

# The Role of Bottom-Up Processing in Perceptual Categorization by 3- to 4-Month-Old Infants: Simulations and Data

Robert M. French  
University of Liège

Denis Mareschal  
Birkbeck University of London

Martial Mermillod  
University of Liège

Paul C. Quinn  
University of Delaware

Disentangling bottom-up and top-down processing in adult category learning is notoriously difficult. Studying category learning in infancy provides a simple way of exploring category learning while minimizing the contribution of top-down information. Three- to 4-month-old infants presented with cat or dog images will form a perceptual category representation for *cat* that excludes dogs and for *dog* that includes cats. The authors argue that an inclusion relationship in the distribution of features in the images explains the asymmetry. Using computational modeling and behavioral testing, the authors show that the asymmetry can be reversed or removed by using stimulus images that reverse or remove the inclusion relationship. The findings suggest that categorization of nonhuman animal images by young infants is essentially a bottom-up process.

Few in cognitive science would dispute the argument that both bottom-up (i.e., perceptually driven) and top-down (i.e., conceptually driven) processes are involved in adult categorization. Numerous studies have discussed the relationship between these two mechanisms of categorization (e.g., French, 1995; Murphy & Kaplan, 2000; Schyns, Goldstone, & Thibaut, 1998). However, in adults, perceptual and conceptual processes are deeply intertwined, making them difficult to isolate and study independently (Goldstone & Barsalou, 1998).

This problem is particularly acute when studying category learning as opposed to category retrieval. When studying category learning, it is necessary to present participants with unfamiliar categories and to observe their responses during the period when classification behaviors rise from chance to some stable level (Ashby & Ell, 2001). The importance of exemplar distributions and of the number of familiarization items on adult category learning has been widely studied (e.g., Ashby, Alfonso-Reese,

Turken, & Waldron, 1998; Ashby & Waldron, 1999; Fried & Holyoak, 1984; Maddox & Ashby, 1993; Nosofsky, 1986; Posner & Keele, 1970; Reed, 1972). However, these investigations have used simple artificial stimuli such as dot patterns, checkerboards, simple monochromatic shapes, or lines of varying lengths and orientations. Indeed, in experiments with adults, the prevailing method of ensuring unfamiliarity is to create new arbitrary categories of objects (artificial categories). Of course, this begs the question of whether the patterns of behavior observed have any ecological validity. Although investigations of learning of these simple categories most certainly reflect some aspects of how new categories are learned in the real world and have been successful in launching the debate over whether concepts are represented by summary-level prototypes or individual exemplars (Nosofsky, 1991), they may fail to capture the richness and multidimensionality of the process of natural category formation.

Studying category learning in infancy provides one way of escaping this impasse. If one makes the reasonable assumption that young infants have little or no prior knowledge of many natural categories in their environment, it then becomes possible to study how these natural categories are acquired. Complex realistic stimuli that form the basis of semantically relevant adult categories can be used to train young infant participants without much risk of the category learning process being contaminated by category retrieval.

Furthermore, studying category learning in infancy can help identify the mechanisms by which category representations emerge as a result of exposure to a structured environment. Young infants are highly sensitive to the statistical regularities of their surroundings (e.g., Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996). We have argued elsewhere that early infant perceptual categorization reflects the internalization of statistical co-occurrences in the environment by an unsupervised associative memory system (Mareschal & French,

---

Robert M. French and Martial Mermillod, Psychology Department, University of Liège, Liège, Belgium; Denis Mareschal, School of Psychology, Birkbeck University of London, London, England; Paul C. Quinn, Department of Psychology, University of Delaware.

Martial Mermillod is now at the Laboratory of Psychology and Neuro-Cognition, University of Grenoble, Grenoble, France.

The order of authorship is alphabetical.

This work has been supported in part by European Commission Grant HPRN-CT-1999-00065 to Robert M. French, National Science Foundation Grant BCS-0096300 and National Institutes of Health Grant HD-42451 to Paul C. Quinn, and Economic and Social Research Council Grant R000239112 to Denis Mareschal.

Correspondence concerning this article should be addressed to Robert M. French, who is now at L.E.A.D., CNRS UMR 5022, Université de Bourgogne, BP 26513 21065, Dijon Cedex, France. E-mail: rfrench@u-bourgogne.fr

1997, 2000; Mareschal, French, & Quinn, 2000; Mareschal, Quinn, & French, 2002). Thus, if one takes the simplest position that the underlying mechanisms of this sensitivity to regular statistical properties in the environment remain largely unchanged from infancy to adulthood, then examining the structural basis of perceptual categorization in 3- to 4-month-old infants also provides a unique window on the bottom-up categorization processes that underlie category learning throughout an individual's lifetime.

In the experiments and computer simulations reported in this article, the stimuli were photographic images of dogs and cats rather than random dot patterns (Posner & Keele, 1970), simple geometric patterns (Bomba & Siqueland, 1983), or stick-figure animals (Younger, 1985). This choice was motivated by a number of factors. Dot-pattern or stick-figure images do not, in general, have the correlated attribute structure that is the hallmark of natural categories (e.g., dogs have fur and four legs and they bark, whereas birds have feathers and two legs and they chirp). Further, dot patterns consist of black dots on white backgrounds and thus lack the multidimensional complexity found on the surfaces of the realistic cat and dog exemplars (e.g., variation in texture, contrast, and color). Also, infants of approximately this age (i.e., 5-month-olds) do transfer successfully from some three-dimensional objects to their depictions in two-dimensional photographs (DeLoache, Strauss, & Maynard, 1979). Finally, infant category representations for cats and dogs, unlike simpler dot-pattern or geometric-figure stimuli, will later develop into the mature concepts of cats and dogs of older children and adults—categories that will include a wealth of conceptual information.

Thus, by studying infants' formation of realistic categories, one can begin to address the critical problem of how the subsequent knowledge-based representations of children and adults develop from the initial perceptual substrate of category representations formed in early infancy. In what follows, we show not only that infants can categorize visually presented photographic images of cats and dogs solely on the basis of perceived differences in the statistical distributions of the perceptual attributes of the different categories but, more important, that by manipulating these distributions without modifying the category identity (i.e., cat vs. dog), we can radically affect the way the infants categorize the images.

We turn now to a brief overview of the infant categorization literature. Early studies in which researchers used novelty preference procedures to investigate the emergence of categorization in human infants have provided evidence that infants under 1 year of age can form perceptual category representations for visual patterns such as schematic faces and geometric forms (Bomba & Siqueland, 1983; Quinn, 1987; Strauss, 1979; Younger & Gotlieb, 1988). For example, when presented with dot-pattern exemplars generated from diamond, square, or triangle prototypes, 3- to 4-month-olds generalized looking times to novel instances from the familiar form category and displayed visual preferences for novel instances from novel form categories. Moreover, research inquiring into the nature of early category representation has shown that under certain conditions known to facilitate prototype abstraction in adults (i.e., increased numbers of exemplars to be classified, delay between initial exposure and the recognition test), infants also displayed a prototype effect; that is, they responded as if an unfamiliar prototype were more familiar than a previously observed exemplar (Bomba & Siqueland, 1983).

Subsequent investigations explored whether young infants could form perceptual category representations for more computationally complex visual stimuli (reviewed in Quinn & Eimas, 1996). In these studies, realistic color photographs of individual exemplars of basic-level animal categories (e.g., cats) and furniture (e.g., chairs) were presented to 3- and 4-month-olds. At a narrowly tuned, basic level, infants were found to form a category representation of domestic cats that included novel cats but excluded birds, dogs, horses, and tigers and a category representation for chairs that included novel chairs but excluded couches, beds, and tables. At a broader, more global level, infants were observed to form a category representation of mammals that included instances of novel mammal categories but excluded birds, fish, and furniture and a category representation of furniture that included exemplars of novel furniture categories but excluded mammals. The category-formation processes of infants can thus be viewed as flexible and responsive to the variability of the characteristics of the input.

These impressive categorization abilities in such young infants raise the question of what causes category-specific looking preferences to emerge. One possibility is that the looking preferences in testing reflect category-specific knowledge of the outside world that the infants bring to the experiment. In its extreme form, proponents of this view would argue that infants recognize the photographs as representations of objects in the world with which they are already familiar and for which they have previous categorical knowledge. Proponents of a less extreme version would argue that although young infants do not actually possess categories of *cat*, *dog*, *horse*, and so on, they recognize the images as representations of animals and can partition those animals into very specific subsets. According to either of these arguments, the familiarization phase in these studies serves to prime the knowledge that the infants have already acquired outside the laboratory.

However, an alternative account based on somewhat unexpected empirical data casts doubt on these hypotheses. Under some familiarization conditions, infants presented with cats will form a perceptual category representation of *cat* that excludes dogs but will not form a perceptual category representation for *dog* that excludes cats (Quinn, Eimas, & Rosenkrantz, 1993). That is, the *cat* and *dog* categories have asymmetric exclusivity: *Cat* excludes dogs, but *dog* does not exclude cats. We have suggested that the exclusivity difference might reflect an asymmetric relation in the distribution of feature values used to characterize the two sets of images (Mareschal & French, 1997; Mareschal et al., 2000).

To provide support for this hypothesis, Mareschal and colleagues (French, Mermillod, Quinn, & Mareschal, 2001; Mareschal et al., 2000, 2002) used a connectionist autoencoder network to simulate the category-based looking-time behaviors of 3- to 4-month-olds, including the subtle asymmetric exclusivity in the extensions of the categories tested, such as *cat* and *dog* (Quinn et al., 1993). Autoencoder networks are feedforward networks with one layer of hidden units and an equal number of input and output units. The task of the network is to reproduce the input pattern on the output pattern. Because there is no explicit training signal, this is sometimes called *self-supervised learning* (Japkowicz, 2001).

The network was trained on measurements of 10 features of the actual stimuli presented to the infants (i.e., horizontal extent,

vertical extent, leg length, head length, head width, eye separation, ear separation, nose length, nose width, and ear length). We selected features on the basis of what is known about where infants look when observing pictures of animals (specifically, with a focus on parts of the head and face region; see French et al., 2001). Note that our coding of features (e.g., “legs”) is not intended to imply that infants possess conceptual categories for such parts; rather, the features refer to surface attributes of the stimuli that are potentially available to low-level parsing routines that segment a shape into a number of component parts (Biederman, 1987; Marr & Nishihara, 1978; Zhu & Yuille, 1996). The simulations revealed that in the image set used by Quinn et al. (1993), the large majority of cat feature values were subsumed within the distribution of dog feature values. In addition, the majority of dog feature values did not fall within the distribution of cat feature values. Thus, at the level of individual features, most cats were plausible dogs, but most dogs were not plausible cats. This difference in the distribution of features for the two categories could thus explain the observed behaviors of the infants if the infants were assumed to acquire their looking-time behaviors solely from the statistical distributions of the perceptual features encountered during familiarization, as was the case with the networks.

