Tracking Variability in Learning: Contrasting Statistical and Similarity-Based Accounts

Yasuaki Sakamoto (yasu@psy.utexas.edu)
Bradley C. Love (love@psy.utexas.edu)
Matt Jones (mattj@psy.utexas.edu)
Department of Psychology, The University of Texas at Austin
Austin, TX 78712 USA

Abstract

Learning to categorize objects involves learning which sources of variability are meaningful and which should be ignored or generalized. In this light, theories and models of category learning can be viewed as accounts of how people capture and represent meaningful variation. Similarity-based models, such as prototype and exemplar models, cannot correctly predict that humans classify a stimulus halfway between the nearest members of a low-variability and high-variability category into the high-variability category. Distributional accounts, descending from the unequal variance signal detection model, can accommodate the result. We present a simple extension to similarity-based models that allows them to display the sensitivity to category variability that humans display. We conclude by discussing what constitutes similarity-based representations and processes and noting the points of convergence between similarity-based and distributional approaches.

Humans operate in environments marked by variability. For instance, categorizing a novel stimulus (e.g., determining whether a person is friend or foe) involves generalizing from past experiences that differ from one another and the current situation.

In this light, models of category learning are accounts of which sources of variability are meaningful and which should be ignored (i.e., generalized). For instance, prototype models abstract (i.e., average) across previous category members to form a central tendency or prototype (Posner & Keele, 1968). In prototype models, the meaningful way in which items vary is in their similarity (i.e., distance) to category prototypes.

Exemplar models deem other sources of variability meaningful. Rather than storing a summary of previously experienced items as prototype models do, exemplar models store every experienced example in memory (Medin & Schaffer, 1978). In exemplar models, the meaningful way in which items vary is in the sum of their pairwise similarities (i.e., distances) to the exemplars representing each category.

Although prototype and exemplar models offer quite different accounts of how categories are represented, they both use similarity-based processing and can make overlapping predictions. Figure 1 illustrates a case in which these models' predictions converge. Subjects learned to classify lines varying in length into one of two categories. Training items are illustrated as dark triangles. The six items (L1–L6) forming one category are relatively less variable than the six items (H1–H6) forming the contrasting category. Following training, subjects classified a variety of items, including some items that were not experienced during training, such as item N6. These novel items are tests of how subjects generalize. Item N6 is of particular interest as it is halfway between the nearest trained members (L6 and H1) of the low-variability and high-variability categories.

Both prototype and exemplar models strongly predict that subjects will classify border item N6 into the low-variance category because the same similarity metric is used for the low-variance and the high-variance categories, and the prototype for the low-variance category is closer to N6 as are the exemplars forming the low-variance category. In contrast, distributional approaches, such as general recognition theory (Ashby & Townsend, 1986) and the category density model (Fried & Holyoak, 1984), predict that item N6 should belong to the high-variance category. These distributional approaches are descendants of the unequal variance signal detection model (Green & Swets, 1966) and represent variability information separately for each category. Distributional approaches seem normative in that they use information about how members of a category vary from one another and this information can potentially improve accuracy. In Figure 1, the density functions of unequal variance depict the category representations of a distributional model. The density function for the high-variability category is above the curve for the low-variability category at N6's location. Therefore, the distributional model predicts N6 belongs to the high-variability category.

To foreshadow the results, subjects are sensitive to the variability across category members as predicted by distributional models and classify the border item N6 into the high-variance category. This result seems to undermine existing similarity-based approaches and favor distributional approaches. However, given the remarkable success of similarity-based models of categorization, it would be imprudent to discard this class of models out of hand.

