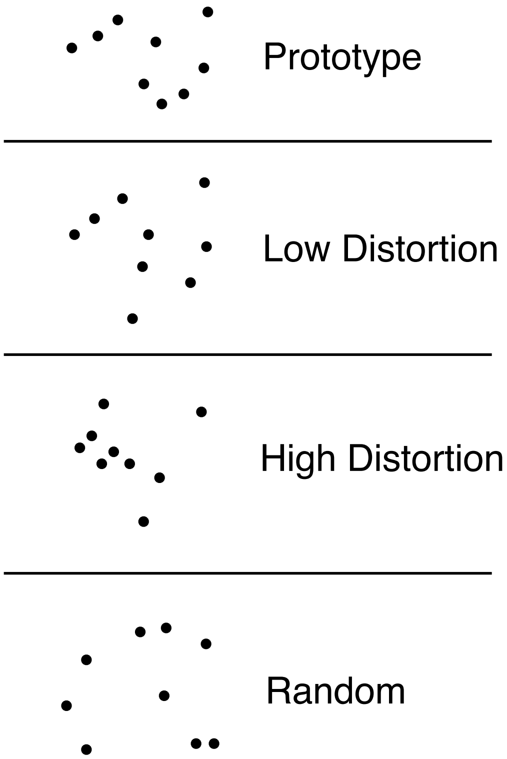

Figure 1. This figure shows an example of four kinds of dot-pattern items. The prototype is the original configuration of dots. Low-distortion items result from a smaller probabilistic move for each dot and the high-distortion items result from a larger probabilistic move for each dot. The random items are not related to the prototype and are a new arrangement of the nine dots.

Cultice, 1984; Posner & Keele, 1968). In a series of elegant experiments, subjects were shown distortions of a dot pattern or polygon (i.e., the prototype). The details for the creation of the stimuli can be found in any of the papers, and an example can be seen in Figure 1. In the figure, the prototype is shown at the top. The distortions were similar to, but not exactly like, the originating prototype. To create the distortions, each dot was subjected to a probabilistic function to determine whether it would keep the same position it had in the prototype and, if not, how far its position would change. Small adjustments of the location of some dots resulted in items that were “low distortions” of the originating prototype, and larger adjustments resulted in “high distortions”.

Subjects were generally trained on high-distortion items. Crucially, subjects were never shown the prototype during the training session. Later, during a test phase, subjects were usually shown the old patterns, some new distortions of varying levels of typicality, and the originating prototype. Studies using these dot patterns have generally found consistent results. First, subjects often performed as well on the prototype as they did on the old

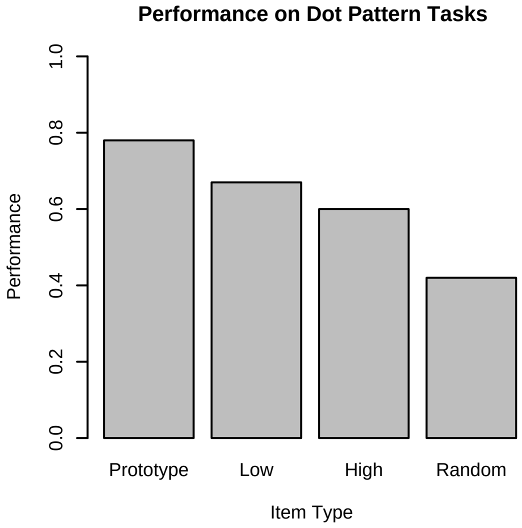

Figure 2. This figure shows the average dot pattern performance by control subjects from Knowlton and Squire (1993) along with the subjects in two papers by Reber and colleagues (Reber et al., 1998a, 1998b). The performance on the prototype pattern is best, followed by performance on the low distortions (which are most like the prototype), the high distortions and the random items.

patterns, even though the prototype was originally unseen. Second, if the test was delayed by several hours or days, performance on the training items declined whereas performance on the prototype remained strong (or declined less). Finally, the endorsement of new items showed a predictable typicality effect (like that shown in Figure 2), such that items that are physically closer to the prototype are endorsed more strongly as category members than items that are physically more distant (Homa et al., 1973; Homa & Cultice, 1984; Knowlton & Squire, 1993; Posner & Keele, 1968; Reber, Stark, & Squire, 1998b, 1998a; J. D. Smith & Minda, 2001; J. D. Smith et al., 2008). In addition, some of this work suggested that prototypes were especially important for larger categories (Homa & Cultice, 1984).

One of the most important contributions of this research program was the notion that the prototype is abstracted from experience with individual exemplars. By this account, there is no need to store every training exemplar, but the average of the exemplar experience is stored and used for subsequent classification decisions. Not surprisingly, the theoretical work with dot-pattern stimuli has generally favoured prototype theory (Ashby & Maddox, 2005; J. D. Smith & Minda, 2001; J. D. Smith et al., 2008).

A second key development in cognition was the influential work in the 1970's of Eleanor Rosch (Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Rosch followed Wittgenstein (Wittgenstein, 1958/2001) by introducing to cognitive psychology the idea of "family resemblance" as an alternative to the classical rule-based models that were dominant at the time. Rosch argued that for many categories, the prototype was an abstract representation with the highest family resemblance to other category members. In some cases this prototype might correspond to an actual category member, but in other cases it is simply a person's best idea about a category. Rosch's use of prototype-centered categories seemed uniquely able to explain the strong typicality gradients that characterized many natural-kind categories. Rosch's work, in conjunction with Posner's and Homa's research, provided the groundwork for prototype theory's dominance in the 1970's and 1980's.

