Asymmetric interference in 3- to 4-month-olds' sequential category learning

Denis Mareschal
Birkbeck College

Paul C. Quinn
Washington & Jefferson College

Robert M. French
Université de Liege

[Reference to cite: Mareschal, D., Quinn, P., & French, R. M., (2002). Asymmetric interference in 3- to 4-month olds' sequential category learning. *Cognitive Science, 79*, 1-13.]

Running Head: Asymmetric interference in infancy

This work was funded by an award from the Royal Society and the Economic and Social Research Council (R000239112) to the first author, European Commission RTN grant HPRN-CT-2000-00065 to the first and third authors and by Grant BCS-0096300 from the National Science Foundation to the second author. We are grateful to Les Cohen, Carolyn Rovee-Collier, Tom Shultz, and Marius Usher for helpful comments on an earlier version of this work.

Address all correspondence to Denis Mareschal, Centre for Brain and Cognitive Development, School of Psychology, Birkbeck College, University of London, Malet Street, London, WC1E 7HX, UK. Email: d.mareschal@bbk.ac.uk

Abstract

Three- to 4-month-old infants show asymmetric exclusivity in the acquisition of Cat and Dog perceptual categories. We describe a connectionist autoencoder model of perceptual categorization that shows the same asymmetries as infants. The model predicts the presence of asymmetric catastrophic interference (retroactive interference) when infants acquire Cat and Dog categories sequentially. A subsequent a study with 3- to 4-month-olds verifies this predicted pattern of behavior. We argue that bottom-up, associative learning systems with distributed representations are appropriate for modeling the operation of short-term visual memory in early perceptual category learning.

Asymmetric interference in 3- to 4-month-olds' sequential category learning

Young infants can form well-defined perceptual category representations when presented with a series of static visual stimuli (e.g., Cohen & Strauss, 1979; Bomba & Siqueland, 1983; Quinn, 1987; Mareschal & Quinn, 2001). Even newborns can form primitive category representations for simple visual forms (Slater, 1995) and by 3 to 4 months infants can categorize a range of real world images of cats, dogs, horses, couches, and chairs (see Quinn & Eimas, 1996 for a complete review). However, the perceptual categories formed by young infants do not always have the characteristics that might be expected from the corresponding adult categories. For example, Quinn, Eimas, and Rosenkrantz (1993) have found that, when shown a series of cat photographs, 3- to 4-month-olds will form a perceptual category representation of Cat that excludes dogs, but when shown a series of dog photographs, the same infants will form a perceptual category representation of Dog that does not exclude cats. Similar asymmetries have now been found with a range of different stimulus sets (e.g., Younger & Fearing, 1999).

In previous work, we have shown that early infant perceptual categorization is well accounted for by the computational properties of an associative learning system with distributed representations (Mareschal & French, 2000; Mareschal, French, & Quinn, 2000). We used a connectionist autoencoder network to model Cat and Dog perceptual category learning by 3- to 4-month-olds. When exposed to the same stimulus set as the infants, the model developed Cat and Dog perceptual categories that had the same exclusivity asymmetry as observed in infant perceptual categories.

In distributed associative neural networks (such as the autoecoder model) categorization arises as a bi-product of information storage in a dynamic associative memory system (Knapp & Anderson, 1984). As the associative learning mechanism attempts to encode individual stimuli, features with predictive value that are repeatedly presented are reinforced while features that are unique to individual exemplars are overwritten by the successive presentation of different exemplars. As a result, such networks develop an implicit prototype category representation that reflects the distribution of features in the environment.

In this paper we extend our examination of simple connectionist networks as models of infant perceptual category learning by investigating how these networks and young infants cope

with sequential learning of two similar but separable perceptual categories. In the real world, categories are rarely learned in isolation. Hence, it is important to consider how the order in which categories are learned may impact on their acquisition. The prior learning of one category may facilitate the subsequent learning of a second category by providing a contrasting reference that helps define the extension of the category. Alternatively, the prior learning of one category may inhibit the subsequent learning of a second category in a manner analogous to proactive interference. Equally, the subsequent learning of a second category may enhance the memory of the first category, or it may interfere with the stored representation of the first category.