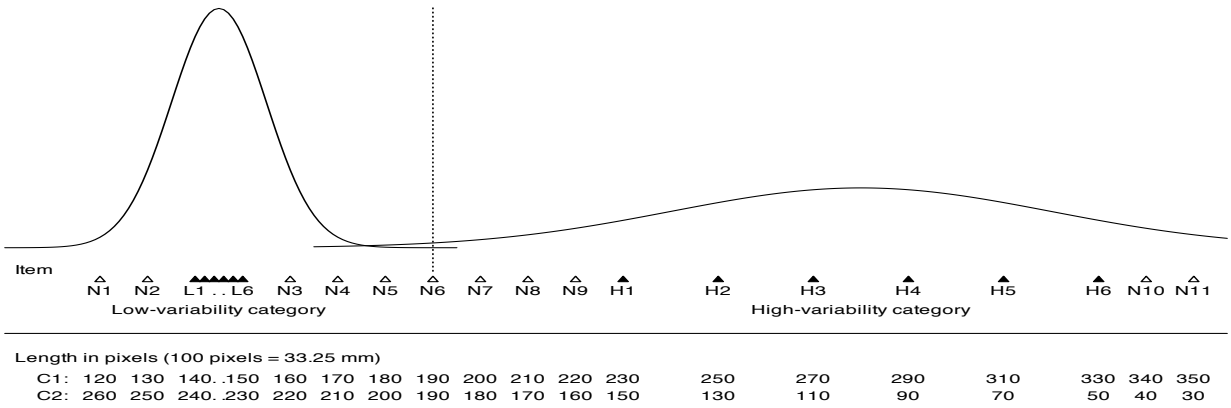

Figure 1: Two categories differing in variability are shown. Dark triangles (L1–L6 and H1–H6) represent training items and light triangles (N1–N11) represent novel items that did not appear during learning. The item lengths are spaced to scale. Item N6 is exactly midway between the nearest studied members (L6 and H1) of the low- and high-variability categories. Items in the low-variability category differ from the nearest studied member by 2 pixels, whereas items in the high-variability category differ from the nearest studied member by 20 pixels. To eliminate possible influences of absolute line length on performance, whether the high-variability category had longer (C1) or shorter lines (C2) than the low-variability category was counterbalanced between subjects. Border item N6 has the same length in both conditions. The two density curves illustrate possible category representations for a distributional model and do not indicate information about the frequency of presentation during the experiment.

The core intuitions underlying similarity-based models encompass constructs like the representativeness heuristic (Tversky & Kahneman, 1974). Moreover, findings like the inverse base rate effect (Medin & Edelson, 1988) are problematic for distributional approaches. To reconcile this impasse, we present a similarity-based model that develops category representations that are sensitive to distributional information that unequal variance models can exploit.

In the General Discussion, we will present related work in light of our findings. We should briefly note that although numerous studies have explored the effects of variability on categorization, the true nature and extent of these effects is far from clear. For example, Fried and Holyoak (1984) found that the border item is more likely to be assigned to the high-variability category, but in their design this item is also more similar to specific training examples from the high-variability category. Likewise, other work exploring the influence of category variability has not been diagnostic in evaluating similarity-based and distributional accounts (e.g., Homa & Vosburgh, 1976; Posner & Keele, 1968). Despite these open issues, findings that are consistent with both similarity-based and distributional accounts are often marshaled to support distributional accounts. For example, Heit and Feeney (2005) stated that "there is well-established evidence from categorization research that more variable observations promote broader or stronger generalizations (e.g., Fried & Holyoak, 1984; Homa & Vosburgh, 1976; Posner & Keele, 1968)" (p. 340). Such citations are common and, while not factually incorrect, do incorrectly imply that the literature has established that people are sensitive to distributional information above and beyond what similarity-based models track. Here, we provide a strong test that distinguishes between existing similarity-based and distributional accounts.

Experiment

Fifty University of Texas undergraduates learned to correctly assign 12 line stimuli (represented by dark triangles labeled L1–L6 and H1–H6 in Figure 1) into category A or B through trial by trial classification learning with corrective feedback. The members of one category (L1–L6) varied relatively little in their lengths, whereas the members of the other category (H1–H6) were highly variable (see Figure 1).