An early mathematical formulation of the prototype model was provided by Edward Smith and Doug Medin in their book *Categories and Concepts* (1981). In this text, the authors presented the prototype model as a plausible alternative to the classical view. The classical view assumed that concepts were essentially defined by necessary and sufficient conditions. Unlike the classical view, the prototype model accounted for the typicality effects observed by Rosch and others. The prototype model's dominance waned, however, as research by Medin, Nosofsky, and others (Medin & Schaffer, 1978; Medin & Schwanenflugel, 1981; Nosofsky, 1986, 1987, 1988) suggested that exemplar models such as the *Generalized Context Model* sometimes provided a better account of categorization behavior than did the prototype model. The prototype model was generally regarded to be inferior to the exemplar model during this period.

It is critical to point out three aspects of that later research that should have tempered the theoretical climate of that era. First, researchers were often relying on small categories with only a few exemplars in their studies. These categories were easily memorizable as specific exemplars, and moreover they were repeated dozens of times during category learning, so participants had little impetus to abstract prototypes. Second, the categories in use at the time were poorly structured with weak family-resemblance relationships, so that once again participants had little to gain from prototype abstraction. Indeed, research by Blair and Homa (2004) showed that in one of the most commonly used category sets, participants appeared not even to categorize, but simply to memorize the frequently recurring specific items.[2] Third, it was the practice of the time to model the performance of whole subject groups. But it is now known that this homogenizes performance profiles in a way that disfavors the prototype model for non-psychological reasons. For example, Smith et al. (1997) used simulations to demonstrate that the common practice of averaging data together nearly always produced performance that was a better match to the predictions of the exemplar model, even when the average was created from known individual prototype-based performances.

Accordingly, the prototype model has enjoyed renewed interest since the late 1990's. In a series of papers, Smith and Minda showed that the prototype model often provided a better account of categorization behavior than the exemplar model (or rule-based models). For example, Smith, Murray, and Minda (1997) found that prototype models often had the advantage over an exemplar model in fitting the data from individual subjects (not whole groups). They found that averaging the individual data together can smooth away the steeper typicality gradients associated with prototype-based performance and present the models with shallower typicality gradients of the kind usually produced by exemplar-based performance. This was true even when the average samples contained known prototype-based performances. Of course the fit of a model to an individual's data is the appropriate level of analysis in the psychological study of category learning. Other research demon-

[2] Blair and Homa demonstrated that there was no advantage for category learning verses individual-stimulus identification learning for the Medin and Schaffer 5–4 stimuli.

strated that when fitting data at earlier stages of leaning, the prototype often fit better than the exemplar model (J. D. Smith & Minda, 1998). Smith and Minda found that early in the learning of certain kinds of categories known an "non linearly separable" categories, subjects often miscategorized exception items, even while they performed very well on more prototypical items. The prototype model fit this data pattern better than the exemplar model, because the prototype model predicts a linear separability constraint. Early in learning, many subjects showed evidence of this linear separability constraint. However, after many practice trials, performance on the exception items improved and the exemplar model tended to fit better than the prototype model. In both of these examples, as well as others (Minda & Smith, 2001), the advantage for the prototype model was strongest when the categories being learned were larger and well-differentiated.

The prototype model (and prototype theory in general) continues to influence the field. For example, recent work with the prototype model has shown that it can be augmented with some exemplar memory to account for recognition of specific items (J. D. Smith & Minda, 2000). Other work has shown that subjects learning via inference and prediction, instead of classification, may show stronger prototype effects when compared to classification learners (Chin-Parker & Ross, 2004; Minda & Ross, 2004; Yamauchi & Markman, 1998). Researchers have also demonstrated that the the prototype model provides a satisfactory account of category learning in humans and nonhuman species (J. D. Smith & Minda, 2001; J. D. Smith et al., 2008). Finally, evidence from the field of cognitive neuroscience has suggested that prototype learning may be strongly tied to specific areas in the visual cortex (Ashby & Maddox, 2005; Reber et al., 1998b; Zeithamova, Maddox, & Schnyer, 2008).

Implementation

In this section we show how the basic prototype model can be used to create specific predictions for individual stimuli in a category set, and how simulations can reveal basic properties of prototype-based categorization. We also describe how the model can be fit to observed data.

Generating Predictions with the Prototype Model

The prototype model can be used to make predictions about what kind of performance is expected for some category set, or it can be used to fit the data collected from experimental studies. In either case, the fitting work is done by the attentional weight parameters and the scaling parameter c. The power metric r also plays a role in how the model works, but is typically set before any model fitting is done. In any event, we do not consider its use here, because we examine only separable dimension stimuli and we set $r = 1$.

To understand how the prototype model generates predictions for the stimuli in a categorization task, consider the influential category set used by Medin and Schaffer (Medin & Schaffer, 1978) and shown in Table 1. Category A has five training items and Category B has four training items. Each item is made up of four binary dimensions that correspond to features. In the table, these are shown as 0's and 1's but can be instantiated in an experiment as colors, shapes, sizes, orientations, and so forth. In addition to the nine training items, the category set also has seven transfer items. In an experiment, subjects would be trained on the nine training stimuli (usually trial by trial with feedback after each response) and would then enter the transfer phase in which they would make a classification decision for all the old and new items. Because there are two categories, the prototype model assumes that there are two prototypes. For Category A, we can define the prototype as 1 1 1 1, because those values will represent the most frequently occurring features in the A category. Notice that the Category A prototype appears as Stimulus 12. The prototype for Category B can be defined as 0 0 0 0, because this is the item opposite to the prototype for Category A, and closely matches the most frequently occurring values for that category. Alternatively, Category B can be defined as 0 ? 0 0, because the 0-value of the second dimension has low category validity for Category B. Notice that the Category B prototype appears as Stimulus 9.

