

# Reversing Category Exclusivities in Infant Perceptual Categorization: Simulations and Data

**Robert M. French, Martial Mermilliod (rfrench, mmermillod@ulg.ac.be)**

Psychology Department, Université de Liège, Belgium

**Paul C. Quinn (pquinn@washjeff.edu)**

Psychology Department, Washington & Jefferson College, Washington, PA 15301 USA

**Denis Mareschal (d.mareschal@bbk.ac.uk)**

Centre for Brain & Cognitive Development, School of Psychology, Birkbeck College, London, U.K

## Abstract

Three- to four-month-old infants presented with a series of cat or dog photographs show an unusual asymmetry in the exclusivity of the perceptual category representations formed. We have previously accounted for this asymmetry in terms of an inclusion asymmetry in the distribution of features present in the cat and dog images used during familiarization (Mareschal, French, & Quinn, 2000). We use a combination of connectionist modeling and experimental testing of infants to show that the asymmetry can be reversed by an appropriate pre-selection and minor image modification of cat and dog exemplars used for familiarization. The reversal of the asymmetry adds weight to the feature distribution explanation put forward by Mareschal et al. (2000).

## Introduction

The ability to categorize is, without question, one of the central pillars of cognition. It is, therefore, not surprising that categorization abilities are present in humans from the very earliest age. Indeed, infants only a few months old are able to separate complex visual stimuli into generic object categories (e.g., Quinn & Eimas, 1996). In previous work, we have presented a simple connectionist model of perceptual categorization during early infancy (Mareschal & French, 1997; Mareschal & French, 2000; Mareschal et al., 2000). The model provided a mechanistic account of early infant category learning in terms of the data compression properties of connectionist autoencoder networks. Not only did this model capture standard infant categorization phenomena such as prototype formation and the use of feature co-variation information to form categories (Mareschal & French, 2000), but it also captured some of the more subtle idiosyncratic characteristics of infants' categorization behavior.

In particular, 3- to 4-month-olds show an unexpected asymmetry in the exclusivity of the perceptual category representations formed for cats versus dogs (Quinn, Eimas, & Rosenkrantz, 1993; Eimas, Quinn, & Cowan, 1994). Following exposure to a series of cat photographs, these infants will form a perceptual

representation for cats that excludes dogs. In contrast, following exposure to a series of dog photographs, the same infants will form a category representation for dogs that does NOT exclude cats. Thus, there is an asymmetry in the exclusivity of the cat and dog categories: dogs are excluded from the representation for cats, whereas dog do not exclude cats.

We extend these results by showing how an opposite exclusivity asymmetry can be induced in 3- to 4-month-olds by a judicious choice of cat and dog exemplars presented to the infants prior to testing. Success in reversing the asymmetry between the Cat and Dog categories would lend strong support to a bottom-up account of early infant perceptual categorization.

## Asymmetric exclusivity in infant categorization

Quinn et al. (1993) reported the following surprising categorization asymmetry. When 3- to 4-month-old infants were shown different photographs of either cats or dogs, they formed perceptual category representations for either groups of pictures. Infants were first shown a number of different photographs of cats and were then shown a picture of a dog paired with a novel picture of a cat. During the preference trials, the infants were much more attentive to the dog than to the novel cat. This was interpreted as showing that the infants had formed a category representation of Cat that excludes dogs. The dog, in other words, was perceived by the infants as not belonging to the category of cats. In sharp contrast, infants who were first shown a series of photographs of different dogs and were then shown a picture of a cat along with a novel dog were not preferentially attentive to either picture. When coupled with the finding that infants did not show a prior preference for cat photographs over dog photographs, and that infants familiarized with either cats or dogs looked longer at a bird photograph, the overall pattern of results was interpreted as showing that infants had formed a category representation of Dog that did not exclude cats. In short, the Dog category included cats, but the Cat category did not include dogs.