On each learning trial, one line was presented horizontally at the center of a display and the text "Category A or B?" appeared at the top left corner of the display. After responding A or B, subjects received visual (e.g., "Right! The correct answer is A.", "Wrong! The correct answer is B.") and auditory corrective feedback (i.e., a low-pitch tone for errors and a high-pitch tone for correct responses). The visual feedback (presented at the bottom left corner of the display) and the stimulus were displayed for 2000 ms. Subjects completed 10 blocks of learning trials. A block was the presentation of each learning item in a random order.

Following learning, subjects answered three addition problems to prevent rehearsal of information from the learning phase. Finally, subjects completed two blocks of transfer classification. In each transfer block, subjects classified the 12 studied and 11 novel items (represented by light triangles labeled N1–N11 in Figure 1) in a random order as they did in the learning phase except that no corrective feedback was provided. Our main interest was subjects' performance on the border transfer item (N6) that was midway between the nearest studied members (L6 and H1) of the two categories.

Results

As shown in Figure 2, border item N6 was more likely to be classified into the high-variability than into the low-variability category. Averaged across the two transfer blocks, subjects assigned item N6 to the high-variability category with greater than chance probability (.69 vs. .5), $t(49) = 3.86$, $p < .001$. In the first transfer block, more subjects (33 of 50) classified item N6 to the high-variability category than was expected by chance, exact binomial $p = .033$ (two-tailed). The same pattern (36 of 50) was found for item N6 in the second transfer block, exact binomial $p = .0026$ (two-tailed).

Extending Similarity-Based Models

As discussed in the Introduction, existing similarity-based models, such as prototype and exemplar models, cannot accommodate the current finding demonstrating that humans are sensitive to the variability across a set of category members. In this section, a simple extension to similarity-based models that use error-driven learning (e.g., Kruschke, 1992; Love, Medin, & Gureckis, 2004) is proposed. The simulations of a prototype (e.g., Smith & Minda, 1998) and an exemplar (e.g., Nosofsky, 1986) version of the model serve as an existence proof that the similarity-based approaches can be readily extended to account for findings supporting distributional approaches. Thus, the main goal of the modeling is to evaluate the distribution learning mechanisms while keeping other variables constant.

Prototype model The prototype version of the model represents each category with a single cluster (i.e., the prototype). Activation of cluster i, a_i, is a Gaussian function of the presented stimulus, x:

$$a_i = \frac{1}{\sqrt{2\pi}s_i} e^{-\frac{(x-\mu_i)^2}{2s_i^2}} \quad (1)$$

where μ_i and s_i are the cluster's mean and standard deviation, respectively. The response probability for each category is proportional to the activation of the corresponding cluster (i.e., the probability matching response rule). For simplicity it is assumed that μ_i corresponds to the true category mean and is not subject to learning.

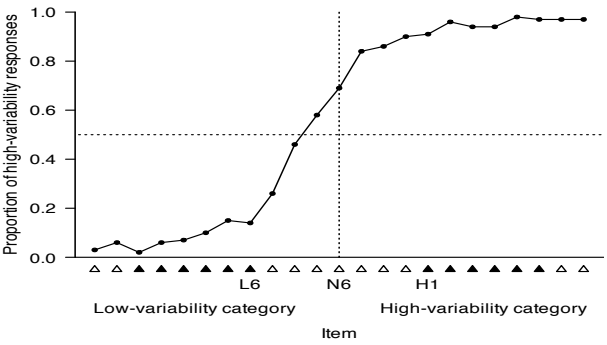

Figure 2: Mean proportion of trials in which each stimulus item was assigned to the high-variability category during the transfer phase is shown. Dark triangles represent studied items and light triangles represent novel items. Item N6 is midway between the nearest studied members (L6 and H1) of the low- and high-variability categories. Items are not spaced to scale (see Figure 1 for the physical scale).