If it is true that infant looking-time behaviors are closely yoked to the statistical distribution of features in the familiarization stimuli, then it should be possible to manipulate infant looking-time patterns by manipulating the distribution of features in the familiarization stimuli. Specifically, if young infant category learning can be explained purely by the observed statistical distributions of the perceptual patterns, then by manipulating nothing more than the statistics of these distributions we should be able to reverse this categorization asymmetry or make it disappear altogether. We first show via simulations using the autoencoder model that when perceptual input has been manipulated either to reverse or to eliminate the original inclusion relationship (i.e., the *dog* category largely includes cats, but not vice versa), the originally observed categorization asymmetry is reversed or disappears, respectively.

The simulations and corresponding experiments are organized as follows. In Simulation 1, we use the autoencoder network to investigate what happens to the asymmetry when the inclusion relation in the cat and dog stimuli is reversed. With Experiments 1, 2, and 3, we test the network’s predictions with 3- to 4-month-olds. In Simulation 2, we investigate the nature of the internal representations developed by the autoencoder model in Simulation 1 and show how closely they resemble both network output and infant performance. In Simulation 3, we examine what happens to the asymmetry when the inclusion relation is removed from the cat and dog stimuli but unequal variances are maintained. Finally, in Experiments 4 and 5, we test the predictions of the model with the removal of the inclusion relation.

### Simulation 1: Reversing the Cat–Dog Asymmetry

In this simulation, we explored the nature of the asymmetrical exclusivity relation when networks are trained on data derived from images in which the distributions of dog feature values are subsumed by the corresponding distributions of cat values. If, as argued in Mareschal and French (1997) and Mareschal et al. (2000, 2002), autoencoder networks develop representations that reflect

the distribution of features in the training set, then we would expect networks trained on this new set of cat exemplars to develop a category of *cat* that includes novel cats and novel dogs. In contrast, networks trained on the new dog exemplars should develop a category of *dog* that includes novel dogs but excludes novel cats.

The images from which the data for this simulation are derived are shown in Figure 1. These image sets were created to reverse the variance and inclusion relations observed in the original images used by both Quinn et al. (1993) and Mareschal et al. (2000, 2002). This was achieved by carefully selecting from Schuler (1980) and Siegal (1983) various perceptually similar breeds of dogs and perceptually highly distinct breeds of cats. In this way, the 18 cats selected constituted the broad category, whereas the 18 dogs chosen were the narrow category. When measuring the features, we normalized the raw values across both sets of animals such that the maximum value was 1.0 for all features. In addition, if the feature could not be measured because, for example, only one ear was showing in an image, that feature was given a value of 0. Finally, we very slightly modified a number of these images by changing the values of the features to be measured (8 from each category). To ensure that modified images did not introduce a bias, we included an identical number of modified images in both the dog and the cat stimulus sets. The feature-value distributions of the resulting cat and dog image sets are shown in Figure 2. In the original image sets from Quinn et al. (1993), the average variance over all 10 features of the dogs was 1.63 times that of the cats, whereas in this image set, the average variance of cats was 3.12 times that of dogs. Note especially that the inclusion relation found in the original image set has been reversed.

To validate the images used here and in later studies, we presented 10 adult participants (4 men and 6 women, mean age = 28.9 years, range 23–40 years) with two blocks of 36 images. The first block contained the cat and dog images used in Experiment 1 (Figure 1) and the second block contained the cat and dog images used in Experiment 4 (Figure 3). The images were randomized and presented one at a time using PowerPoint software. The block presentation order was counterbalanced. Participants were told that they would see 36 images from two categories. They were then asked to look through the slides and to write down on an answer sheet the slide number of the image that they were viewing and into which category they would sort the image. They were given no indication as to what the categories might be. The results were as follows. First, all 10 participants spontaneously used the cat and dog labels to identify the sorting categories. Nine participants correctly identified 36 out of 36 images as cat or dog, and 1 participant correctly identified 35 out of 36 images as cat or dog in Block 1. Nine participants correctly identified 36 out of 36 images as cat or dog, and 1 participant correctly identified 35 out of 36 images as cat or dog in Block 2. By the binomial test, the probability of observing 35 or more correct responses out of 36 is  $p < .0001$ . Thus, adults correctly identify these images as cats or dogs equally well in both the image set used for Experiment 1 and that used for Experiment 4.

In all other respects, the network simulations reported here were identical to those in Mareschal and French (1997) and Mareschal et al. (2000, 2002). All simulations are based on the performance of a standard 10–8–10 feedforward backpropagation network with a learning rate of 0.1, momentum of 0.9, and a sigmoid prime

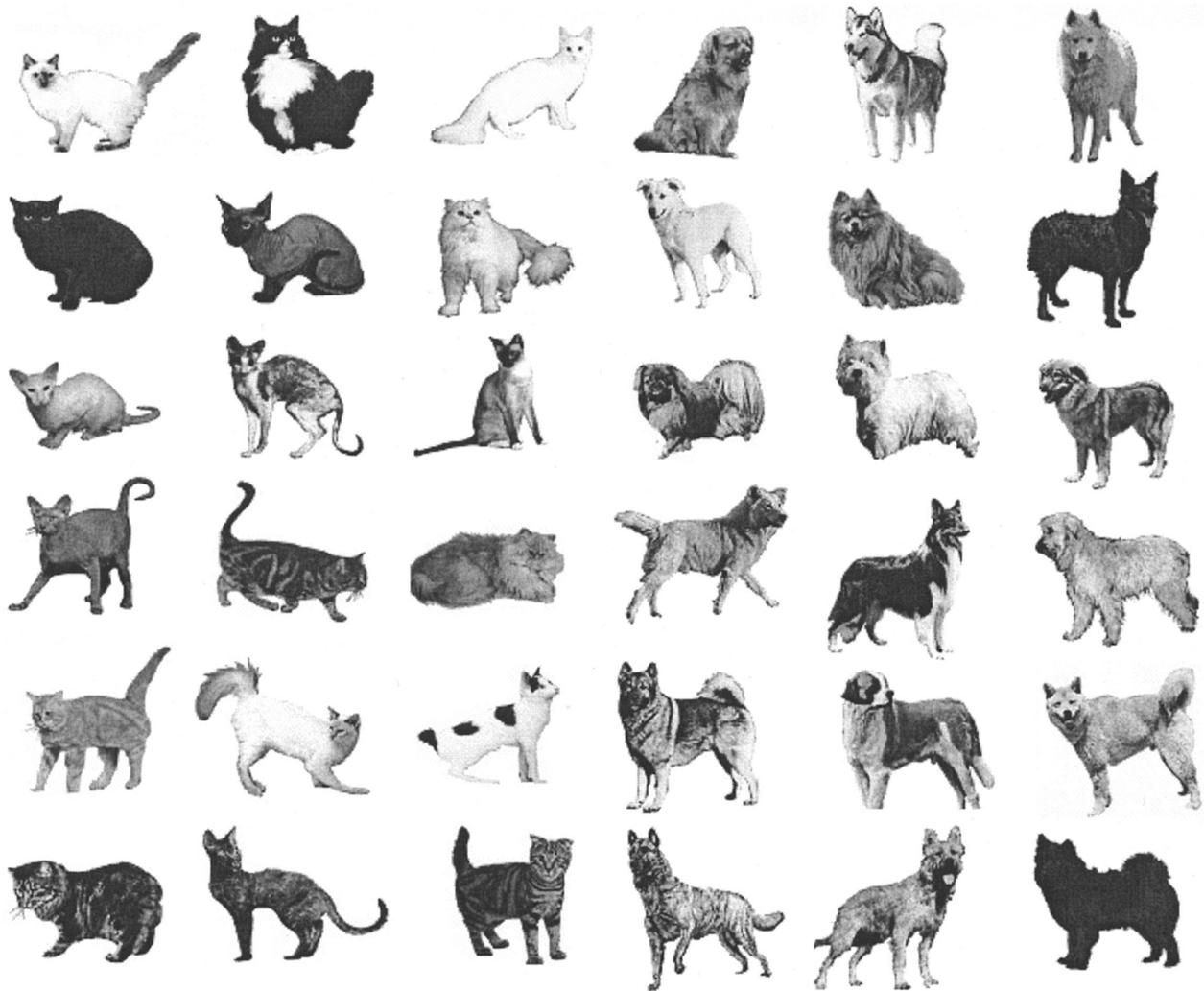


Figure 1. Cat and dog exemplars used as input in Simulation 1. Cat pictures are reprinted from *Simon & Schuster's Guide to Cats*, by M. Siegal (Ed.), 1983, New York: Simon & Schuster. Copyright 1983 by Monadori Publishers, Milan, Italy. Reprinted with permission. Dog pictures are reprinted from *Simon & Schuster's Guide to Dogs*, by E. M. Schuler (Ed.), 1980, New York: Simon & Schuster. Copyright 1980 by Monadori Publishers, Milan, Italy. Reprinted with permission.

offset (Fahlman, 1988) of 0.1. Twelve items from one category were presented sequentially to the network in groups of two (i.e., weights were updated in batches of two) to capture the fact that pairs of pictures were presented to the infants during the familiarization trials. Networks were trained for 250 epochs (weight updates) on one pair of patterns before being presented with the next pair. This was done to reflect the fact that in the Quinn and Eimas (1996) studies, infants were shown pairs of pictures for a fixed duration of time. The total amount of training was  $6 \times 250 = 1,500$  weight updates. The results are averaged over 50 network replications, each with random initial weights. The remaining 6 items from the training category and all 18 items from the non-training category were used to test whether the networks had formed category representations.

Finally, the model rests on an analogy between the weight adjustments that occur when a network learns to autoencode a

stimulus and the representation construction hypothesis proposed by Sokolov (1963) in his comparator theory of habituation to explain what drives infant looking in visual familiarization tasks. A stimulus that is poorly autoencoded by a network (i.e., one for which the output of the autoencoder differs significantly from the input stimulus) will take longer (in terms of the number of iterations) to learn (i.e., autoencode properly) than will a stimulus that produces network output very similar to itself. Similarly, infants look longer at stimuli for which they have poor internal representations (for a detailed explanation of this correspondence between network error and infant looking time, see Mareschal et al., 2000).