Infant perceptual categorization tasks frequently rely on preferential looking techniques based on the finding that infants direct attention more to unfamiliar or unexpected stimuli (e.g., Sokolov, 1963; Charlesworth, 1969; Cohen, 1973). While it is true that infants may sometimes have a preference for familiar stimuli, such as word stress patterns (Jusczyk, Cutler, & Redanz, 1993), it has been repeatedly shown that there is preferential attention directed to *novel* visual stimuli. The standard interpretation of this behavior is that the infants are comparing the input stimuli to an internal representation of those stimuli. As long as there is a discrepancy between the information stored in the internal representation and the visual input the infant continues to attend to the stimuli. While attending to the stimuli the infant updates its internal representation. When the information in the internal representation is no longer discrepant with the visual input, attention is switched elsewhere. When a familiar object is presented there is little or no attending because the infant already has a reliable internal representation of that object. In contrast, when an unfamiliar or unexpected object is presented, there is a lot of attending because an internal representation has to be constructed or adjusted.

When a series of exemplars can be grouped into a consistent category, this account of representation construction implies a progressive decrease in looking time with successive exemplars encountered. Although each exemplar encountered is novel (and therefore attracts the infant's attention), the process of representation construction gradually leads to the extraction of key dimensions of the category. Thus, after some time, a reliable category representation is constructed and new exemplars encountered (although still novel), take little time to be assimilated to the internal category representation and therefore only briefly capture the infant's attention.

### A model of infant perceptual categorization

We used a three-layer autoencoder to model infant categorization behaviors (Mareschal & French, 1997; Mareschal & French, 2000; Mareschal et al., 2000). Learning in an autoencoder consists of developing an internal representation of the input (at the hidden unit level) that is sufficiently reliable to reproduce all the information in the original input (Cottrell, Munro, & Zipser, 1988). Information is first compressed into an internal representation and then expanded to reproduce the original input. The successive cycles of training in the autoencoder are an iterative process by which a reliable internal representation of the input is developed. The reliability of the internal representation is tested by expanding it and comparing the resulting predictions to the actual stimulus being encoded.

The degree to which network error increases on presentation of a novel object depends on the similarity

between the novel object and previously seen (i.e., familiar) objects. Presenting a series of similar objects leads to a progressive drop in error on future similar objects. The modeling assumption that we have therefore made is that network error and infant attention levels correlate: the higher the network error, the longer the looking time of the infant. This is true of both autoassociators (where output error is the measurable quantity) and infants (where looking time is the measurable quantity).

To model the Cat/Dog findings, we obtained data for the networks from measurements of the original cat and dog pictures used by Quinn et al. (1993). There were 18 dogs and 18 cats classified according to the following ten traits: head length, head width, eye separation, ear separation, ear length, nose length, nose width, leg length vertical extent, and horizontal extent. The feature values were normalized over all pictures in both training sets to be within 0 and 1.

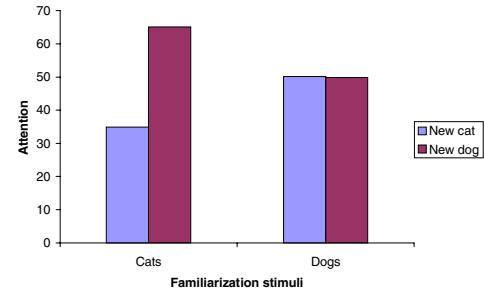


Figure 1. Generalization errors for networks trained on cats and dogs (Mareschal et al. 2000).