Because the model assumes that the prototypes are the only reference standards, items that are closer to the prototype will generally receive a stronger endorsement. For example Stimulus 1 (1 1 1 0) shares three of four features with its prototype (1 1 1 1) and only 1 feature with the Category B prototype. As a result, the prototype model assumes that this item will tend to be classified as a Category A member. On the other hand, Stimulus 2 (1 0 1 0) shares one half of its features with the Category A prototype (and it shares as many features with the prototype for category B). As a result, the prototype model will predict lower performance on Stimulus 2 relative to Stimulus 1 (or equal performance if no attention is paid to Dimension 2). Regardless of how the attention is allocated, the arrangement of these two stimuli and their relationship to their respective prototypes ensure that Stimulus 2 can never be endorsed as a Category A member more strongly than Stimulus 1. This result is a byproduct of prototype-based responding.

As an example of how the model makes a prediction, assume first that the model adopts a homogenous attentional profile such that each attentional weight is .25 (.25 for each of the four dimensions, summing to 1.0). This is just an example, because any attentional configuration is allowed as long as the weights sum to 1.0. The model compares each stimulus to the A and B prototype. For example, Stimulus 1 is 1 1 1 0 and the prototype for A is 1 1 1 1. Equation 1 results in dimensional distances of 0, 0, 0, and 1, which are multiplied by the weights to arrive at 0, 0, 0, and .25. These are summed to get a distance of .25. Equation 2 multiples the distance by the scaling parameter (assume 2.0 for now) and calculates the similarity between Stimulus 1 and the Category A prototype as being .61. The same procedure calculates the similarity of Stimulus 1 to the B prototype (0 0 0 0) as .22. The model then uses Equation 3 to generate the probability of Category A membership as being .73. This is intuitive, since the item shares 3/4 features with its own prototype and 1/4 features with the prototype in the opposite category. A different configuration of attentional weights produces a different result. If the weights .50 .20, .20, .10 are used, the similarity of the items to the A prototype is .82 and the similarity of the item to the Category B prototype is .16. The model calculates the probability of Category A membership as .83, once again an intuitive value because now the discrepant 4th feature has been relatively underweighted in attention.

A simple way to simulate performance on a given category set is to choose a large number of attentional weights and scaling parameter values and to generate a prediction for each set. For example, using the category set in Table 1, we can select 50,000 configurations of the standard prototype model. Each configuration is a set of four attention weights and

Table 1: Category Set from Medin and Schaffer (1978)

Stimulus	d1	d2	d3	d4
Category A				
1	1	1	1	0
2	1	0	1	0
3	1	0	1	1
4	1	1	0	1
5	0	1	1	1
Category B				
6	1	1	0	0
7	0	1	1	0
8	0	0	0	1
9	0	0	0	0
Transfer				
10	1	0	0	1
11	1	0	0	0
12	1	1	1	1
13	0	0	1	0
14	0	1	0	1
15	0	0	1	1
16	0	1	0	0

a value of the scaling parameter. For this simulation, we bounded the c parameter between 1.00 and 10.00, though in principle higher values are possible. Averaging across the 50,000 sets of predictions will give a general prediction for each stimulus. The resulting predictions for a simulation like this are shown in Figure 3.

Figure 3 can also be used to illustrate two points about the prototype model. First, notice that each of the stimuli is classified in accordance with its family resemblance to its prototype. This results in strong typicality effects. Regarding the B category, notice that Stimulus 9 (0 0 0 0, the Category B prototype) is categorized as A only .035 of the time—that is, it is correctly assigned to Category B .965 of the time. In contrast, Stimuli 6 and 7 (1 1 0 0 and 0 1 1 0, each with only two features characteristic of Category B), are only correctly assigned to Category B .499 and .501 of the time, respectively). This is a typicality gradient of about 46%. Regarding Category A, the obedience of the prototype model to family resemblance also results in the predicted advantage for Stimulus 1 over Stimulus 2 (Medin & Schaffer, 1978; J. D. Smith & Minda, 2000). Stimulus 1 is more prototypical than Stimulus 2 and across 1000's of configurations of the prototype model, it receives a stronger Category A prediction. This is a fundamental prediction of the prototype model and it comes about because of the model's basic assumptions about category representation. Any model with a prototype as the representational core and comparison standard must make this prediction. A model that assumes an alternative kind of representation (exemplar based, rule and exception, or clustering) is not required to make this prediction.