Figure 4A shows the networks' response when presented a novel cat and a novel dog during testing. A low output error means that the novel exemplar has been correctly autoencoded; that is, the network was able to assimilate the exemplar to its internal repre-

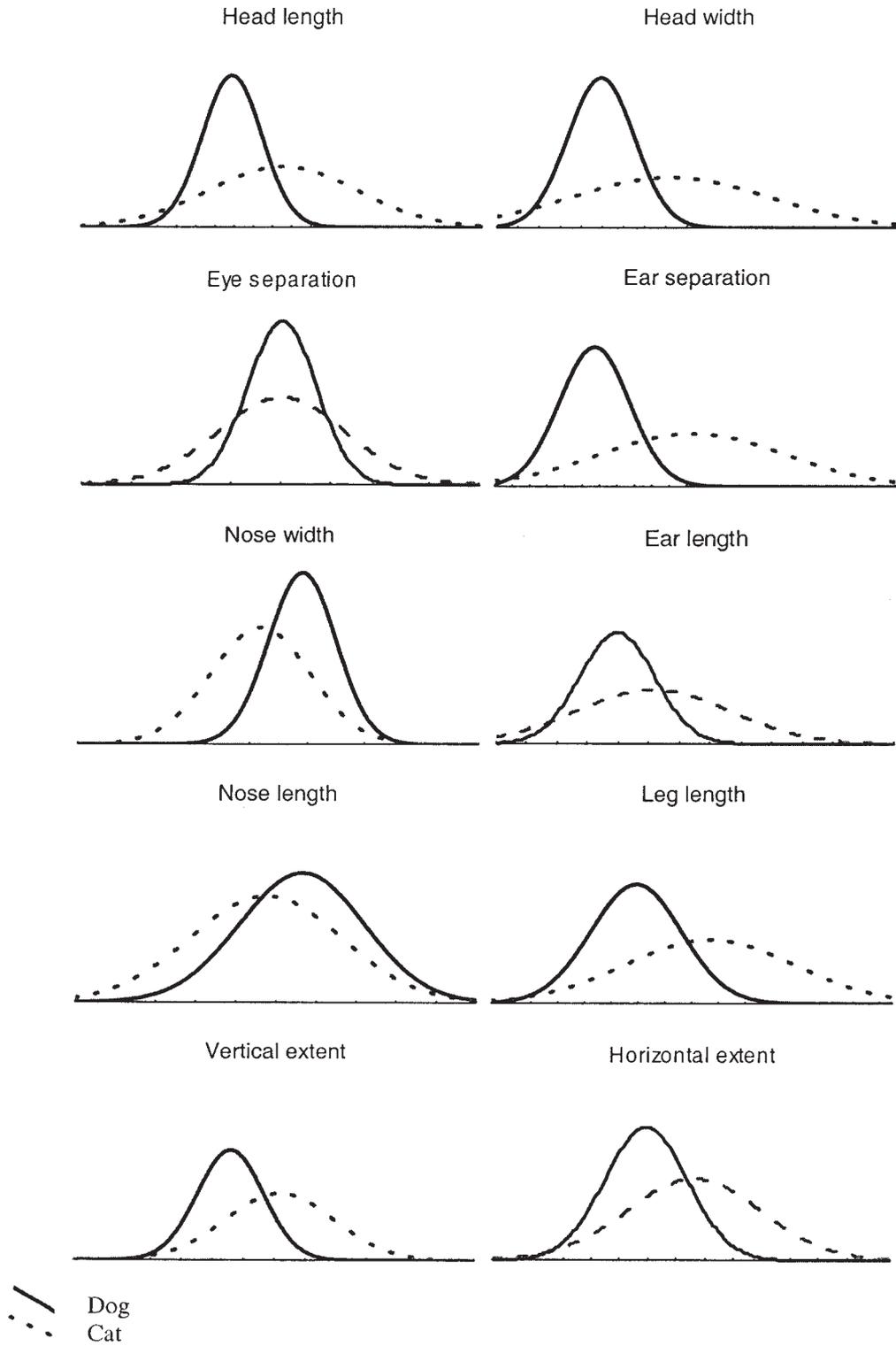


Figure 2. Feature-value distributions for exemplars in Figure 1. The cat distributions subsume the dog distributions.

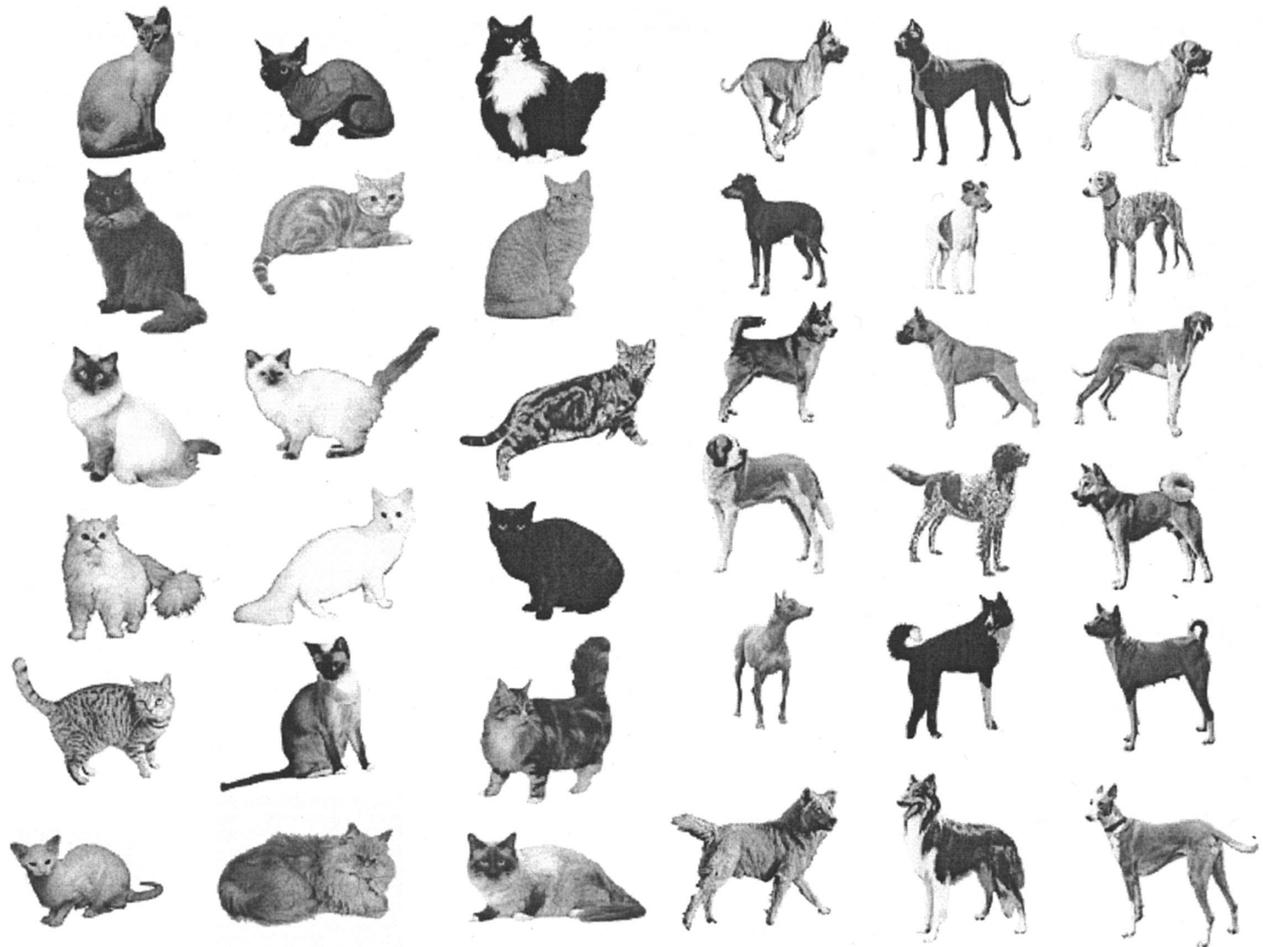


Figure 3. Cat and dog exemplars used as input in Simulation 3. Cat pictures are reprinted from *Simon & Schuster's Guide to Cats*, by M. Siegal (Ed.), 1983, New York: Simon & Schuster. Copyright 1983 by Monadori Publishers, Milan, Italy. Reprinted with permission. Dog pictures are reprinted from *Simon & Schuster's Guide to Dogs*, by E. M. Schuler (Ed.), 1980, New York: Simon & Schuster. Copyright 1980 by Monadori Publishers, Milan, Italy. Reprinted with permission.

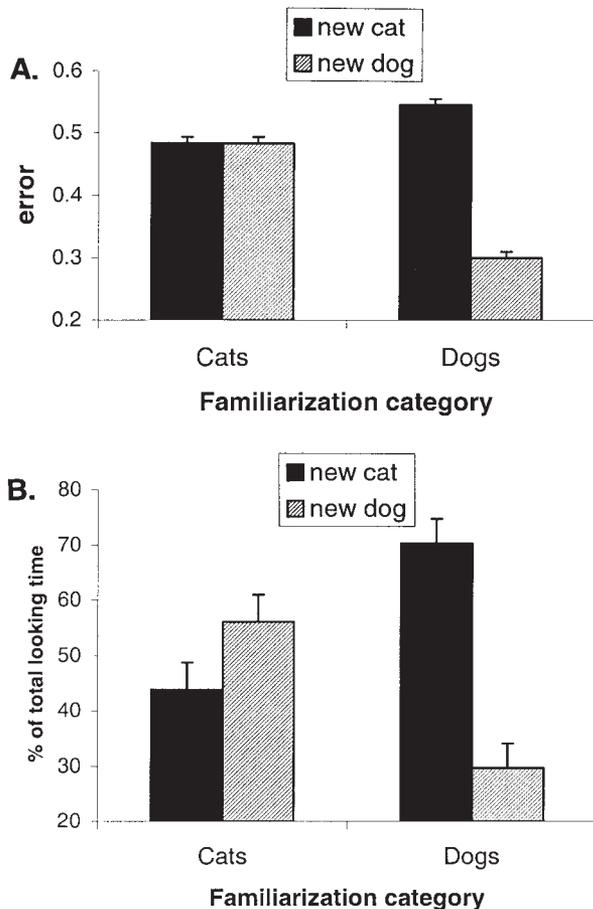
sentation and respond appropriately (Mareschal et al., 2000).<sup>1</sup> Networks trained with cats showed an equivalent amount of output error for both the novel cat and the novel dog exemplars. In contrast, networks trained on dogs show much more output error for the novel cat exemplar than the novel dog exemplar. Thus, the networks appear to have formed a category of *cat* that includes novel cats and novel dogs, and a category of *dog* that includes novel dogs but excludes novel cats. As predicted by the inclusion relation account of the exclusion asymmetry, these networks show an asymmetric exclusivity that is the reverse of that shown by the networks trained on the original set of images.