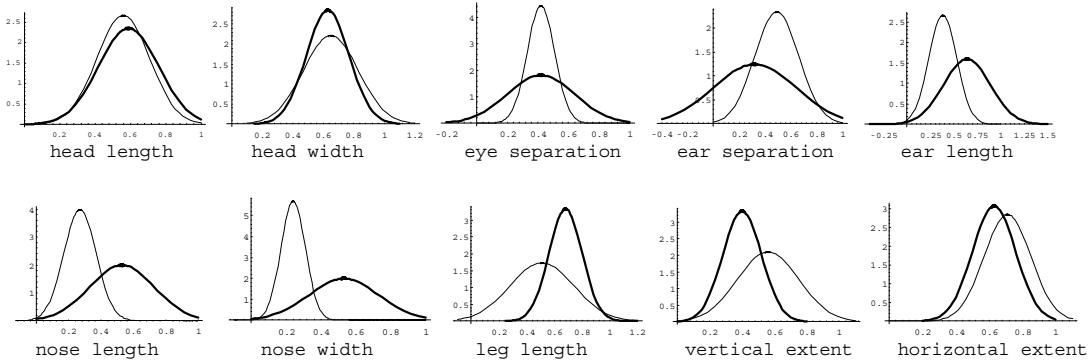
Figure 1 shows what happens when networks trained on cats are presented with a novel cat and a dog, and when networks trained on dogs are tested with a novel dog and a cat. When the networks are initially trained (i.e., familiarized) on cats, the presentation of a dog results in a large error score (corresponding to infants' longer looking time). Dogs are not included in the category representation of cats. In contrast, when the networks are initially trained on dogs, the presentation of a cat will result in essentially the same error as a novel dog, suggesting that the cats have been included in the category representation for dogs.

Because autoencoders extract the distribution statistics of the exemplars they have encountered, this led us to explore the distribution of feature values in the data measured from the original photographs in order to explain the asymmetry. Figure 2 shows the probability distributions of the 10 traits for both cats and dogs. Some of the traits are very similar in terms of their means and distribution of both cats and dogs (e.g., head length and head width). Others, especially nose length and nose width, are very different and will provide the crucial explanation of the unexpected attentional asymmetries reported by Quinn et al. (1993) and Eimas

et al. (1994). It is clear that in almost all cases the distribution for each Dog trait (represented by the dark line) subsumes the distribution for corresponding trait for cats. The narrower distributions for most Cat traits, on the other hand, do not subsume the range of values for the corresponding Dog traits. In other words, cats are possible dogs but the reverse is not the case: most dogs are not possible cats. Specifically, when we examine all of the members of the two populations, we see that the values of all ten traits for 9 (i.e., 50%) of the members of the Cat category fall within a  $2\sigma$  cut-off

for those traits for the Dog category. Fully half of the cats in the population could be reasonably classified as dogs. In contrast, the smaller means and lower variances of a number of traits (especially, nose length and nose width) for cats compared to dogs means that only 2 of the 18 dogs could conceivably be classified as being members of the Cat category.

Hence, it seems that the exclusivity asymmetry is driven by (1) an associative learning mechanism that is sensitive to feature distributions, and (2) a distribution



*Figure 2.* Frequency distributions for the ten defining traits of 18 dogs and 18 cats in Mareschal et al., 2000. The variance of Dog traits is, on average, 1.6 times that of Cats. Dogs' features largely subsume cats'.

profile in which the Dog feature values largely subsume the Cat feature values. A direct implication of this is that if the distribution statistics were reversed, then we should observe a reversed categorization asymmetry. In this new case, infants should develop a perceptual category representation of Dog that excludes cats and a perceptual category representation of Cat that does not exclude dogs. The simulation and experiments reported below test this prediction directly.

### Reversing asymmetric exclusivity

To explore whether the asymmetry could be experimentally reversed, we began by artificially manipulating the naturally occurring variance of the two categories. In the original experiment the within-category variability of the dog photographs was greater than that of the cat photographs and, crucially, the feature set for dogs largely subsumed that of cats. However, by carefully selecting sets of cat and dog photographs and then morphing a number of these images, we were able to reverse the variance of the categories. In the original experiment (Mareschal et al., 2000) the average variance over all ten features of the Dogs was 1.63 times that of the Cats, whereas for the modified images the average variance of Cats was 3.12 times that of Dogs. Figure 3 shows the feature distributions for these modified exemplar sets. A comparison with the original data plotted in Fig. 2 shows how the distributions have been reversed. This is

especially clear for features such as "Leg length." There were an identical number of morphed images (8 out of 18) in both the Dog and Cat stimuli sets.