Standard deviations are learned by gradient descent:

$$\Delta s_i = -\epsilon \frac{\partial}{\partial s_i} \{\frac{1}{2}(t_i - a_i)^2\} \quad (2)$$

where ϵ is a learning rate and t_i is the feedback to cluster i, equal to α if the stimulus is in category i and 0 otherwise. Equation 2 yields the following learning rule:

$$\Delta s_i = \epsilon(t_i - a_i) \frac{(x-\mu_i)^2 - s_i^2}{s_i^4\sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2s_i^2}}. \quad (3)$$

The model was trained and tested in a trial-by-trial fashion like the human subjects. Figure 3 illustrates the dynamics of the model simulated on the present experiment. This figure is based on an average over 100 separate runs, using the parameter values $\epsilon = 28000$, $\alpha = .04$, and $s_0 = 14$.

Prior to training, both clusters have the same standard deviation of 14 and border item N6 is more similar to the cluster representing the low-variability category because its prototype is closer to item N6. The model thus predicts that item N6 should be classified into the low-variability category before training. Following training, the cluster encoding the low-variability category is tightened to minimize unwanted activations by items from the high-variability category, leading to a learned standard deviation of 9.5 (averaged across runs). The cluster encoding the high-variability category is similarly widened, leading to an average standard deviation of 20.4. These effects are illustrated in the bottom panel of Figure 3 (as predicted in Figure 1). Consequently, item N6 more strongly activates the high-variability category's cluster after learning. The ratio of cluster activations for stimulus N6 leads to a

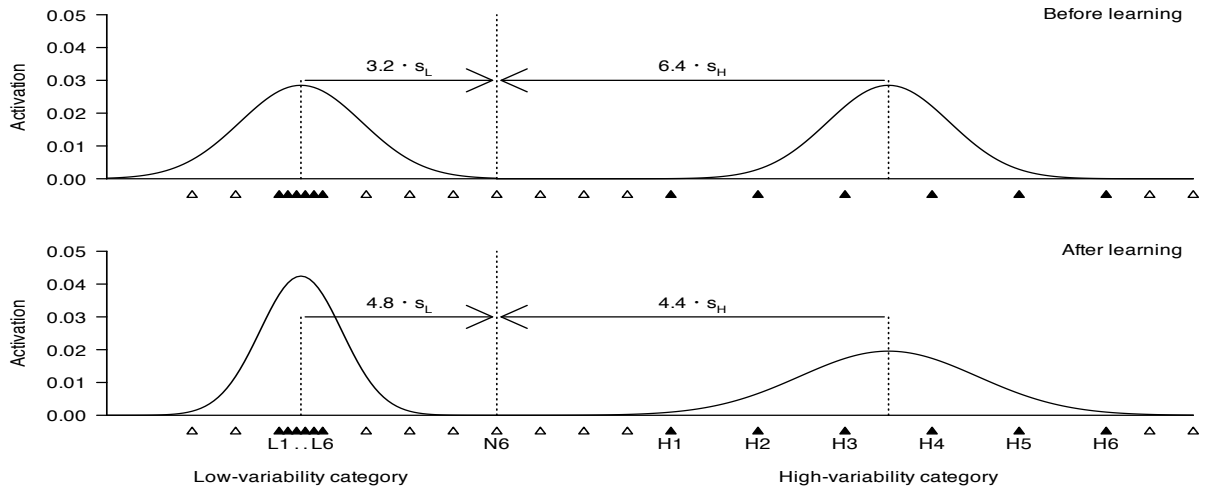

Figure 3: The activations of the clusters encoding the low-variability (cluster L) and high-variability categories (cluster H) in the prototype model are shown for each stimulus item. Before learning, each cluster's dispersion is equal and item N6 is fewer standard deviations from cluster L's than cluster H's mean (indicated by arrows), leading to greater activation and higher response probability for the low-variability category. The opposite pattern is observed after learning due to the tightening of cluster L and the widening of cluster H, which makes the clusters relatively more responsive to their members, and item N6 is now more likely to be assigned to the high-variability category. This process occurs for every exemplar in the exemplar model.

.68 probability of selecting the high-variability category, in close agreement with the empirical data.

Exemplar model The exemplar version of the model represents categories by a separate cluster centered at each stimulus and responds based on summing activations from each category. The model is otherwise identical to the prototype version.