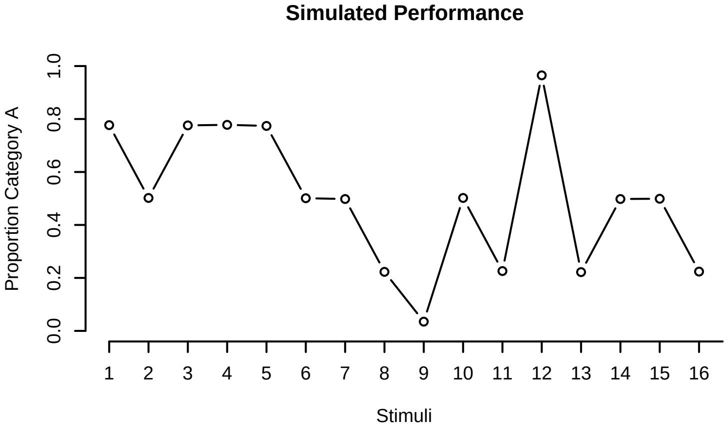

Figure 3. The prediction of simulated prototype-based participants for the Medin and Schaffer (1978) category set. Each point represents the average prediction of 50,000 configurations of the prototype model.

The second main point concerns the predictions for Stimuli 9 and 12. Both are prototypes for their categories (B and A), and they receive the lowest and highest Category A predictions, respectively. The prototype enhancement effect is another core prediction of the prototype model (see also J. D. Smith & Minda, 2000). Again, this prediction comes about because of the assumptions of the model and the fact that only prototypes are referenced for categorization decisions (i.e., presented prototypes will have perfect similarity to their prototype category representations). Together, the strength of typicality gradients and the size of prototype-enhancement effects have become critical diagnostic tools in analyzing whether, and when, humans and nonhuman animals abstract prototypes in category learning.

The Effects of the Scaling Parameter

Simulations can also reveal other properties and predictions of the model. For example, recall from the earlier section that the model makes use of a similarity scaling parameter c. This scaling parameter is a freely-estimated parameter that can take on values from 1 to ∞ and reflects how discriminable each prototype is from the other prototypes in psychological space. Low values of c indicate that the prototypes are not distinct from each other; higher values of the sensitivity parameter magnify the psychological space and increase the steepness of the similarity gradient around the prototypes. We can illustrate the effects of different values of c by conducting the following simulations. Again we use the stimuli from Medin and Schaffer (1978) shown in Table 1. We can select 30,000 configurations of the standard prototype model. Each configuration is a set of four attentional weights and a value of the scaling parameter. For the first 2000 configurations, we set the value of c at 1.00 and then we average across the resulting 2000 sets of predictions to find the average prediction for the prototype model with $c = 1.00$. Next we do the same with 2000

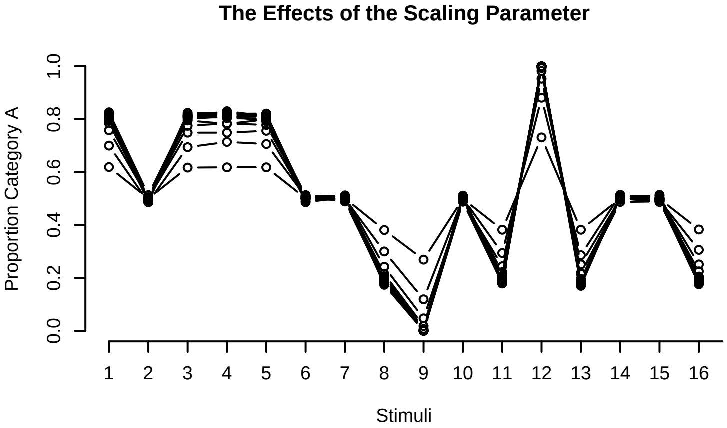

Figure 4. The results of 15 simulations of the prototype model on the Medin and Schaffer (1978) category set. Each line represents the performance profile of 2000 configurations of the prototype model with c increasing from 1.00 to 15.00. Performance increases with higher values of c for prototypical items but not for the less prototypical items.

configurations and $c = 2.00$ and continue until $c = 15.00$. In this way, we can examine the predictions for 15 increasing values of c (Figure 4).

Notice that there is a line of predictions running around the .40 -.60 range that corresponds to $c = 1.00$. This is what happens when a category set is represented in a low-sensitivity psychological space in which the prototypes are not very distinct from each other. Though stimuli may be fairly close to one prototype, they will not be that far from the other prototype in a low-sensitivity space, and categorization will remain uncertain with resulting weak category endorsements. However, as the scaling parameter's value increases, so does the differentiation between the prototypes and the typicality gradient around them. For example, performance on Stimulus 1 increases to above .80. Performance on the prototypes (Stimuli 9 and 12) increases even more to 1.0 (or 0.0). On the other hand, for less typical stimuli like Stimulus 2, the increases are very small. In other words, increasing the scaling parameter's value increases performance on items proportionally to their prototypicality. Furthermore, increasing c changes the character of performance, but eventually, the model reaches a settling point.

Fitting the Model to Observed Data

This method of using the attention weights and the scaling parameter to run simulations can be used to generate predictions for any stimulus set. However, the model can be used with greater precision to fit observed data. In this case, the researcher has typically collected data from a set of subjects, and wishes to determine if a model can account for the performance, often in comparison with other models. Model fitting involves estimating and adjusting the model's parameters in order to generate predictions that approximate the observed data. There are several possible algorithms for parameter estimation, but most will find the same set of best-fitting parameters. In this chapter, we discuss a hill-climbing algorithm that minimizes the Root Mean Square Deviation ($RMSD$) between the observed data and the model's predictions (Minda & Smith, 2001; J. D. Smith & Minda, 2000). Other methods of model fitting that have been commonly used with the prototype model are minimizing the sum of squared deviations and maximizing the log-likelihood. The model fitting index, the $RMSD$ or another index, can be used to compare and evaluate alternative models (see Myung, 2000, for details on model fitting). [3]

To find the best-fitting parameter settings of each model, a single parameter configuration (attention weights and c) is chosen at random and the predicted categorization probabilities for the stimuli in an experiment are calculated according to that configuration. The $RMSD$ is calculated as shown in Equation 4.