The model results confirm that the exclusivity of the categories acquired by the networks is predicted by the inclusion relations among the feature distributions of the two sets of stimuli. The prediction of the model, then, is that we should observe this same reversal with young infants. In Experiments 1, 2, and 3 below, we examine whether the asymmetric exclusivity in the category representations formed by 3- to 4-month-olds is equally dependent on the inclusion relation in the stimulus sets.

### Experiment 1: Reversing the Quinn et al. (1993) Categorization Asymmetry

In Experiment 1, we presented infants with stimuli drawn from the cat and dog images used to train the networks in Simulation 1. In these images, the dog feature value distribution is generally subsumed within the corresponding cat feature value distribution. The prediction derived from the network simulation is that infants familiarized with images of cats will show no significant preference when subsequently tested with a novel dog or a novel cat

<sup>1</sup> It is important not to confuse the notion of network error with the term *error rate*, commonly used in experimental psychology. Network error for the autoencoder network is determined by the maximum error across all output nodes when they are compared with the input stimulus. This measures how well the network is able to autoencode each individual stimulus. The larger this error, the longer, in general, it will take the network to find a set of weights such that this error falls below a preset criterion.



**Figure 4.** A. Average output error in response to a novel cat and a novel dog in Simulation 1. The error measure used (shown by the error bars) is the standard error of the mean. B. Reversal of the asymmetric categorization observed in Quinn et al. (1993). The error measure used (shown by the error bars) is the standard error of the mean.

image. In contrast, infants familiarized with dogs will show a significant preference for a novel cat over a novel dog in subsequent testing.

### Method

**Participants.** The participants in Experiment 1 were 24 infants approximately 3 to 4 months of age ( $M = 3$  months 18 days;  $SD = 9.29$  days). Fifteen of the infants were girls and 9 were boys.

**Stimuli.** The stimuli were 36 color photographs of cats and dogs (18 exemplars of each category) described in Simulation 1 and shown in Figure 1. We varied the stimuli for body orientation (i.e., left or right facing), and the animals were chosen so as to have a relatively uniform color distribution over the set of animals. The sizes of the images were all made approximately the same, so, for example, German shepherds were approximately the same size as Pekingese. Each picture consisted of a single animal that had been cut away from its background and mounted onto a white 17.7- × 17.7-cm posterboard for presentation.

**Apparatus.** All infants in each experiment were tested in a visual preference apparatus modeled on the one described by Fagan (1970). The apparatus is a large, three-sided gray viewing chamber that is on wheels. It has a hinged gray display panel onto which were attached two compart-

ments to hold the posterboard stimuli. The stimuli were illuminated by a fluorescent lamp that was shielded from the infant's view. The center-to-center distance between compartments was 30.5 cm, and on all trials the display panel was situated approximately 30.5 cm in front of the infant. There was a 0.62-cm peephole located midway between the two display compartments that permitted an observer to record the infant's visual fixations. A second peephole, 0.90 cm in diameter, was located directly below the first peephole and permitted a Pro Video CVC-120PH pinhole camera (Provideo, Amityville, NY) and a JVC video recorder to record infants' gaze duration.

**Procedure.** All infants underwent the following general procedure. They were brought to the laboratory by a parent and seated in a reclining position on the parent's lap. There were two experimenters, both of whom were naive to the hypotheses under investigation. The first experimenter positioned the apparatus so that the midline of the infant's head was aligned with the midline of the display panel. When the display panel was open, the infant could see the experimenter from the midsection up in addition to a portion of the room that was a light background color. The experimenter selected the appropriate stimuli as previously determined for the forthcoming trial and loaded them into the compartments of the display panel from a nearby table. The experimenter then elicited the infant's attention and closed the panel, thereby exposing the stimuli to the infant. The parent was unable to see the stimuli.

During each trial, the first experimenter observed the infant through the small peephole and recorded visual fixations to the left and right stimuli by means of two 605 XE Accusplit electronic stop watches (Accusplit, San Jose, CA), one of which was held in each hand. The second experimenter timed the fixed duration of the trial, signaled the end of the trial, and was positioned at a location in the testing room that prevented her from observing the stimuli. Between trials, the first experimenter opened the panel, changed the stimuli, obtained the infant's attention, centered his or her gaze, and closed the panel. When estimates of spontaneous preference were obtained (i.e., in Experiment 3), the second experimenter did not participate other than to time the trials and signal when a trial was to end. In experiments in which the discrimination or categorization of stimuli was measured (Experiments 1, 2, 4, and 5), the first and second experimenters changed places for the test trials. The experimenter who presented stimuli and measured infant fixations during familiarization now measured trial duration and signaled the end of each test trial, whereas the second experimenter presented the test stimuli and measured infant fixations. The second experimenter was always naive with respect to the familiar category (Experiments 1 and 4) or familiar stimulus that was presented when the discriminability of two exemplars was measured (Experiments 2 and 5). The two experimenters changed roles across infants.

To handle possible side biases (i.e., position preferences) sometimes displayed by individual infants, over all the trials, both familiarization and test, the looking time to one side (i.e., left or right compartment) of the display stage could not exceed 95% of the total looking time for the infant to be included in the data analysis. In addition, on preference test trials, each infant was required to look at both of the stimuli. Interobserver reliability was later determined by comparing the looking times measured by the experimenter using the center peephole and an additional naive observer recording the looking times offline from videotape records; the interobserver reliability averaged .94.

Twelve infants were randomly assigned to each group, defined by the familiar category, *cats* or *dogs*. Within each group, each infant was presented with 12 randomly selected pictures of cats or dogs. On each of six 15-s familiarization trials, 2 of the 12 stimuli, again randomly selected, were presented. Immediately after the familiarization trials, two 10-s preference test trials were administered in which a novel cat was presented along with a novel dog. There were 12 such pairs, randomly selected, and each pair, which was seen on both test trials, was assigned to 1 infant who had seen dogs and 1 infant who had seen cats during the familiarization trials. The test-trial stimuli were thus identical for both groups of infants.

Left–right positioning of the novel exemplar from the novel category was counterbalanced across infants on the first test trial and reversed on the second test trial.

**Results and Discussion**

*Familiarization trials.* Individual looking times were summed over the two stimuli that were presented on each trial and then averaged across the first three and the last three trials. The mean looking times and standard deviations are shown in Table 1. An analysis of variance with familiar category (*cats* vs. *dogs*) as a between-subject factor and trials (1–3 vs. 4–6) as a within-subject factor revealed no significant effects. For the effect of familiar category,  $F(1, 22) = 2.01, MSE = 14.99, p = .167$ . For the effect of trials,  $F(1, 22) = 0.22, MSE = 3.39$ , and for the interaction of familiar category and trials,  $F(1, 22) = 0.09, MSE = 3.39$ . As has been the case in previous studies using cat and dog stimuli, the multidimensional variation among the pictorial exemplars of the cats and dogs was sufficient to maintain infant attention throughout the familiarization trials (e.g., Eimas & Quinn, 1994). Important, though, was the fact that there were no significant differences observed in the degree of encoding for either the cat or the dog familiarization stimuli.

*Preference test trials.* The total looking time of each infant across the two test trials to the novel stimulus from the novel category was divided by the total looking time to both test stimuli and converted to a percentage score. The mean novel category preference scores are shown in Table 1. The  $t$  tests comparing the scores with chance performance (50%) revealed that infants familiarized with dogs preferred the novel cats, but infants familiarized with cats did not prefer the novel dogs. In addition, the two means were significantly different from each other,  $t(22) = 2.51, p < .02, \eta^2 = .22$ . These results are consistent with the model prediction. They are the opposite of what has previously been reported when infants were familiarized with the original images (Mareschal et al., 2000, 2002; Quinn et al., 1993). Thus, reversing the feature inclusion relation reverses the novel category preferences in 3- to 4-month-olds (Figure 4B). However, it is possible that the looking patterns observed reflect either an inability to discriminate dog exemplars—the novel dog presented on the test trial being perceived simply as the familiar exemplar and not as a different member of the familiar category—or a spontaneous preference for the cat exemplars over the dog exemplars. We tested these possibilities in two control experiments, Experiments 2 and 3.

**Experiment 2: Within-Category Discrimination Control for Experiment 1**

In Experiment 2, we tested whether infants were able to discriminate among the exemplars within the cat and dog categories used in Experiment 1.

*Method*

*Participants.* The participants in Experiment 2 were 24 infants approximating 3 to 4 months of age ( $M = 3$  months 19 days;  $SD = 7.01$  days). Fourteen of the infants were girls and 10 were boys. Two additional infants were tested but did not complete the procedure because of fussiness.

*Stimuli and apparatus.* The stimuli and apparatus were the same as those used in Experiment 1.

*Procedure.* Each infant was randomly allocated to a cat-discrimination or a dog-discrimination group. Infants in the cat-discrimination group were presented with a different randomly selected pair of cats, one of which was arbitrarily designated the familiar stimulus and the other the novel stimulus. The familiar stimulus was duplicated and shown in both compartments of the display panel for a single 15-s trial. There were then two 10-s trials during which the familiar and novel cats were displayed, with their left–right positions counterbalanced across trials. Infants in the dog-discrimination group were presented with the same sequence of events, but with dog images rather than cat images.

**Results and Discussion**

*Familiarization trial.* Individual looking times were summed over the left and right copies of the stimulus on the familiarization trial and then averaged across infants. Mean looking times to the individual cats and dogs on the single familiarization trial are shown in Table 2. The difference in mean looking times between the two experimental groups was not significant,  $t(22) = -0.11, p > .20$ .