There has been little research on sequential perceptual category learning in infancy. One notable exception is the work of Eimas, Quinn, and Cowan, (1994) who suggested that learning a second category should be easier than learning a first category because the first category serves as a frame of reference against which to judge the instances of the second category. However, our network account makes a different prediction. Connectionist networks are susceptible to a form of retroactive interference called catastrophic interference in which subsequently acquired material overwrites previously acquired material (McCloskey & Cohen, 1989; Ratcliff, 1990; for a review, see French, 1999). The degree of interference depends on the degree of similarity between the old and new material. As a result, the degree to which learning a second category will interfere with a previously acquired category will depend on the amount of feature overlap in the two categories. Mareschal et al (2000) reported that, in the images used to test 3- to 4-month-olds on Cat and Dog categories, the distribution of cat values were largely subsumed within the distribution of dog values. This suggests that the learning of Dog following the initial learning of Cat will interfere with the initial Cat perceptual category (because the new dog exemplars will constitute items outside the Cat perceptual category). However, the learning of Cat following the initial learning of Dog will not interfere with the initial Dog perceptual category (because the cat exemplars constitute items within the Dog perceptual category). We tested these hypotheses, first in connectionist autoencoder networks, then with 3- to 4-month-olds.

Building the model

Infant visual categorization tasks rely on preferential looking techniques based on the finding that infants direct more attention to unfamiliar or unexpected stimuli. The standard interpretation of this behavior is that the infants are comparing an input stimulus to an internal representation of the same stimulus (e.g., Charlesworth, 1969; Cohen, 1973; Sokolov, 1963). 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 stimulus. While attending to the stimulus the infant updates its internal representation. When the information in the internal representation is no longer discrepant with respect to 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 much attending because an internal representation has to be constructed or adjusted. The degree to which the novel object differs from existing internal representations determines the amount of adjusting that has to be done, and hence the duration of attention.

We used a connectionist autoencoder to model the relation between attention and representation construction (Mareschal & French, 2000; Mareschal et al., 2000; Schafer & Mareschal, 2001). The network learns to reproduce on the output units the pattern of activation presented to the input units. Learning in such networks is unsupervised because the (perceptual)

input signal also serves as the training signal for the output. Unsupervised autoencoder networks have been found to match or outperform supervised networks on a range of natural classification tasks (Japkowicz, 2001). 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 representation is tested by expanding it, and comparing the resulting predictions to the actual stimulus being encoded. Similar networks have been used to produce compressed representations of video images (Cottrell, Munro, & Zipser, 1988).

We suggest that during the period of captured attention infants are actively involved in an iterative process of encoding visual input into an internal representation and then assessing that representation against continuing perceptual input. This is accomplished by using the internal representation to predict what the properties of the stimulus are. As long as the representation fails to predict the stimulus properties, the infant continues to fixate the stimulus and to update the internal representation. Our modeling is based on the assumption that infant looking time is positively correlated with network error[1].

Sequential category learning in autoencoder networks

Data for training the networks were obtained from the original cat and dog pictures used by Quinn et al. (1993) to test infant categorization[2]. The 18 dog and 18 cat photographs were measured along the following ten perceptual dimensions: head length, head width, eye separation, ear separation, ear length, nose length, nose width, leg length, vertical extent, and horizontal extent. Although it is difficult to say for certain which features the infants are using during categorization, it is well known that infants can segregate items into categories on the basis of attributes with different values (e.g., Younger, 1985). The feature values (measured in millimeters) were then normalized to range within 0 and 1. The resulting 36 (18 dogs and 18 cats), 10 dimensional continuous valued arrays were used to train and test the networks.

The networks used were standard 10-8-10 feedforward autoencoders trained using the backpropagation algorithm with the following parameter values: learning rate =0.2, momentum =0.9, Fahlman offset =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 are presented to the infants in experimental studies of perceptual categorization (Quinn & Eimas, 1996). A network was 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 original Quinn et al. (1993) studies, infants were shown pairs of pictures for a fixed duration of time. The network was then tested a first time (T1) with a novel exemplar of the same category and a novel exemplar of the novel category. As in the experimental procedures with infants, higher error on (a preference for looking at) the exemplar from the novel category rather than a novel exemplar from the same category is taken as evidence that the network has formed a perceptual category representation that excludes exemplars from a novel category but includes novel exemplars from the familiar category.

Following this initial phase, the network was trained on 4 exemplars (2 pairs) of the contrasting category. If the network had initially been trained on cats it was presented with 2 pairs of dogs. If it had originally learned dogs, the network was presented with 2 pairs of cats. Finally, the network was tested a second time (T2) with novel exemplars as in the first test session. The network's ability to autoencode a novel test stimulus accurately (i.e., to have low

output error when presented with the novel stimulus) reflects the extent to which the category specific information is encoded accurately in the connection weights. Hence, the difference in the network's performance at T2 as compared to T1 is a measure of the amount of interference (or distortion of the category representations) that occurred as a consequence of learning the intervening exemplars. The results reported below are averages over 50 networks.