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}(O_i - P_i)^2}{N}} \tag{4}$$

The $RMSD$ between the observed (O_i) and predicted (P_i) probabilities is then minimized with a hill-climbing algorithm that makes a small adjustment to the provisional best-fitting parameter settings and chooses the new settings if they produce a better fit (i.e., a smaller $RMSD$ between predicted and observed performance). During each iteration of the algorithm, a parameter and a directional change are chosen at random. These changes are usually very small: gradations of 1/100, and they always respect the upper and lower bounds of the parameters. To ensure that attentional weights always sum to 1.0, the weight parameters are always adjusted in randomly chosen complementary pairs (e.g., when increasing the weight on dimension 1 by .01 the algorithm must decrease the weight on dimension 2 by the same amount). The hill-climbing algorithm continues to adjust the weights and the scaling parameter until no change can produce a better fit. To ensure that local minima are not a problem, the fitting procedure can be repeated by choosing additional starting configurations of the model and hill climbing from there, and choosing the best-fitting parameters of the multiple fittings. [4]

As an example of model fitting, consider again the stimuli shown in Table 1. If a classification experiment is conducted using these stimuli, the resulting data can be fit with the prototype model. For example, a hypothetical data[5] set is shown in Figure 5. The observed data are graphed as the dark symbols and solid lines and are expressed in terms of Category A performance. The prototype model was fit to this data set using the hill-climbing algorithm described above. The resulting best-fitting profile is shown in open circles on top of the observed data. The prototype model accommodates the data fairly well.

[3] We use $RMSD$ because it expresses fit in an intuitive way and it allows for comparisons between training sets and transfer sets that have different numbers of items. In general, when the models are fit by minimizing the $RMSD$ or maximizing the log likelihood, the resulting best-fitting parameters and predictions are nearly the same (J. D. Smith & Minda, 2000).

[4] Although many statistics and mathematical program like Matlab, R, and others provide model-fitting routines, we've implemented the model and the fitting procedure in a number of basic computing languages. The simulations and fitting in this paper were all programmed by the first author in REALbasic for Mac OS.

[5] These data are reflective of the data that have been shown by many actual subjects in some studies. We generated these specific probabilities for this chapter as a way of illustrating the fit of the prototype model to data.

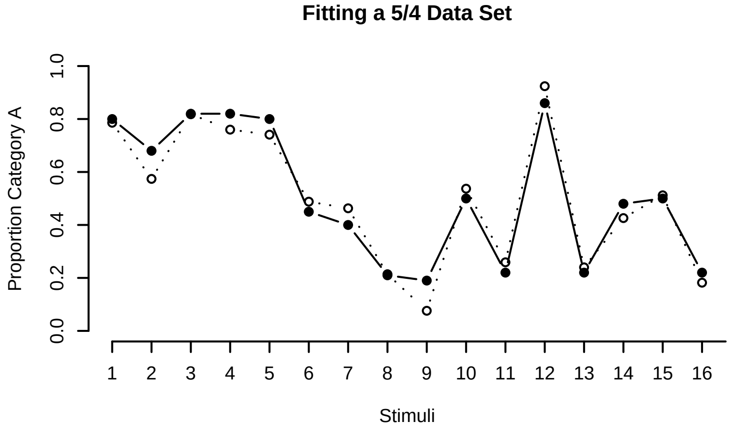

Figure 5. This figure shows the fit of the prototype model to a hypothetical set of classification data. The observed data are illustrated with filled circles and solid lines; the predictions of the prototype model are shown with open circles and dotted lines.

For example, it reproduces the prototype-enhancement effect for Stimuli 9 and 12 (though, in line with its representational assumption, it actually predicts a stronger enhancement than is shown). In this particular case, the $RMSD$ was 0.05.

A second example illustrates the linear separability constraint of the prototype model. This example uses the category set first used by Smith and colleagues (J. D. Smith et al., 1997; J. D. Smith & Minda, 1998) and shown in Table 2. Each exemplar is defined by 6 binary dimensions, and each category contains 7 exemplars. Notice that Stimuli 1 and 8 are the actual prototypes for their respective categories. Also notice that Stimuli 2-6 and Stimuli 9-13 are all high-typicality exemplars that share 5/6 of their features with their own prototype. However, Stimulus 7 and Stimulus 14 are exception items that only share 1/6 features with their prototype and 5/6 with the opposite category prototype (think of them as BATS in the RODENT category). Furthermore, there is no combination of attention weights that allow the exception to be correctly classified without making an error on another stimulus. For example, Stimuli 5 and 7 in Category A are exact featural opposites. Any attentional allocation that weighted Dimension 5 strongly enough to let Stimulus 7 be correctly assigned to Category A, would force Stimulus 5 to be incorrectly assigned to Category B. The presence of the exceptions along with their respective opposite in the same categories is what breaks the linear separability constraint for these stimuli. As with the previous example, an experiment using this category set would result in a set of classification probabilities. One set of classification probabilities is shown in Figure 6A. The solid dots are the Category A probabilities for a hypothetical subject who performed well on the prototypes and high-typicality exemplars, but misclassified the exception items (subjects in J. D. Smith and Minda (1998) did show this pattern). The many errors in the data that occur on the exception items suggest a linear separability constraint. As can been seen in the figure, the prototype model accommodates well this pattern ($RMSD$ = .06). This is