*Preference test trials.* Each infant’s looking time to the novel stimulus was divided by the looking time to both test stimuli and then converted to a percentage score. The mean novelty preference scores are shown in Table 2. The  $t$  tests comparing the scores with chance performance revealed that infants in both experimental groups displayed reliable novelty preferences. In addition, the two means were not significantly different from each other,  $t(22) = 1.01, p > .20$ . These preferences show that the infants were capable of discriminating between the individual cats and also between the individual dogs presented in Experiment 1.

Table 1  
Mean Fixation Times (in Seconds) During the Familiarization Trials and Mean Novel-Category Preference Scores (in Percentages) During the Preference Test Trials in Experiment 1

Familiarization category	Fixation time				Novelty preference		$t^a$	$\eta^2$
	Trials 1–3		Trials 4–6		$M$	$SD$		
	$M$	$SD$	$M$	$SD$				
Cats	7.33	3.17	6.92	3.28	56.14	16.98	1.25	
Dogs	8.75	2.21	8.66	3.32	72.73	15.36	5.13****	.69

<sup>a</sup> The  $t$  tests compared the mean scores with chance performance.  
\*\*\*\*  $p < .0005$ , one-tailed.

Table 2  
*Mean Fixation Times (in Seconds) During the Familiarization Trial and Mean Novelty Preference Scores (in Percentages) During the Preference Test Trials in Experiment 2*

Experimental group	Fixation time		Novelty preference		$t^a$	$\eta^2$
	$M$	$SD$	$M$	$SD$		
Cat	7.81	3.04	62.30	17.16	2.48**	.34
Dog	7.97	3.91	56.39	10.69	2.07*	.26

<sup>a</sup> The  $t$  tests compared the mean scores with chance performance.  
 \*  $p < .05$ , one-tailed. \*\*  $p < .025$ , one-tailed.

### Experiment 3: Spontaneous Preference Control for Experiment 1

In Experiment 3, we tested whether infants had a prior preference for either the dog or the cat stimuli.

#### Method

**Participants.** The participants in this experiment were 12 infants approximating 3 to 4 months of age ( $M = 3$  months 17 days;  $SD = 10.81$  days). Five of the infants were girls and 7 were boys.

**Stimuli and apparatus.** The stimuli and apparatus were the same as those used in Experiment 1.

**Procedure.** All infants were presented with the preference test trials described for Experiment 1 but not the prior familiarization trials.

#### Results and Discussion

The looking times to cats were summed over both trials for each infant, divided by the total looking time to cats and dogs, and converted to percentages. The mean preference for cats was 49.31%,  $SD = 15.68$ . This preference was not significantly different from chance,  $t(11) = -0.15$ ,  $p > .20$ . In addition, the preference for cats after familiarization with dogs in Experiment 1 was found to be reliably higher than the spontaneous preference for cats with no familiarization with dogs,  $t(22) = 3.70$ ,  $p < .01$ ,  $\eta^2 = .39$ . Thus, infants showed no spontaneous preference to look at the cat or dog stimuli used in Experiment 1.

### Discussion of Experiments 1, 2, and 3

In Experiment 1, we found that when 3- to 4-month-olds were familiarized with the cat stimuli selected from those in Figure 1, they showed no subsequent novel category preference when tested with a novel cat and a novel dog. In contrast, when 3- to 4-month-olds were familiarized with the dog stimuli selected from those in Figure 1, they displayed a subsequent novel category preference for cats when tested with a novel cat and a novel dog. In addition, Experiment 2 revealed that 3- to 4-month-olds could discriminate between individual exemplars of the cat and dog stimuli, and Experiment 3 showed that they had no a priori preference for looking at either the cat or the dog stimuli. Thus, the asymmetric exclusivity found in Experiment 1 does not arise from either an inability to discriminate items within a category or a prior preference for one category over the other.

Taken together, the results of Experiments 1, 2, and 3 confirm the model predictions made in Simulation 1. That is, when presented with stimuli in which the distribution of dog feature values is subsumed within the distribution of cat feature values, infants will form a category of *dog* that excludes cats and a category of *cat* that does not exclude dogs. These results support the argument made by Mareschal and French (1997) and Mareschal et al. (2000) that infants are generating the *cat* and *dog* categories online and in response to the particular exemplars encountered during familiarization. Changing the distribution of features in the familiarization exemplars changes the pattern of novel category preferences of the infants. In the next simulation, we explore how this behavior relates to the way the networks represent the categories internally.

### Simulation 2: Internal Representations of the Autoencoder

As the autoencoder learns, it develops internal representations of the external stimuli. When it is presented with new stimuli, it also represents them internally. We therefore examined the internal representations of the autoencoder to determine to what extent they reflected the categorization performance of the network. The technique we used was as follows. With the same settings as in Simulation 1, we first trained the network on 12 items from one category. We then tested the network on all 36 stimuli (18 dogs, 18 cats), the 12 training exemplars, and the 24 novel items (6 from the training category, 18 from the unseen category).

We next examined the relationship between the hidden-unit representations of all the stimuli. To do this, we computed the Euclidean distance from the location in the  $n$ -dimensional hidden unit space for each of the novel stimuli to the location of the nearest training stimulus. This provides an indication of the distribution of the internal representations of the new stimuli with respect to the internal representations of the previously learned training stimuli. The key idea is that the farther away a new representation is from any previously learned representation, the greater the error on output of the network for the corresponding stimulus. The results reported are averaged values over 20 runs of the program.

In Figure 5, it can be seen that when trained on the broad category—in this case, cats—the distances of the internal representations of new cats and new dogs to the closest representation of an exemplar in the training set are approximately the same. In contrast, when trained on the narrow category *dogs*, largely included in the broader *cat* category, there are more novel cats than novel dogs whose internal representations are farther from the representations of the previously learned dogs. In short, after training on cats, when presented with a novel cat or a dog, the network is internally at approximately equal distance from a previously learned instance in the *cat* category (because dogs are, in this case, included in the *cat* category). However, after training on the narrow category (*dog*), the internal representation for any previously learned dog is considerably closer to a new dog than to any cat among the test stimuli.

In sum, the performance of the network in Simulation 1 (Figure 4A) and the corresponding performance of young infants (Table 1 and Figure 4B) reflect the structure of the internal category representations acquired during training. In particular, they reflect the distances between internal representations of the previously learned items to the novel items of both categories.

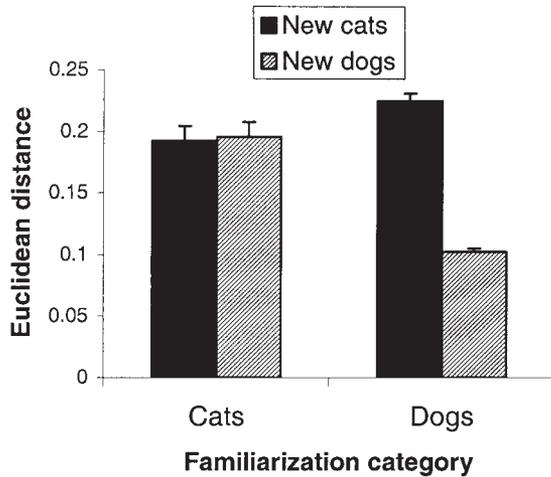


Figure 5. The average distances of the internal representations of novel items to the internal representations of previously learned exemplars. The error measure used (shown by the error bars) is the standard error of the mean.

### Simulation 3: Removing the Cat–Dog Asymmetry

An issue that remains unclear from the experiments and simulations reported thus far centers on what aspects of the feature distribution cause the asymmetry to appear. In the two image sets used (i.e., the original images used by Quinn et al., 1993, and the one described in Simulation 1), the subsumed distribution was also narrower than the subsuming distribution. Thus, infants may have been responding to the differences in variability of the image sets rather than the inclusion relations. Although the fact that there are no significant differences between the two groups during familiarization argues against this interpretation, it is possible that the infants simply did not learn the broad category as deeply as the narrow category and hence were unable to exclude relatively similar exemplars that were nevertheless members of the contrasting category. In Simulation 3, we use the autoencoder model to examine whether the asymmetric exclusivity is removed when the inclusion relation between the image sets is also removed.

In this simulation, we explore the nature of the exclusivity asymmetry when networks are trained on data derived from images in which the variance of the distributions of cat and dog feature values are unequal but do not subsume one another. If, as argued in Mareschal et al. (2000), the autoencoder networks develop representations that reflect the distribution of features in the training set, then we would expect networks trained on this new set of cat exemplars to develop a category of *cat* that includes novel cats but excludes novel dogs. Similarly, networks trained on the new dog exemplars will develop a category of *dog* that includes novel dogs but excludes novel cats. In contrast, if the asymmetric exclusivity is derived from the differences in variance between the two feature distributions, then we would expect the networks to show an asymmetry, with the narrower category excluding exemplars of the broader category.

The images from which the data for this simulation are derived are shown in Figure 3. These image sets were created to remove the inclusion relation observed in the original images used by

Quinn et al. (1993), Mareschal and French (1997), and Mareschal et al. (2000, 2002). As in Simulation 1, the images were produced by selecting sets of 18 dogs and 18 cats from Schuler (1980) and Siegal (1983). Eleven of the images from each category were very slightly modified using the same methods described in Simulation 1. As before, to ensure that modified images did not introduce a bias, we included an identical number of modified images in both the dog and the cat stimulus sets. We compared the distributions of each of the 10 features to ensure that there was minimal overlap of the distributions of feature values.

The feature-value distributions of the resulting cat and dog image sets are shown in Figure 6. In the original image sets used by Quinn et al. (1993), the average variance over all 10 features of the dogs was 1.63 times that of the cats. In the present image set, the average variance of dogs was intentionally kept greater than that of cats. Dogs had an average variance across the 10 features that was 1.40 times that of cats. It is crucial to note that with these two sets of images, both the inclusion relation found in the original image set used by Quinn et al. (1993) and the reverse inclusion relation found in the images used in Experiment 1 have been largely removed.

In all other respects, the network simulations reported here were identical to those in Simulation 1. The results are averaged over 50 network replications, each with random initial weights. The remaining six items from each category were used to test whether the networks had formed category representations.

Figure 7A shows the networks' response when presented with a novel cat and a novel dog following training with 12 exemplars of cats or dogs. Networks trained with cats show more output error for the novel dog exemplar than the novel cat exemplar. Similarly, networks trained on dogs show more output error for the novel cat exemplar than the novel dog exemplar. Thus, the networks appear to have formed a category of *cat* that includes novel cats but excludes novel dogs and a category of *dog* that includes novel dogs but excludes novel cats. As predicted by the inclusion relation account of the exclusion asymmetry, these networks do not show an asymmetry in the exclusivity of the categories acquired. They have acquired an exclusive category of *cat* and an exclusive category of *dog*.