### Reversing the network's learning

The simulation reported was done on a standard 10-8-10 feedforward backpropagation autoencoder network (learning rate: 0.1, momentum: 0.9, Fahlman offset: 0.1). Training was identical to that in Mareschal et al. (2000). Networks were trained in batches of 2 patterns for a maximum of 250 epochs. This simulated familiarization with pairs of pictures for a fixed period before being presented with a new familiarization pair. Results were averaged over 50 runs.

Figure 4 shows the model's generalization error to novel exemplars of cats and dogs as a function of whether they were trained on cats or on dogs. Networks trained with cats show no difference in error (hence predict no difference in looking times) when tested with a novel cat or a dog. In contrast, the networks originally trained with dogs show much greater error when tested with a novel cat than when tested with a novel dog (suggesting a strong preference for looking at a cat vs. a novel dog). This asymmetry is exactly the opposite of the one found in the original Mareschal et al. (2000) study and constitutes an explicit prediction of the autoencoder model.

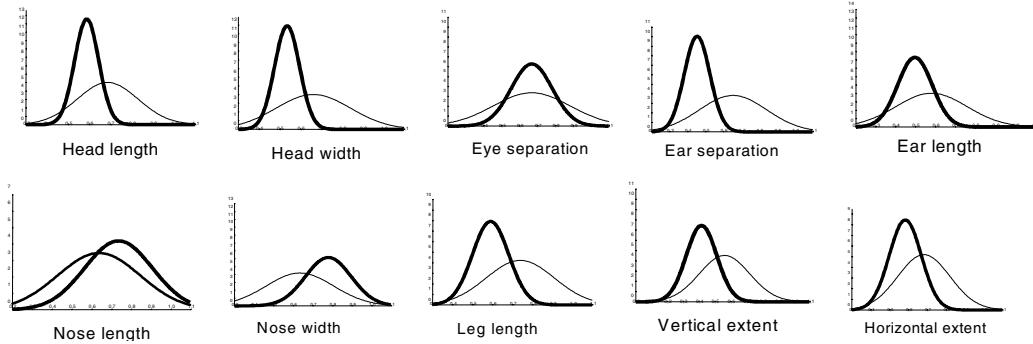


Figure 3. Frequency distributions for Exp. 1 for the 18 dogs and 18 cats. The variance has been artificially reversed and, crucially, Dog features are largely subsumed by Cat features. Compare with Figure 2

### Experiment 1: Reversal of exclusivity

This prediction was tested with two groups of 3- and 4-month-old infants that were presented with a set of 12 exemplars from the same category, cats or dogs, during a series of familiarization trials, and were then presented with preference test trials consisting of a novel cat paired with a novel dog. The model predicts that infants familiarized with dogs should display a novel category preference for cats on the preference test trials, whereas infants familiarized with cats should display looking times divided evenly between the dogs and cats displayed on the preference test trials.

### Method

**Participants.** The participants in Exp. 1 were 12 infants approximating 3 and 4 months of age ( $M = 3$  months, 20 days;  $SD = 8.30$  days). Seven of the infants were females and five were males.