The exemplar model learns to tighten the clusters in the low-variability category and broaden those in the high-variability category. The model was simulated for 100 runs using the parameter values $\epsilon = 150000$, $\alpha = .025$, and $s_0 = 20$. After learning, the mean cluster dispersions were 15.4 for the low-variability category and 26.1 for the high-variability category, leading to a .69 probability of classifying the border stimulus into the high-variability category. Consistent with the current modeling, Nosofsky and Johansen (2000) extended Nosofsky's (1986) Generalized Context Model to include exemplars with varying dispersions, though they did not specify a learning rule for updating the dispersions.

A Further Test of the Modeling Approach Both the prototype and exemplar versions of the model predict that subjects should initially assign the border item to the low-variability category but reverse their preference after training and favor the high-variability category. Our experiment demonstrated that subjects assign the border item to the high-variability category after training, but did not test whether subjects favor the low-variability category prior to training. Twenty-five University of Texas undergraduates completed a one trial experiment that evaluated their pre-training preference to assign the border item N6 to the low- or high-variability category. Subjects were shown the two category prototypes and chose to which category the item N6 belonged. In this single triad task, subjects preferred (22 of 25) to classify item N6 into the low-variability category, exact binomial $p = .00016$ (two-tailed), consistent with the model's predictions.

General Discussion

The current experiment examined the effect of category variability on classification behavior. Subjects learned about two artificial categories with one category more variable than the other. When transferred to novel stimuli, subjects classified an item halfway between the nearest members of the two categories into the high-variability category, suggesting that humans develop distributional knowledge for categories, which they use when making category judgments. Existing similarity-based models, such as exemplar and prototype models, incorrectly predict the border item should be assigned to the low-variability category after training. Distributional accounts can accommodate the current result.

The present work provides the first experimental demonstration of a preference for one category over another based solely on differences in variability. Stewart and Chater (2002) observed a preference for the higher-variability category only when instruc-

tional manipulations and simultaneous presentation of all stimuli alerted subjects to the category structures. Under more standard learning conditions, Stewart and Chater found a preference to assign the border item to the low-variability category, consistent with similarity-based accounts, and thus their design was unable to discern between similarity-based and distributional accounts. Rips (1989) has shown preferences for the higher-variability category in line with distributional accounts, but only by relying on pre-existing knowledge and categories, as opposed to utilizing well-controlled experimental manipulations and training procedures. Other studies have shown a decrease in similarity-based responding with distributional manipulations, though not to the point where the high-variability category was preferred (Cohen, Nosofsky, & Zaki, 2001; Hahn, Bailey, & Elvin, 2005). Cohen et al. did observe a preference for the higher-variability category utilizing two-dimensional stimuli, but these stimuli invite attentional explanations of performance.

Certain characteristics of the current experiment's design are likely responsible for the clear pattern of empirical findings. One positive aspect of the design is that the spacing between the low- and high-variability categories is large relative to that in the previous studies (e.g., Cohen et al., 2001), which could lead to a stronger influence of variability by decreasing the border item's similarity to either category. The difference in internal variance between categories is also larger and thus more apparent in the current design than in the previous studies (e.g., Stewart & Chater, 2002). At the same time, both categories have discriminable internal variance (i.e., subjects can appreciate that the categories contain multiple members that vary from one another like real-world categories), in contrast to Cohen et al.'s Experiment 1 in which the low-variability category consists of a single item. Further, the category variances in the present experiment are learned through direct experience with the items. In transfer, no corrective feedback is provided and the space of possible stimuli is uniformly sampled, which stresses generalizing previous knowledge and reduces the chances of significant unsupervised learning during transfer.

We presented a simple extension to similarity-based models that allows them to address the current findings. Prototype and exemplar models were conceived as extreme cases of clustering solutions in which prototype models devote one cluster for each category whereas exemplar models devote one cluster for each item. In both limiting cases, the clusters adaptively adjusted their tolerance of variability through an online training procedure that maximized correct responding. Both the prototype and exemplar versions of the model correctly predicted that people's preference to assign the border item should shift from the low- to the high-variability category with training. Because clustering approaches span the gamut from prototype to exemplar models, the simulation results have broad applicability to similarity-based approaches. We should note that although an exemplar model was constructed to demonstrate the generality of our extension to similarity-based models, such an extension seems antithetical to the fundamental tenets of the exemplar approach as variability is not a characteristic of an individual example.