These results are consistent with the claim that the exclusivity of the categories acquired is predicted by the inclusion relation among the feature distributions of the two sets of stimuli and not the difference in variance between the two image sets. In Experiments 4 and 5 below, we confirm experimentally that the exclusivity asymmetry observed in 3- to 4-month-olds is equally dependent on the inclusion relation in the stimulus sets by showing that, as predicted by the model, this asymmetry disappears when we eliminate the inclusion relation.

### Experiment 4: Experimental Evidence for the Removal of the Categorization Asymmetry

In Experiment 4, we presented infants with images drawn from the cat and dog stimulus sets used to train the networks in Simulation 3 and depicted in Figure 3. In these images, the feature-value distributions for cats and dogs did not subsume one another. In addition, the dog distribution was broader than the cat distribution. The predictions derived from the network simulation are that infants familiarized with images of cats will show a significant

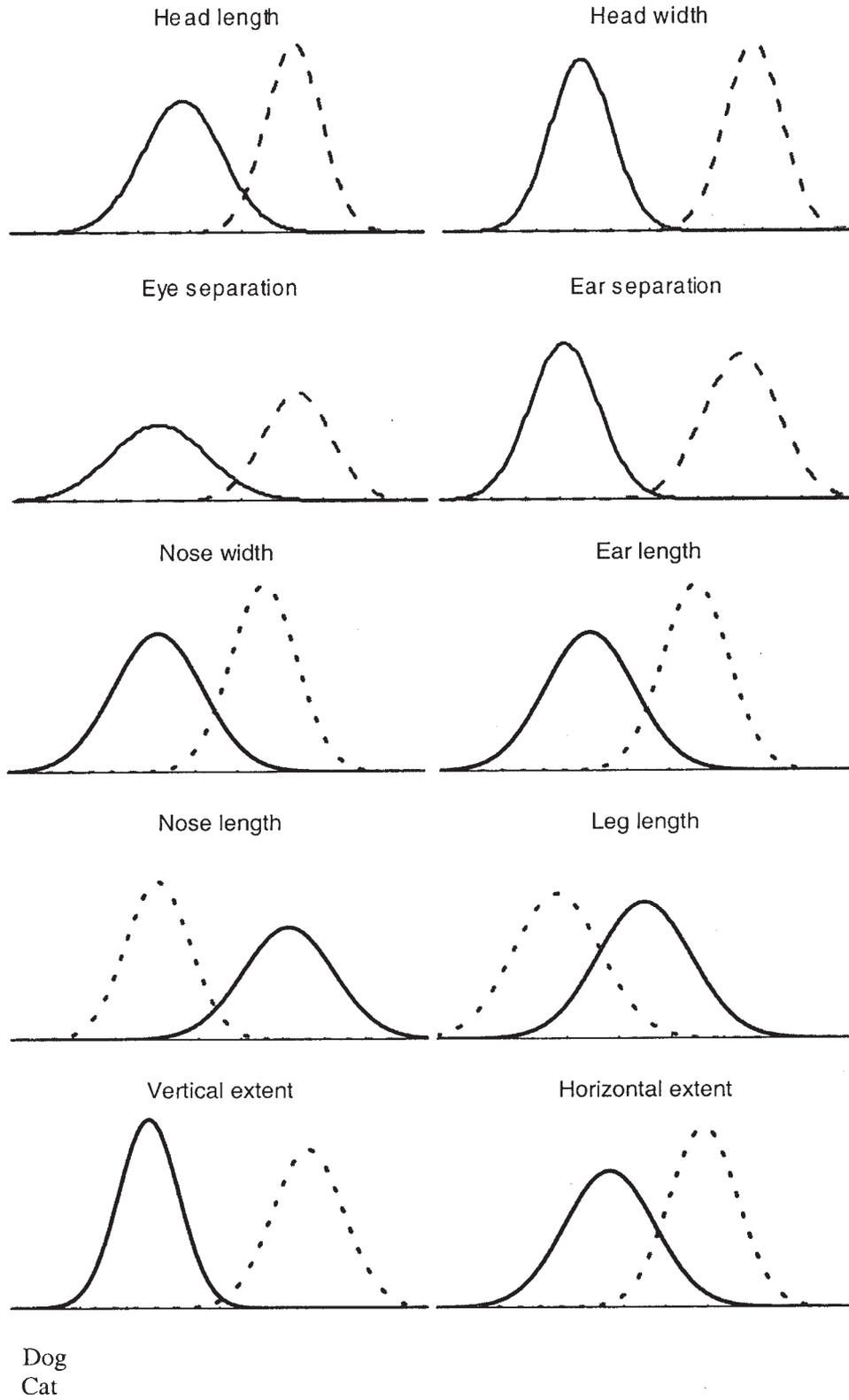


Figure 6. Feature value distributions for exemplars in Simulation 3. The cat distributions do not subsume the dog distributions and vice versa.

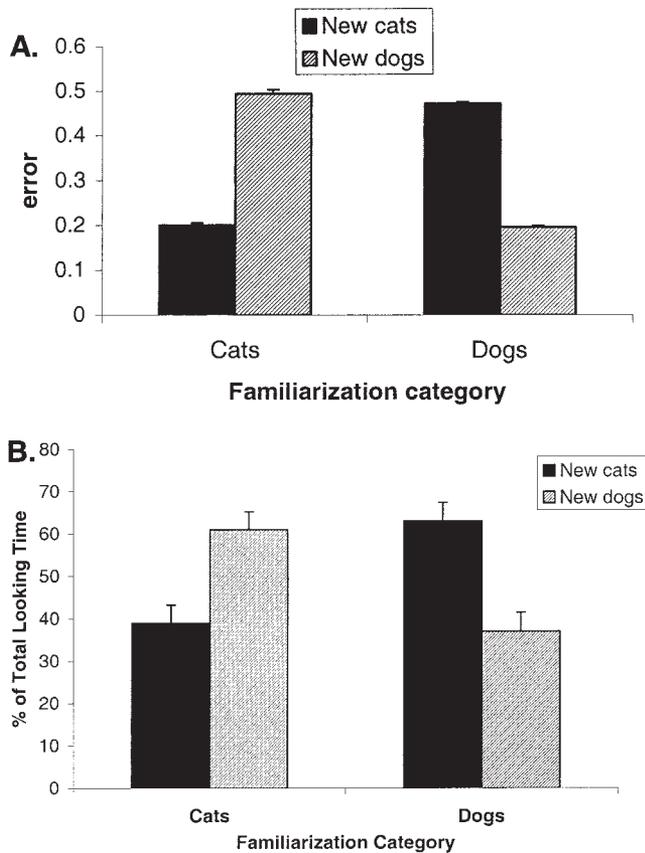


Figure 7. A. Average output error in response to a novel cat and a novel dog in Simulation 3. The error measure used (shown by the error bars) is the standard error of the mean. B. Disappearance of the categorization asymmetry for infants in Experiment 4. The error measure used (shown by the error bars) is the standard error of the mean.

novel category preference for a novel dog over a novel cat and that infants familiarized with dogs will show a significant novel category preference for a novel cat over a novel dog.

**Method**

*Participants.* The participants in Experiment 4 were 24 infants approximating 3 to 4 months of age ( $M = 3$  months 23 days;  $SD = 9.43$  days).

Thirteen of the infants were girls and 11 were boys. One additional infant was tested but did not complete the procedure because of fussiness.

*Stimuli.* The stimuli were 36 color photographs of cats and dogs (18 exemplars of each category) described in Simulation 3. Each picture consisted of a single animal that had been cut away from its background and mounted onto a white 17.7- × 17.7-cm posterboard for presentation.

*Apparatus and procedure.* The apparatus and procedure were identical to those used in Experiment 1.

**Results and Discussion**

*Familiarization trials.* Individual looking times were summed over the two stimuli that were presented on each trial and then averaged across the first three and the last three trials. The mean looking times and standard deviations are shown in Table 3. An analysis of variance with familiar category (*cats* vs. *dogs*) as a between-subjects factor and trials (1–3 vs. 4–6) as a within-subjects factor revealed no significant effects. For the effect of familiar category,  $F(1, 22) = 0.01, MSE = 7.69$ . For the effect of trials,  $F(1, 22) = 3.12, MSE = 3.21, p = .09$ , and for the interaction of familiar category and trials,  $F(1, 22) = 0.42, MSE = 3.21$ . As in Experiment 1 and in previous studies using cat and dog stimuli (e.g., Eimas & Quinn, 1994), the multidimensional variation among the pictorial exemplars of both cats and dogs was sufficient to maintain infant attention throughout the familiarization trials. It is important to note, though, that there were no differences in the degree of encoding observed for either cat or dog familiarization stimuli.

*Preference test trials.* The total looking time of each infant across the two test trials to the novel stimulus from the novel category was divided by the total looking time to both test stimuli and converted to a percentage score. The mean novel category preference scores are shown in Table 3. The  $t$  tests comparing the scores with chance performance revealed that infants familiarized with dogs preferred the novel cats and infants familiarized with cats preferred the novel dogs (Figure 7B). In addition, the two means were not significantly different from each other,  $t(22) = -0.28, p > .20$ . These results are consistent with the model prediction. Removing the feature inclusion relation removes the novel category preference asymmetry in 3- to 4-month-olds. It should be noted that this pattern of novel category preference cannot be due to an a priori preference for looking at one set of images over the other, as we find a preference for either novel cats or novel dogs depending on the familiarization regime.

Table 3  
Mean Fixation Times (in Seconds) During the Familiarization Trials and Mean Novel-Category Preference Scores (in Percentages) During the Preference Test Trials in Experiment 4

Familiarization category	Fixation time				Novelty preference		$t^a$	$\eta^2$
	Trials 1–3		Trials 4–6		$M$	$SD$		
	$M$	$SD$	$M$	$SD$				
Cats	7.02	2.54	5.92	2.57	61.33	14.76	2.66**	.37
Dogs	6.77	1.55	6.05	2.53	63.03	15.43	2.93***	.42

<sup>a</sup> The  $t$  tests compared the mean scores with chance performance.  
\*\*  $p < .025$ , one-tailed. \*\*\*  $p < .01$ , one-tailed.