========== Insert Figure 1 about here ==========

Figure 1a shows the difference between the network's performance at T1 and T2, when (a) the initial category was Cat and the intervening category was Dog, and (b) the original category was Dog and the intervening category was Cat. For networks initially trained on cats, performance at T1 shows much higher error for (a clear preference for looking at) a novel dog over a novel cat. However, at T2 (following the intervening presentation of dog exemplars) this difference is greatly reduced implying that there is no clear preference for looking at a novel dog over a novel cat. The difference in response patterns at T1 and T2 is due to a substantial increase in error to a novel cat at T2 as compared to T1. In short, the learning of dogs during the intervening period has strongly interfered with the previously acquired internal representation of cats.

A markedly different pattern of behavior emerges from networks originally trained with dogs. At T1, there is only a small difference in error for (no preference for looking at) a novel dog over a novel cat. This finding replicates the results of Mareschal et al. (2000) and is consistent with the empirical finding that 3- to 4 month-olds will show no significant preference for cats over dogs under these conditions (Quinn et al., 1993). At T2, the networks still show only a small difference in error for (no preference for looking at) a novel cat over a novel dog. The subsequent learning of Cats during the intervening period has not interfered with the previously acquired internal representation for Dog. The difference in response patterns at T1 and T2 arises from a drop in error to novel cats at T2 as compared to T1 that is due to the decreased novelty of cats. It is not due to a change in the error for dogs[3].

In summary, the subsequent learning of Dog will interfere with a previously acquired representation of Cat, but the subsequent learning of Cat will not interfere with the prior representation of Dog.

The asymmetric interference can be traced to (1) the distribution of feature values in the examplars used to train the networks, and (b) the fact that network responses are based on an internal representation that captures the distribution statistics in the original data (Baldi & Hornik, 1989; Japkowicz, Hanson, & Gluck, 2001). In this data set, most cat exemplars have feature values that fall within 2 standard deviations of dog values, whereas most dog exemplars have feature values that fall outside two standard deviations of cat values (Mareschal et al., 2000). Thus, most cat exemplars are plausible dogs, whereas most dog exemplars are not plausible cats. There is no interference when cats are acquired following dogs because this is equivalent to reinforcing exemplars that are consistent with the internal representation for dogs. In contrast, learning novel dog exemplars interferes with the existing cat internal representation because the new exemplars fall outside the internal representation for cats. The asymmetric interference in sequential category learning is an explicit model prediction about how 3- to 4-month-olds will respond to the sequential presentation of cat and dog photographs.

Sequential category learning in 3- to 4-month-olds

The model predicts that learning to categorize dogs will disrupt a previously learned category of Cat (as measured by a lack of preferential looking towards a novel dog at T2), whereas learning to categorize cats will not disrupt a previously learned category of Dog (as measured by the absence of a change in novelty preference at T2). This stands in contrast to the representation account suggested by Eimas et al. (1994).

Method

Participants. Forty-eight 3- to 4-month-olds (26 boys, 22 girls) were participants (mean age = 3 months, 13 days; SD = 10 days). Nine additional infants were not included in the analyses because of fussiness (n=6), a position bias of > 95% looking to one side of the display (n=1), or a failure to look at both test stimuli (n=2).

Stimuli. The stimuli were thirty-six color photographs of cats and dogs (18 exemplars of each category) previously used in Quinn et al. (1993) and Eimas et al. (1994). The pictures were cut from Simon and Schuster's Guide to Cats (Siegal, 1983) and Simon and Schuster's Guide to Dogs (Schuler, 1980), and chosen to represent a variety of shapes, colors, and stances of both categories of animals. Each picture contained a single animal that had been cut away from its background and mounted onto a white 17.7 by 17.7-cm posterboard for presentation.

Apparatus. Infants were tested by means of a portable visual preference apparatus, adapted from that used by Fagan (1970). The apparatus is an enclosed viewing box with a grey display stage (85 cm wide and 29 cm high) that contains two compartments to hold the two 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 the two compartments was 30.5 cm. A 0.625-cm peephole located midway between the stimulus compartments permitted observation and recording of the infant's visual fixations.