**Stimuli.** The stimuli were 36 colored photographs of cats and dogs (18 exemplars per category, representing 18 different breeds for each category). The photographs were obtained from Siegal (1983) and Schuler (1980). In order to obtain dogs with low perceptual variance and cats with high variance, certain stimuli were slightly modified using computer imaging processing software (Rubber v.2.0). None of the stimuli were “morphed” to the point of giving the impression of a strange animal. The same number of animals (8) were morphed in both groups. The variance of the Cat category was modified so that the average variance of Cats was 3.1 times that of Dogs (compared to the original experiment where the average variance of Dogs was 1.6 times that of Cats). As in Quinn et al. (1993) and Mareschal et al. (2000), the pictures selected were chosen to represent a variety of shapes, colors, and orientations of each type of animal. The size of the animal in each picture was nearly the same, and thus not a reflection of its actual size (so that any categorization effects observed would not be the result of simple size discrimination.)

Each stimulus contained a single animal, cut away

from its background, centered, and mounted onto a white 17.7 x 17.7 cm posterboard for presentation.

**Apparatus.** Infants 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 (85 cm high and 29 cm wide) onto which were attached two compartments to hold the posterboard stimuli. The stimuli were illuminated by a 60-Hz 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's face. There was a 0.625 cm peephole midway between the two display compartments allowing observation of the infant's visual fixations.

**Procedure.** In both experiments, infants were placed in a reclining position on their seated parent's lap. An experimenter positioned the apparatus such that the midline of the infant's head was aligned with the midline of the display panel. The experimenter loaded the appropriate stimuli into the display panel, elicited the infant's attention and exposed the stimuli to the infant. During the course of a trial, the experimenter observed the infant through the peephole and recorded visual fixations to the left and right stimuli by means of two 605 XE Accusplit electronic stop watches, one of which was held in each hand. Interobserver reliability for this procedure was determined by comparing the looking times measured by the experimenter using the center peephole and additional observers using peepholes to the left of the left stimulus compartment and to the right of the right stimulus compartment is high (Pearson  $r = .97$ ), a value comparable to values obtained in other laboratories that measured visual fixation duration with the corneal reflection procedure (e.g., O'Neil, Jacobson, & Jacobson, 1994). Two experimenters recorded fixations, one during familiarization and another during preference test trials. Importantly, the person recording during preference test trials was unaware of the category information that was presented during the familiarization trials.

Each infant was assigned twelve randomly selected

pictures of cats or dogs. On each of six 15s familiarization trials, two of the twelve stimuli, again randomly selected, were presented. Six infants were randomly assigned to each group, defined by the familiarization category, cats or dogs. Immediately after the familiarization trials, two 10s preference test trials were administered in which a novel cat was presented with a novel dog. There were six such pairs, randomly selected, and each pair, which was seen on both test trials, was assigned to one infant who had seen dogs and one infant who had seen cats during the familiarization trials. The test-trial stimuli were thus identical for both groups of infants. The left-right positioning of the novel animal from the novel category was appropriately counterbalanced across infants.

	Familiarization phase (average fixation time in secs.)		Novelty preference (% of viewing time for unfamiliar category)	t
Familiarization category	First 3 trials	Last 3 trials		
Cats	7.8(3.8)	6.9(3.6)	49.5% (16.7)	-0.08
Dogs	7.9(1.4)	9.2(3.1)	70.4% (9.7)	5.1*

t for mean vs. chance      \*p < .005, one-tailed.

Table 1. Mean fixation times 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 1. Novelty preference is expressed in percentage of time that the infant looks at the exemplar from the unfamiliar category compared to the total time regarding the pair of test stimuli. An analysis of variance, familiar category (cats vs. dogs) x trial block (1-3 vs. 4-6), performed on the individual scores, revealed no significant effects,  $F(1, 10) < 2.28$ ,  $p > .15$ , in each instance. As has been the case in previous infant categorization studies using the same procedures and similar stimuli (Eimas & Quinn, 1994; Eimas, et al. 1994; Mareschal et al., 2000; Quinn & Eimas, 1996, 1998; Quinn, et al., 1993), no evidence of habituation was obtained. We believe the complexity and variety of stimuli were sufficient to maintain infant attention during the familiarization trials.