Because all of the dogs in our pictures were in a standing posture and the cats in our pictures were depicted in both standing and sitting postures, it is possible that some of our novelty preference effect is driven by a contrast between images of animals that are standing and those with animals that are sitting. The critical cell for this analysis is the dog familiarization condition, because this is the cell where one would go from all standing during familiarization to either standing versus standing or standing versus sitting in the preference test trials. If the standing–sitting contrast were facilitating performance, this is where one should observe a clear difference in performance, with higher novel category preference scores in the standing–sitting contrast. However, even with small samples, the percentages of looking times toward the novel category are virtually identical whether the test pair consists of standing–standing or standing–sitting images (for the standing–standing pairs, novelty preference = 63.43%,  $SD = 16.50$ ,  $N = 7$ ; for the standing–sitting pairs = 62.46%,  $SD = 15.68$ ,  $N = 5$ ). These data are consistent with the findings that infants focus more on features associated with the heads of animals (Quinn & Eimas, 1996; Spencer, Quinn, Johnson, & Karmiloff-Smith, 1997).

To compare the novel category preferences for infants familiarized with dogs in Experiments 1 and 4 and to ensure that any significant preference differences were not due to a very small subset of babies showing the novelty preference, we also ran nonparametric tests (Wilcoxon signed rank and chi-square tests) on the raw looking-time data. Wilcoxon signed rank tests performed on these data showed that in Experiment 1, there is significantly more looking at the novel than the familiar category exemplar ( $z = 2.67$ ,  $p < .008$ , two-tailed). Similarly, in Experiment 4, this same test showed that there is significantly more looking at the novel than the familiar category exemplar ( $z = 2.12$ ,  $p < .04$ , two-tailed).

In addition, when the two experiments are compared, we find that in Experiment 1, 11 of 12 infants familiarized with dogs had preference scores for cats above the 50% chance level, whereas in Experiment 4, 8 of 12 infants familiarized with dogs had preference scores for cats above the 50% chance level. A chi-square test revealed that these distributions did not differ significantly between the two experiments,  $\chi^2(1, N = 24) = 2.27$ ,  $p > .10$ .

It is still possible that the looking pattern observed in Experiment 4 reflects an inability to discriminate cat or dog exemplars, the novel exemplar from the familiar category presented on the test trials being perceived simply as the familiar exemplar and not a different member of the familiar category. We tested this possibility in Experiment 5.

#### Experiment 5: Within-Category Discrimination Control for Experiment 4

In Experiment 5, we tested whether infants were able to discriminate exemplars chosen from within the cat and dog categories used in Experiment 4.

#### Method

**Participants.** The participants in Experiment 5 were 24 infants approximately 3 to 4 months of age ( $M = 3$  months 22 days;  $SD = 8.57$  days). Fourteen of the infants were girls and 10 were boys.

**Stimuli and apparatus.** The stimuli and apparatus were the same as those used in Experiment 4.

**Procedure.** The procedure was the same as that used in Experiment 2 (the previous within-category discrimination study).

#### Results and Discussion

**Familiarization trial.** Individual looking times were summed over the left and right copies of the stimulus on the familiarization trial and then averaged across infants. Mean looking times to the individual cats and dogs are shown in Table 4. The difference in mean looking times between the two experimental groups was not significant,  $t(22) = -0.67$ ,  $p > .20$ .

**Preference test trials.** Each infant's looking time to the novel stimulus was divided by the looking time to both test stimuli and then converted to a percentage score. The mean novelty preference scores are shown in Table 4. The  $t$  tests comparing the scores with chance revealed that infants in both experimental groups displayed reliable novelty preferences. In addition, the two means were not significantly different from each other,  $t(22) = -0.31$ ,  $p > .20$ . These preferences show that the infants were capable of discriminating between the individual cats and also between the individual dogs presented in Experiment 4.

#### Discussion of Experiments 4 and 5

In Experiment 4, we found that when 3- to 4-month-olds were familiarized with the cat stimuli in Figure 3, they showed a subsequent novel category preference for novel dogs when tested with a novel cat and a novel dog. Similarly, when 3- to 4-month-olds were familiarized with the dog stimuli in Figure 3, they also displayed a subsequent novel category preference for novel cats when tested with a novel cat and a novel dog. In addition, Experiment 5 revealed that the infants could discriminate between individual exemplars of cats and also between the individual instances of dogs.

Taken together, the results of Experiment 4 and 5 confirm the model predictions made in Simulation 3. When familiarized with stimuli in which neither the dog nor the cat feature value distributions subsumed one another, infants formed perceptual categories that are fully exclusive. That is, the *cat* category excluded novel dogs and the *dog* category excluded novel cats. This is true even if the variances of the two categories are unequal. These results support the argument made by Mareschal et al. (2000) that infants are generating the *cat* and *dog* categories online and in

Table 4  
Mean Fixation Times (in Seconds) During the Familiarization Trial and Mean Novelty Preference Scores (in Percentages) During the Preference Test Trials in Experiment 5

Experimental group	Fixation time		Novelty preference		$t^a$	$\eta^2$
	$M$	$SD$	$M$	$SD$		
Cat	8.35	2.73	60.32	14.85	2.41**	.33
Dog	9.14	3.01	62.35	16.99	2.52**	.35

<sup>a</sup>The  $t$  tests compared the mean scores with chance performance.  
\*\*  $p < .025$ , one-tailed.

response to the particular exemplars encountered during familiarization. Changing the distribution of features in the familiarization exemplars changes the pattern of novel category preferences observed in the infants.

### General Discussion

Quinn et al. (1993) observed a striking asymmetry in the categorization of cat and dog images by 3- and 4-month-old infants. An initial simulation by Mareschal and French (1997) and Mareschal et al. (2000) was able to reproduce the original experimental results by focusing on the inclusion relationship of the two categories of animals. This simulation led to a prediction: Namely, that if the inclusion relationship of shared features (i.e., images of cats being subsumed under a broader distribution of images of dogs) was the key to explaining this categorization asymmetry, then artificially reversing the nature of the inclusion relationship for shared features should reverse the infant categorization asymmetry (French et al., 2001). In Simulation 1 and Experiment 1, we were able to reverse the inclusion relationship and, as predicted by the model, we observed a reversal of the categorization asymmetry in the infants.

The model also predicted in Simulation 3 that if the inclusion relationship of the statistical distributions of the feature values for exemplars from the two categories was removed by a judicious choice of familiarization exemplars of cats and dogs, even if the original difference in overall variance between the two categories was maintained, then there would no longer be a categorization asymmetry. This prediction was confirmed experimentally in Experiment 4.

The asymmetric exclusivity of *cat* and *dog* categories is not limited to photographic stimuli nor to 3- to 4-month-olds. Mareschal, Powell, and Volein (2004) found that asymmetries persisted even in infants as old as 9 months when tested with an object examination task. This method is similar to a visual habituation method except that it involves presenting the infants with three-dimensional toy replicas of cats and dogs rather than two-dimensional photographs. Thus, we believe that the processes that underlie category learning in this article reflect general-purpose processes that operate on a range of stimuli whenever category retrieval is not possible. As such, this work has a number of implications that cut across a broad range of levels.

First, it provides an example of the productive interaction between computational modeling and empirical research. The somewhat surprising data on categorization asymmetry in young infants reported by Quinn et al. (1993) led to the development of the autoencoder model by Mareschal and French (1997) and Mareschal et al. (2000). This model was initially able to demonstrate how a simple statistical (i.e., bottom-up) online mechanism sensitive to perceptual distribution information could provide an explanation for the original categorization-asymmetry data. The model was then used to make a number of specific and falsifiable predictions about how this asymmetry could be experimentally manipulated in infants. These predictions were then tested and confirmed empirically in the presently reported findings. This illustrates how a deeper understanding of behavioral phenomena can be achieved by shifting the emphasis of research away from asking what participants can do to asking how they do it. Imple-

mented computational models are the appropriate research tools for addressing this latter type of question.

Second, the work has implications for the understanding of early infant categorization. In particular, the research presented in this article argues for the online bottom-up nature of category acquisition in young infants. The fact that we were able to predict (and manipulate) categorization asymmetries solely by manipulating the inclusion relationship of perceptual features between the two categories supports this conclusion. In addition, the mechanism of category acquisition proposed here relies only on processing the contents of an associative short-term memory consisting of the statistical distributions of immediate perceptual information about the stimuli. Significantly, this mechanism does not require any contribution from prior knowledge stored in some kind of long-term memory. That is, what we are witnessing in the network—and we believe in the infants—is online perceptually driven learning that does not use prior knowledge. Although we do not claim that this is the sole mechanism of young infant category acquisition (see, e.g., the work of Quinn & Eimas, 1998, on how infants represent humans), we do believe that it predominates in early infancy for these nonhuman animal images of cats and dogs and presumably for other generic object categories (e.g., chairs, cars; Behl-Chadha, 1996). Thus, we argue that behavioral exclusivity asymmetries that can be shown to depend on the distribution characteristics of the images encountered during familiarization are markers for bottom-up processing (see also Mareschal et al., 2004).

Indeed, similar sensitivity to distribution information is found with adults learning artificial categories (e.g., Fried & Holyoak, 1984; Posner & Keele, 1970; Reed, 1972). However, the apparent failure to retrieve appropriate category representations and to apply them in a top-down fashion does not necessarily mean that the infants do not possess the appropriate category representations. One possibility is that the infants lack appropriate conceptual or semantic representations of cats and dogs in their long-term memory. However, a second possibility is that the infants do possess the appropriate category representations but that they are failing to retrieve that information in this task.

Of course, at some point during later infancy or early childhood, category learning is no longer a wholly online process dependent almost exclusively on the statistical distribution of perceptual feature information in the environment (except, of course, for artificial categories specifically devised to avoid top-down processing in adults; Ashby & Ell, 2001). How the emergence of prior knowledge interacts with online bottom-up driven category learning is by no means a straightforward question (Murphy, 2002). It seems unlikely that as soon as infants possess relevant conceptual-category information, this information overrides the effects of distribution information in the familiarization information. Indeed, this is not the case even in adults. Prior knowledge can limit attention to statistical information about a category, but, in some cases, it even appears to increase sensitivity to statistical information (Kaplan & Murphy, 1999; Spalding & Murphy, 1999), perhaps through a process of knowledge-based focused attention to diagnostic features (Heit, 1998). One of the challenges that lies ahead is to understand how the kind of computational model described in the present article must be modified for it to integrate not only statistical information from the environment but also prior knowledge stored in some kind of long-term memory. A number of

small steps have been taken in this direction (Mermillod, French, Quinn, & Mareschal, 2003), but the bulk of this research still remains to be done.