The success of the current simulations raises the issue of what counts as true similarity-based representations and processes. The notion of similarity itself has evolved over the last few decades. Similarity is now more of a construct to be studied in its own right than an explanatory element. For instance, to explain the perceived similarity of two objects or scenes, theories in the analogy literature posit representations containing relations and complex comparison operations that put these representations into alignment (Gentner & Markman, 1997).

Closer to the present work, studies of variability and categorization have also found that stimulus representations and the comparison processes stressed at test play major roles in shaping performance. For example, Smith and Sloman (1994) found that Rips' (1989) results favoring high-variance responses to border items do not replicate when people are given perceptually rich stimuli or do not engage in verbal descriptions.

In conclusion, similarity-based accounts of category learning have the virtue of orienting research toward fundamental representational and processing issues. It should not be surprising if, as in the current work, the set of representations and processes on which the field converges bears characteristics of normative accounts, such as the distributional accounts considered here. Indeed, a Bayesian account with a properly specified prior on the category variability should be able to accommodate the observed shift in people's preference to assign the border stimulus to the low-variability category before training and to the high-variability category after training. Augmenting similarity-based models has the potential to ease theoretical conflicts between researchers advocating similarity-based representations (e.g., McKinley & Nosofsky, 1996) and those advocating distributional representations (e.g., Maddox & Ashby, 1998).

In addition, the present line of research can shed light on how people categorize more naturalistic stimuli and how the expertise to succeed at such tasks develops over training. Though not tested under supervised learning, one explanation for asymmetries in infant categorization (e.g., generalizing from dogs to cats but not from cats to dogs) is that infants are sensitive to category variability (French, Mareschal, Mermillod, & Quinn, 2004). Many critical tasks, such as training medical students to detect cancer, depend on the learner's experiencing a vari-

ety of category members. Theories and models that help us understand how people come to appreciate and represent meaningful variation should provide useful guidance in devising training regimens.

Acknowledgments

This work was supported by AFOSR FA9550-04-1-0226 and NSF CAREER 0349101 to B. C. Love and NRSA F32-MH068965 from NIMH to M. Jones.

References

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154–179.

Cohen, A. L., Nosofsky, R. M., & Zaki, S. R. (2001). Category variability, exemplar similarity, and perceptual classification. *Memory & Cognition*, *29*(8), 1165–1175.

French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The role of bottom-up processing in perceptual categorization by 3- to 4-month-old infants: Simulations and data. *Journal of Experimental Psychology: General*, *133*, 382–397.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*, 234–257.

Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, *52*, 45–56.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Hahn, U., Bailey, T. M., & Elvin, L. B. C. (2005). Effects of category diversity on learning, memory, and generalization. *Memory & Cognition*, *33*(2), 289–302.

Heit, E., & Feeney, A. (2005). Relations between premise similarity and inductive strength. *Psychonomic Bulletin & Review*, *12*(2), 340–344.

Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(3), 322–330.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*, 22–44.

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

Maddox, W. T., & Ashby, F. G. (1998). Selective attention and the formation of linear decision boundaries: Comment on McKinley and Nosofsky (1996). *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 301–321.

McKinley, S. C., & Nosofsky, R. M. (1996). Selective attention and the formation of linear decision boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 294–317.

Medin, D. L., & Edelson, S. M. (1988). Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General*, *117*, 68–85.

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

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

Nosofsky, R. M., & Johansen, M. K. (2000). Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review*, *7*(3), 375–402.

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

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 21–59). New York: Cambridge University Press.

Smith, E. E., & Sloman, S. A. (1994). Similarity-versus rule-based categorization. *Memory & Cognition*, *22*, 377–386.

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*, 1411–1430.

Stewart, N., & Chater, N. (2002). The effect of category variability in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(5), 893–907.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1130.