### 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. Mean novel category preference scores are shown in Table 1 and in Figure 5. t tests versus chance, which were one-tailed because of

the predicted preference in the direction of novelty, revealed that infants familiarized with dogs preferred the novel cat, but infants familiarized with cats did not prefer the novel dog. In addition, the two means were significantly different from each other,  $t(10) = 2.65$ ,  $p < .05$ , two-tailed. As was predicted by the model, infants familiarized with dogs formed a category representation that excluded cats, but infants familiarized with cats did not form a category representation that excluded dogs. The findings are exactly the opposite of those reported in Exp. 4 of Quinn et al. (1993). Thus, we can reasonably conclude that the stimulus manipulations were successful in reversing the inclusion relation between dogs and cats reported by Mareschal et al. (2000).

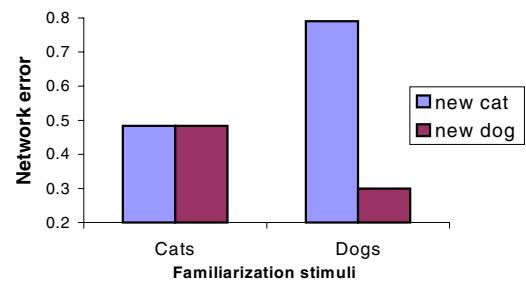


Figure 4. Network generalization errors when Cat features largely subsume Dog features

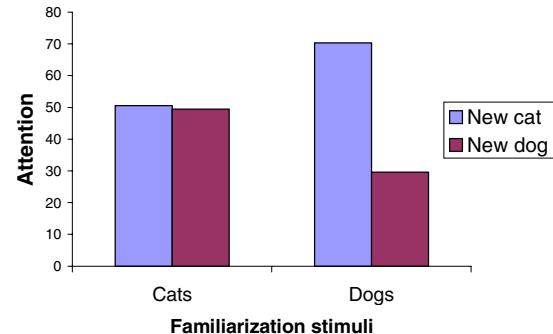


Figure 5. Infant attention when Cat feature distributions largely subsume those for Dogs.

## Experiment 2: No prior preference

An alternative explanation for the outcome of Exp. 1 is that infants might have spontaneously preferred the cats over the dogs. Although no such spontaneous preference was found in Quinn et al. (1993) or Quinn and Eimas (1996), it is possible that the stimulus manipulations could have inadvertently produced one. If there was a preference for cats over dogs in Exp. 1, then it would have facilitated (if not fully explained) any presumed novel category preference for cats after familiarization with dogs, and it would have interfered with any novel category preference for dogs after familiarization with cats. Exp. 2 was thus replication of Exp. 1, but conducted without the familiarization trials.

## Method

Participants. 6 infants approximately 3 and 4 months old ( $M = 3$  months, 17 days;  $SD = 12.71$  days). Three of the infants were females and three were males.

Stimuli and apparatus. Same as Exp. 1.

Procedure. All infants were presented with the preference test trials described for Exp. 1, but without the prior familiarization trials.

## Results and Discussion

A preference score for cats was determined for each infant for the two trials by dividing the looking time that the cat was observed by the total looking time to both the cat and dog. The score was then converted into a percentage. The mean preference for cats was 48.34%,  $SD = 22.03$ . This preference was not significantly different from chance,  $t(5) = -0.18$ ,  $p > .20$ , two-tailed. Further, the preference for cats after familiarization with dogs in Exp. 1 was found to be reliably higher than the spontaneous preference for cats with no familiarization with dogs,  $t(10) = 2.24$ ,  $p < .05$ , two-tailed. The preference results from Exp. 1 are thus unlikely to be reflective of a spontaneous preference for cats and more likely are a consequence of the reversal of the inclusion relation between cats and dogs.