This work also bears on the study of conceptual development in general. The transition from purely perceptually based category representations to knowledge-rich category structures has been the topic of much debate in the literature on infant categorization (e.g., Carey, 2000; Madole & Oakes, 1999; Mandler, 2000; Quinn & Eimas, 1997, 2000; Rakison & Poulin-Dubois, 2001). Our view is that initial concept acquisition in young infants is empirically driven (e.g., Eimas & Quinn, 1994; Madole & Oakes, 1999; Quinn & Eimas, 2000; Rakison, 2003). In this respect, we agree with Jean Mandler (e.g., Mandler, 2000) and disagree with Susan Carey (Carey, 2000). However, we differ from both Mandler and Carey with regard to how the transition from early to later conceptual development occurs. We believe that initially infants form category representations on the basis of early perceptual experience, whereas later development is devoted to enrichment of (i.e., incorporation of nonobvious information into) these early perceptual representations. For recent detailed discussions of the different positions on conceptual development, see Rakison and Oakes (2003).

The shifting reliance from bottom-up to top-down processing that occurs during infancy may also relate to other research that provides evidence for greater perceptual sensitivity in infancy than in adulthood. Our results suggest that 3- to 4-month-olds are more responsive to subtle feature changes in the familiarization stimuli than adults are. Similar behavioral effects have been found with infant face processing (Pascalis, de Haan, & Nelson, 2002). Here, 6-month-olds were found to be sensitive to the subtle differences between different monkey and different human faces; however, both 10-month-olds and adults were no longer able to discriminate between the individual monkey faces while maintaining their accuracy at discriminating human faces. Taken together, the two lines of work suggesting the following: Infants may begin life with perceptual sensitivities that allow them to parse environmental experience into various classes of stimulation. Without continuing experience, at least some level of these sensitivities may be lost and may require extensive reexperience for recovery (Pisoni, Aslin, Perey, & Hennessy, 1982). With continuing experience, such perceptual sensitivities may eventually transition and give rise to conceptual sensitivities.

To conclude, adult category learning has both conceptual and perceptual components, but these are notoriously hard to tease apart. The value of studying perceptual mechanisms in infants is that it gives one a window on the perceptual mechanisms of categorization, largely, if perhaps not completely, undistorted by the later contribution of conceptual knowledge. Our use of realistic images of natural stimuli bears directly on adult categorization. We know that for the categories we have chosen (i.e., *dogs* and *cats*), conceptual knowledge will later play a significant role as these categories develop into mature adult categories; it is much less certain that this would be the case for artificial stimuli, such as dot-pattern stimuli or geometric shapes. In this way, when conceptual knowledge does begin to play a role in categorization, researchers will be in a better position to understand how it contributes to the overall categorization process, having previously acquired an understanding of the purely perceptual mechanisms that may be operative with the onset of environmental experience.

## References

- Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442–481.
- Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences*, *5*, 204–210.
- Ashby, F. G., & Waldron, E. M. (1999). On the nature of implicit categorization. *Psychonomic Bulletin & Review*, *6*, 363–378.
- Behl-Chadha, G. (1996). Basic-level and superordinate-like categorical representations in early infancy. *Cognition*, *60*, 105–141.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.
- Bomba, P. C., & Siqueland, E. R. (1983). The nature and structure of infant form categories. *Journal of Experimental Child Psychology*, *35*, 609–636.
- Carey, S. (2000). The origin of concepts. *Journal of Cognition and Development*, *1*, 37–41.
- DeLoache, J. S., Strauss, M., & Maynard, J. (1979). Picture perception in infancy. *Infant Behavior and Development*, *2*, 77–89.
- Eimas, P. D., & Quinn, P. C. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development*, *65*, 903–917.
- Fagan, J. F. (1970). Memory in the infant. *Journal of Experimental Child Psychology*, *9*, 217–226.
- Fahlman, S. E. (1988). Faster-learning variations on back-propagation: An empirical study. In D. S. Touretsky, G. E. Hinton, & T. J. Sejnowski (Eds.), *Proceedings of the 1988 Connectionist Summer School* (pp. 38–51). Los Altos, CA: Morgan Kaufman.
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*, 15822–15826.
- French, R. (1995). *The subtlety of sameness: A theory and computer model of analogy-making*. Cambridge, MA: MIT Press.
- French, R. M., Mermillod, M., Quinn, P. C., & Mareschal, D. (2001). Reversing category exclusivities in infant perceptual categorization: Simulations and data. In K. Stenning & J. Moore (Eds.), *Proceedings of the 23rd annual conference of the Cognitive Science Society* (pp. 307–312). Mahwah, NJ: Erlbaum.
- 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*, 234–255.
- Goldstone, R. L., & Barsalou, L. W. (1998). Reuniting perception and conception. *Cognition*, *65*, 231–262.
- Heit, E. (1998). Influences of prior knowledge on selective weighting of category members. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 712–731.
- Japkowicz, N. (2001). Supervised and unsupervised binary learning by feedforward neural networks. *Machine Learning*, *42*, 97–122.
- Kaplan, A. S., & Murphy, G. L. (1999). The acquisition of category structure in unsupervised learning. *Memory & Cognition*, *27*, 699–712.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*, B35–B42.
- Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *58*, 49–70.
- Madole, K., & Oakes, L. (1999). Making sense of infant categorization: Stable processes and changing representations. *Developmental Review*, *19*, 263–296.
- Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. *Journal of Cognition and Development*, *1*, 3–36.
- Mareschal, D., & French, R. M. (1997). A connectionist account of interference effects in early infant memory and categorization. In M. G.

- Shafto & P. Langley (Eds.), *Proceedings of the 19th annual conference of the Cognitive Science Society* (pp. 484–489). Mahwah, NJ: Erlbaum.
- Mareschal, D., & French, R. M. (2000). Mechanisms of categorization in infancy. *Infancy, 1*, 59–76.
- Mareschal, D., French, R. M., & Quinn, P. C. (2000). A connectionist account of asymmetric category learning in early infancy. *Developmental Psychology, 36*, 635–645.
- Mareschal, D., Powell, D., & Volein, A. (2004). Basic-level category discrimination by 7- and 9-month-olds in an object examination task. *Journal of Experimental Child Psychology, 86*, 87–107.
- Mareschal, D., Quinn, P. C., & French, R. M. (2002). Asymmetric interference in 3- to 4-month-olds' sequential category learning. *Cognitive Science, 26*, 377–389.
- Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London: Series B. Biological Sciences, 200*, 269–294.
- Mermillod, M., French, R. M., Quinn, P. C., & Mareschal, D. (2003). The importance of long-term memory in infant perceptual categorization. *Proceedings of the 25th annual conference of the Cognitive Science Society* (pp. 804–809). Mahwah, NJ: Erlbaum.
- Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.
- Murphy, G. L., & Kaplan, A. S. (2000). Feature distribution and background knowledge in category learning. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 53A*, 962–982.
- Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115*, 39–57.
- Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 3–27.
- Pascalis, O., de Haan, M., & Nelson, C. A. (2002, May 17). Is face processing species-specific during the first year of life? *Science, 296*, 1321–1323.
- Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 297–314.
- Posner, M. I., & Keele, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology, 83*, 304–308.
- Quinn, P. C. (1987). The categorical representation of visual pattern information by young infants. *Cognition, 27*, 145–179.
- Quinn, P. C., & Eimas, P. D. (1996). Perceptual organization and categorization in young infants. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 10, pp. 1–36). Norwood, NJ: Ablex.
- Quinn, P. C., & Eimas, P. D. (1997). A reexamination of the perceptual-to-conceptual shift in mental representations. *Review of General Psychology, 1*, 271–287.
- Quinn, P. C., & Eimas, P. D. (1998). Evidence for a global categorical representation of humans by young infants. *Journal of Experimental Child Psychology, 69*, 151–174.
- Quinn, P. C., & Eimas, P. D. (2000). The emergence of category representations during infancy: Are separate perceptual and conceptual processes required? *Journal of Cognition and Development, 1*, 55–61.
- Quinn, P. C., Eimas, P. D., & Rosenkrantz, S. L. (1993). Evidence for representations of perceptually similar natural categories by 3- and 4-month-old infants. *Perception, 22*, 463–475.
- Rakison, D. H. (2003). Free association? Why category development requires something more. *Developmental Science, 6*, 20–22.
- Rakison, D., & Oakes, L. M. (Eds.). (2003). *Early category and concept development: Making sense of the blooming, buzzing confusion*. New York: Oxford University Press.
- Rakison, D., & Poulin-Dubois, D. (2001). The developmental origin of the animate–inanimate distinction. *Psychological Bulletin, 127*, 209–228.
- Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology, 3*, 382–407.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996, December 13). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.
- Schuler, E. M. (Ed.). (1980). *Simon & Schuster's guide to dogs*. New York: Simon & Schuster.
- Schyns, P. G., Goldstone, R. L., & Thibaut, J. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21*, 1–54.
- Siegal, M. (Ed.). (1983). *Simon & Schuster's guide to cats*. New York: Simon & Schuster.
- Sokolov, E. N. (1963). *Perception and the conditioned reflex*. Hillsdale, NJ: Erlbaum.
- Spalding, T. L., & Murphy, G. L. (1999). What is learned in knowledge-related categories? Evidence from typicality and feature frequency judgments. *Memory & Cognition, 27*, 856–867.
- Spencer, J., Quinn, P. C., Johnson, M. H., & Karmiloff-Smith, A. (1997). Heads you win, tails you lose: Evidence for young infants categorizing mammals by head and facial attributes. *Early Development and Parenting, 6*, 113–126.
- Strauss, M. S. (1979). Abstraction of prototypical information by adults and 10-month-old infants. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 618–632.
- Younger, B. A. (1985). The segregation of items into categories by ten-month-old infants. *Child Development, 56*, 1574–1583.
- Younger, B. A., & Gotlieb, S. (1988). Development of categorization skills: Changes in the nature or structure of infant form categories? *Developmental Psychology, 24*, 611–619.
- Zhu, S. C., & Yuille, A. L. (1996). FORMS: A flexible object recognition and modeling system. *International Journal of Computer Vision, 20*, 187–212.

Received April 30, 2003

Revision received March 4, 2004

Accepted March 17, 2004 ■

## E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://watson.apa.org/notify/> and you will be notified by e-mail when issues of interest to you become available!