Table 2: Non Linearly Separable Category Set

Stimulus	d1	d2	d3	d4	d5	d6
Category A						
1	0	0	0	0	0	0
2	1	0	0	0	0	0
3	0	1	0	0	0	0
4	0	0	1	0	0	0
5	0	0	0	0	1	0
6	0	0	0	0	0	1
7	1	1	1	1	0	1
Category B						
8	1	1	1	1	1	1
9	0	1	1	1	1	1
10	1	0	1	1	1	1
11	1	1	0	1	1	1
12	1	1	1	0	1	1
13	1	1	1	1	1	0
14	0	0	0	1	0	0

because of the model's assumption that only prototypes serve as the comparison standard. However, as some research has shown, subjects can learn to overcome this linear separability constraint and can learn to classify correctly the exceptions (Medin & Schwanenflugel, 1981; J. D. Smith et al., 1997; J. D. Smith & Minda, 1998). Figure 6B shows the data from a hypothetical subject who performed much better on the exception items (and performed well on all the items). In this case, the model did not fit that data as well ($RMSD = .13$), and it made significant under/over prediction errors in the exception items and their complements (Stimuli 5 and 7 in Category A; Stimuli 12 and 14 in Category B). In this case, the prototype model was still operating under a linearly separable constraint, even while the data did not show this pattern. In other words, prototype models must predict the linear separability constraint as a consequence of their representational assumption.

To summarize, the basic prototype model can be used to create simulations and to make predictions. These predictions can inform experimenters and can be compared to the outcomes of real experiments. The model can also be configured to fit data collected from experiments. We have also described some fundamental predictions of the prototype model, including the prototype-enhancement effect, the strong typicality gradients predicted by prototype-based category representations, the linear separability constraint, and the effect of increasing the value of the scaling parameter. In the final section, we consider the relationship of the prototype model to several other categorization models and also the future of the prototype model.

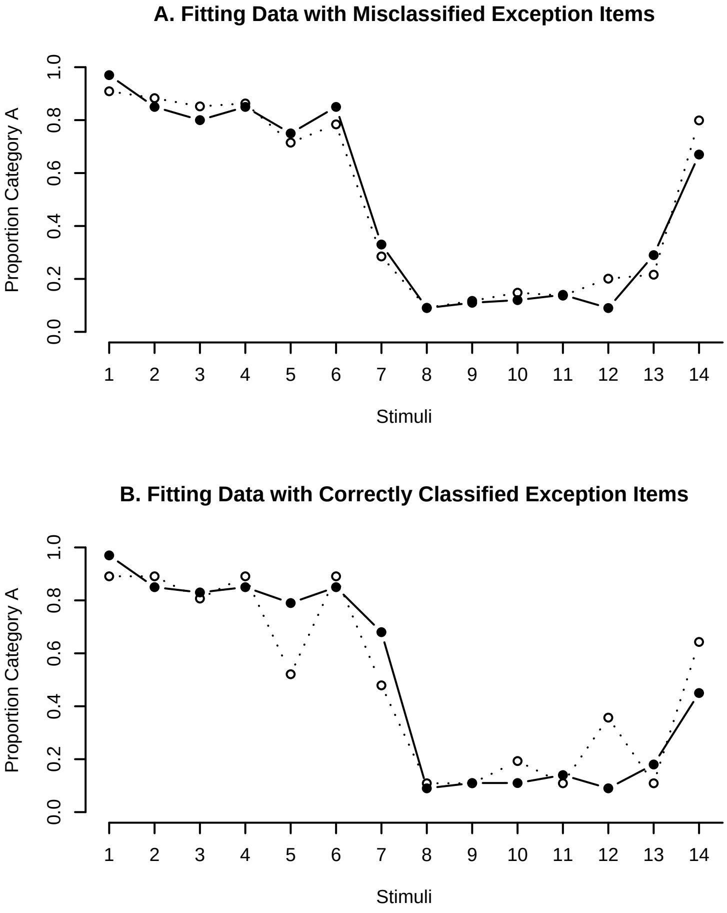

Figure 6. This figure shows the fit of the prototype model to two hypothetical sets of classification data. The observed data are illustrated with filled circles and solid lines; the predictions of the prototype model are shown with open circles and dotted lines. Panel A shows observed data that obey a linear separability constraint and Panel B shows observed data that do not obey a linear separability constraint.

Relationship to Other Models

The prototype model is closely related to two classes of models. First, it bears a computational and historical association with the *Generalized Context Model* (Nosofsky, 1986, 1987), which is also covered in this volume (Chapter 2). Second, the prototype model is related to several models that can assume prototype-based representations, such as SUSTAIN (Love, Medin, & Gureckis, 2004) and KRES (Rehder & Murphy, 2003), both of which have some degree of prototype-based representation, and both of which are covered in this volume (Chapters 10 and 12). We cover a number of these relationships, though not all, and we begin with GCM because of the strong connection between these models.