## Conclusion

Quinn, Eimas, & Rosenkrantz (1993) observed a striking asymmetry in the infant categorization of photos of cats and dogs. 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 *within-category variability and inclusion relation* of the two categories of animals. This simulation led to a prediction — namely, that if the degree of variability and overlap of shared feature distributions was the key to explaining this categorization asymmetry, then artificially reversing the order of the within-category variability for shared features should reverse the infant categorization asymmetry. We were able to reverse this categorical variability and, as predicted by the model, we observed the reverse categorization asymmetry in the infants.

The reversal of the asymmetry makes the point that infants form at least some category representations on-line, rather than tapping into pre-existing concepts that had been formed prior to arriving at the laboratory. If the infants been relying on previously acquired categories, then their responsiveness should not have varied with the variations in the familiar category information presented. The fact that infant responsiveness did vary across experiments suggests that the categories were being formed on-line and that the boundaries can be pushed around depending on the information presented during familiarization.

## Acknowledgments

This work was supported by Grants BCS-0096300 from the NSF to P. Quinn, RS000239112 from the ESRC (UK) to D. Mareschal and HPRN-CT-2000-00065 from the Eur. Commission to R. French. Thanks to C. Laboue for his comments on this work and to L. Yarzab for her assistance with the experiments.

## References

Charlesworth, W. (1969). The role of surprise in cognitive development. In D. Elkind & J. Flavell (Eds.), *Studies in cognitive development. Essays in honor of Jean Piaget*, pp. 257-314, Oxford University Press.

Cohen, L. (1973). A two-process model of infant visual attention. *Merrill-Palmer Quarterly*, 19, 157-180.

Cottrell, G., Munro, P., & Zipser, D. (1988). Image compression by backpropagation: and example of extensional programming. In N. E. Sharkey (Ed.), *Advances in cognitive science*, Vol. 3. Ablex.

Eimas, P., & Quinn, P. (1994). Studies on the formation of perceptually based basic-level categories in young infants. *Child Development*, 65, 903-917.

Eimas, P., Quinn, P., & Cowan, P. (1994). Development of exclusivity in perceptually-based categories of young infants. *J. of Exp. Child Psychology*, 58, 418-431.

Fagan, J. (1970). Memory in the infant. *J. of Experimental Child Psychology*, 9, 217-226.

Jusczyk, P., Cutler, A., & Redanz, N. (1993). Infants' preference for the predominant stress patterns of English verbs. *Child Development*, 64, 675-687.

Mareschal, D. & French, R. (2000). Mechanism of categorization in infancy. *Infancy*, 1, 59-76.

Mareschal, D. & French, R. (1997). A connectionist account of interference effects in early infant memory and categorization. *Proceedings of the 19th Annual Cognitive Science Society Conference*, LEA, 484-489.

Mareschal, D., French, R., & Quinn, P. (2000). A connectionist account of asymmetric category learning in early infancy. *Developmental Psych.*, 36, 635-645.

O'Neill, J., Jacobson, S., & Jacobson, J. (1994). Evidence of observer reliability for the Fagan Test of Infant Intelligence. *Infant Behavior & Dev.*, 17, 465-469.

Quinn, P. & Eimas, P. (1996). Perceptual cues that permit categorical differentiation of animal species by infants. *J. of Experimental Child Psychology*, 63, 189-211.

Quinn, P. & Eimas, P. (1998). Evidence for a global categorical representation of humans by young infants. *J. of Experimental Child Psychology*, 69, 151-174.

Quinn, P., Eimas, P., & Rosenkrantz, S. (1993). Evidence for representations of perceptually similar natural categories by 3- and 4-month-old infants. *Perception*, 22, 463-475.

Schuler, E. M. (Ed.) (1980). *Simon and Schuster's guide to dogs*. New York: Simon and Schuster.

Siegel, M. (Ed.) (1983). *Simon and Schuster's guide to cats*. New York: Simon and Schuster.

Sokolov, E. N. (1963). *Perception and the conditioned reflex*. Hillsdale, NJ: LEA.