Relationship with Exemplar Models

The prototype model that we describe here is a computational partner of the GCM and there are numerous overlaps between these two models, as well as a fundamental difference.[6] First, the GCM and the prototype model rely on the same similarity calculations when deciding category membership (Equations 1 and 2). That is, both models assume that to-be-categorized items are compared to stored representations, that the dimensions can be weighted, and the similarity between items and category representations is an exponential function of distance in psychological space. Furthermore, both models rely on a scaling parameter c to adjust the similarity function. The key difference, of course, is that in the prototype model the to-be-categorized items are compared to stored prototypes whereas in the GCM they are compared to the set of stored exemplar traces. Another similarity between the models is that they make a classification decision in the same way. Once distance and similarity have been computed, the classification choice (Equation 3) is based on the similarity of the items to one category divided by the similarity to all categories. The result is that both of these models produce a decision that corresponds to the probability of category membership. Finally, both models can be fit to subjects' data by the same parameter estimation and minimization / maximization routines (minimizing the *RMSD* or maximizing the log likelihood). As a result, these two models are well suited for comparisons and even for using together to generate predictions about when and how subjects may rely on prototype abstraction and when and how they rely on exemplar learning (Minda & Smith, 2001; J. D. Smith & Minda, 1998).

Given the close correspondence, it should not be surprising that the two models make similar predictions about many categorization phenomena, and indeed both models often provide comparable fits to data sets (J. D. Smith et al., 1997; J. D. Smith & Minda, 1998, 2000). Because of this, researchers have often relied on carefully constructed category sets and experimental designs in order to distinguish these models. Although the literature has several examples of this, including dot patterns and non linearly separable categories, (Medin & Schwanenflugel, 1981; J. D. Smith & Minda, 2001, 1998), we will illustrate our point by discussing the category set shown in Table 1. As we mentioned earlier, the prototype model represents each category by storing the prototype (0 0 0 0) of the category. The GCM represents each category by storing the training exemplars (each of the 5 category A exemplars). Although the two representational schemes are very different (one involves

[6] Because ALOVE is a descendant of the GCM, many of the specifics we discuss here will also be true of ALCOVE and KOVE, which are also discussed in this volume.

abstraction; the other does not) the two models often make similar predictions because the prototype is the average of the exemplars that the GCM is storing. It is still possible to differentiate the two models, however. Recall that the prototype model will always predict that Stimulus 1 is a better (more typical) member of Category A than Stimulus 2, because Stimulus 1 shares more features with the prototype. The GCM, however, makes the opposite prediction. Considering exemplar similarities as the GCM does, Stimulus 1 is only marginally similar (sharing only 2 of 4 features) to three other members of Category A. In contrast, it is very similar (sharing 3 of 4 features) with two members of Category B. Stimulus 2, on the other hand, has substantial featural overlap with Category A members but not with Category B members. As a result, the GCM predicts stronger Category A assignment for Stimulus 2 than for Stimulus 1. Although subjects in some of the early experiments showed the pattern predicted by the exemplar model (Medin & Schaffer, 1978), this has not always been the case (Minda & Smith, 2002; J. D. Smith & Minda, 2000). In other words, the category set really does distinguish between prototype and exemplar based categorization, but there is no clear evidence that subjects will learn these categories as prototypes or as exemplars. More recent work with dot patterns, which can also be used to reliably distinguish prototype from exemplar processing, has found more consistent evidence for prototype-based categorization (J. D. Smith & Minda, 2001; J. D. Smith, 2005; J. D. Smith et al., 2008).

Models with Prototype Assumptions

The prototype model is related to a number of other models discussed in this volume by virtue of the assumption of prototype abstraction. That is, regardless of how classifications are made in other models, and in addition to other assumptions about rules or exemplars, many of the models in this volume contain a mechanism for reproducing prototype-like performance. Take as an example the KRES model (Rehder & Murphy, 2003). This model was originally designed as a prototype-based model and was implemented as a connectionist model that learns direct connections between stimulus features and category labels, amounting to a “feature list” kind of prototype. As a result, KRES makes many of the same predictions as our standard prototype model, including the assumption about linearly separability (Rehder & Murphy, 2003). Of course a key difference between KRES and a basic prototype model is the addition of knowledge to the network (that is the whole point of KRES). Furthermore, the version of KRES discussed in this volume allows for both prototype and exemplar nodes, in order to capture the flexibility and different kinds of knowledge that subjects use.

Another class of models which are discussed in this volume and that make a prototype assumption are the adaptive clustering models, like SUSTAIN (Love & Gureckis, 2007; Love et al., 2004). SUSTAIN is not a prototype model *per se* but rather, it assumes that categories can be learned as clusters of similar stimuli. A single cluster can represent one or many exemplars. As such, SUSTAIN has the ability to represent categories with a single prototype, several sub-prototypes, or with many single exemplars, and can be thought of as a hybrid model. Furthermore, SUSTAIN has a mechanism for supervised learning (i.e., explicit, feedback-driven classification) and unsupervised learning (which is important in dot-pattern learning). SUSTAIN has been successfully applied to a broad range of data, suggesting that prototypes may provide the best solution to a classification problem, even when an exemplar-based solution is possible.

Finally, to some degree, the simplicity model (Pothos & Chater, 2002), which is described in this volume(Chapter 9), is likely to find a category description that results in prototype-like classifications. The simplicity model is designed to maximize category coherence and simplicity, though the model itself is not a prototype model. However, like the adaptive clustering models, this model may arrive at a classification that is intuitive, and possibly the result of a prototype representation. Furthermore, like the prototype model, the simplicity model is sensitive to the linear separability constraint.

Future Directions

The prototype model has a long history in cognitive psychology and the study of categorization. It is not an understatement to say that the currently rich field of categorization models (as can be seen in this volume) owes some debt to the early success of the prototype model and the various models that were formulated as extensions and counters to it. As an example, consider the debate between the prototype and exemplar models (Blair & Homa, 2001; Medin & Schwanenflugel, 1981; Nosofsky & Zaki, 2002; Nosofsky, 1992; J. D. Smith et al., 1997; J. D. Smith & Minda, 1998). Clearly the GCM benefitted from this program of research at least as much as the prototype model. In fact, one development that came out of that debate was the mixture model (Minda & Smith, 2001; J. D. Smith & Minda, 1998, 2000) that combined a prototype abstraction system with exemplar memorization. This model assumes that although the prototype describes much of subjects' performance, subjects may also try to learn specific exemplars. Other models make the same mixed representational assumption (KRES and ATRIUM, for example).

Clearly the prototype model provides an account of category learning that emphasizes the abstraction and storage of prototypes. Other models do this as well, but the prototype model we have described here is explicit in its assumptions about how these prototypes give rise to performance. As a result, this model makes strong claims about humans (and nonhuman) categorization behavior. Prototypes must predict a linear separability constraint. This is a strong claim, and clearly not all categories are linearly separable. But there is considerable research, from natural categories, artificial concept learning, comparative work, and computational modeling suggesting that humans and nonhumans may share this constraint. This constraint can be overcome: people can learn exceptions to a well-structured category. But we argue that this linearly separable constraint may be one of the default assumptions about categories that humans and and nonhumans make when first learning about a new set of categories.

Therefore, we see a constructive future for the prototype model. The prototype model operates via a clear and transparent formalism. Each equation that defines the basic model is intuitive and has strong psychological underpinnings. That is, the assumptions about similarity and psychological space hold up under a variety of circumstances and scenarios. For this reason the prototype model should take a place within a larger understanding of categorization that grants organisms redundant systems and different category representations for different tasks and purposes. This view is not opposed to other categorization models and theories, but rather is inclusive. The minds of humans and nonhumans are not unitary in process or representation but diverse and varied in the approaches they take to the fundamental task of forming categories.

References

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149-178.

Blair, M., & Homa, D. (2001). Expanding the search for a linear separability constraint on category learning. *Memory & Cognition*, *29*, 1153-1164.

Blair, M., & Homa, D. (2004). As easy to memorize as they are to classify: The 5–4 categories and the category advantage. *Memory & Cognition*, *31*, 1293-1301.

Chin-Parker, S., & Ross, B. H. (2004). Diagnosticity and prototypicality in category learning: A comparison of inference learning and classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 216-226.

Homa, D., Cross, J., Cornell, D., & Shwartz, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology*, *101*, 116-122.

Homa, D., & Cultice, J. C. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 83-94.

Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, *262*, 1747-1749.

Love, B. C., & Gureckis, T. M. (2007). Models in search of a brain. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 90-108.

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

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

Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 355-368.

Minda, J. P., & Ross, B. H. (2004). Learning categories by making predictions: An investigation of indirect category learning. *Memory & Cognition*, *32*, 1355-1368.

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

Minda, J. P., & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 275-292.

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

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

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning Memory, and Cognition*, *13*, 87-108.

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

Nosofsky, R. M. (1992). Exemplars, prototypes, and similarity rules. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning theory to connectionist theory: Essays in honor of William K. Estes* (Vol. 1, p. 149-167). Hillsdale, NJ: Lawrence Earlbaum.

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 924-940.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology 77*, 353-363.

Pothos, E., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science*, *26*, 303–343.

Reber, P., Stark, C., & Squire, L. (1998a). Contrasting cortical activity associated with category memory and recognition memory. *Learning & Memory*, *5*, 420-428.

Reber, P., Stark, C., & Squire, L. (1998b). Cortical areas supporting category learning identified using functional MRI. *Proceedings of the National Academy of Sciences*, *95*, 747-750.

Rehder, B., & Murphy, G. L. (2003). A knowledge-resonance (KRES) model of category learning. *Psychonomic Bulletin & Review*, *10*, 759-784.

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

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382-439.

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

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Smith, J. D. (2005). Wanted: A new psychology of exemplars. *Canadian Journal of Experimental Psychology*, *2003 Festschrift for Lee R. Brooks. Vol 59*, 47-53.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning Memory, and Cognition*, *24*, 1411-1436.

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning Memory, and Cognition*, *26*, 3-27.

Smith, J. D., & Minda, J. P. (2001). Journey to the center of the category: The dissociation in amnesia between categorization and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *4*, 501-516.

Smith, J. D., Murray, J., Morgan J., & Minda, J. P. (1997). Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 659-680.

Smith, J. D., Redford, J. S., & Haas, S. M. (2008). Prototype abstraction by monkeys (*Macaca mulatta*). *Journal of Experimental Psychology: General*, *137*, 390-401.

Wittgenstein, L. (1958/2001). *Philosophical investigations.* New York, NY: Blackwell.

Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory & Language*, *39*, 124-148.

Zeithamova, D., Maddox, W. T., & Schnyer, D. M. (2008). Dissociable prototype learning systems: Evidence from brain imaging and behavior. *Journal of Neuroscience*, *28*, 13194-13201.