Procedure. The 48 infants were randomly assigned to one of two category presentation orders. Each infant in the Cat-first group was first familiarized with 12 cats, randomly selected and different for each infant, presented during six 15-s trials (two different cats per trial). After this first familiarization phase the formation of a category representation was tested with a novel cat paired with a novel dog. Looking times to the novel cat and novel dog were recorded. Following this test, the infant was familiarized with 4 dogs, randomly selected and different for each infant, presented during two 15-s trials (two different dogs per trial). Finally, following this second familiarization phase, the infant was again presented with test trials in which a novel cat was paired with a novel dog. This preference test was conducted in two 10-s trials. The left-right positioning of the novel instance from the familiar category and the novel instance from the novel category were counterbalanced across infants on the first test trial and reversed on the second test trial. Looking time to the novel cat and novel dog was recorded a second time. Each infant in the Dog-first group was familiarized and tested in the same way as those in the Cat-first group except that dog pictures were presented during the first familiarization phase and cat pictures during the second.[4]

Results

Familiarization trials. Table 1 shows the mean fixation times averaged across the first three familiarization trials, the second three familiarization trials, and the final two interference trials. An ANOVA with first familiarization category (Cat vs. Dog) by trial block (1-3 vs. 4-6) revealed only a significant effect of trial block on the initial familiarization trials, $F(1, 46)=16.11$,

$p < .001$. This result indicates that there was a reliable decrement in looking time from the first to the second block of the initial familiarization trials, and provides evidence that both groups habituated to the information presented during initial familiarization. Moreover, there was no significant difference in the mean looking times during the interference trials for the Cat-first versus Dog-first groups, $t(46) = -0.37$, $p > .20$, two-tailed. Together, these findings suggest that any differences in the preference test outcomes cannot be attributed to category-specific differential habituation rates.

========== Insert Table 1 about here ==========

Preference test trials. Figure 1b shows the duration of infant looking times towards a novel cat and a novel dog during the first (T1) and second (T2) test trials for both the Cat-first and the Dog-first groups. The pattern of looking time responses is strikingly similar to that of the model output error. In all cases, infants look more at the test stimulus that produces the greatest error in the network. In other words, the model captures exactly all trends in the data. In addition, the model also predicts that the significant preference for the novel dog stimulus at T1 in the Cat-first group (reflecting that the Cat perceptual category excludes dogs) should not be present at T2. However, there should be no significant preference for a novel cat stimulus at both T1 and T2 in the Dog-first group.

To test these predictions, the infant looking times were entered into an ANOVA with three factors: Novel stimulus (Cat vs. Dog) and Test Trial (T1 vs. T2) as within subject factors, and Group (Cat-first vs. Dog-first) as a between subjects factors. This analysis revealed a significant three-way Stimulus x Group x Test Trial interaction, $F(1, 46) = 6.01$, $p<.02$. This was the only significant effect (all other F's <0.42). The three-way interaction was explored by carrying out separate two-way ANOVAS with Stimulus and Group as factors at each level of Trial.

At trial T1, the two-way ANOVA revealed an interaction of Stimulus x Group, $F(1, 46) = 4.71$, $p<.04$, as the only significant effect (all other F's <0.48). This two-way interaction could be explained by comparing the looking times towards a novel cat and a novel dog within each group. This latter analysis revealed that the infants looked significantly longer at the novel dog than the novel cats in the Cat-first group, $F(1, 23) = 6.420$, $p<.02$, but showed no significant difference in looking times in the Dog-first group, $F(1, 23)=.97$, $p>.33$. These results replicate those reported in Quinn et al. (1993). Quinn et al. also reported that infants familiarized with either cats or dogs will show a significant preference for a novel bird over a novel exemplar of the familiarization category, and that the infants can discriminate individual exemplars within the cat and dog categories. Taken together, these results lead to the conclusion that the infants have formed a perceptual category representation of Cats (in the Cat-first group) that excludes novel dogs and novel birds, but that they have formed a perceptual category representation of Dog (in the Dog-first group) that excludes novel bird exemplars and includes novel cat exemplars.

The critical issue is what happens to the pattern of looking times at T2. According to Eimas et al. (1994), there should be a significant novelty preference in both groups, because the presence of a contrasting category will help the infants separate the two categories in the Dog-first group. By contrast, the simulations suggest that there should be no significant novelty-preference in either group because the initial Cat representation in the Cat-first group will have

undergone interference from the newly encountered dog exemplars, whereas in the Dog-first group, the newly encountered cat exemplars will not have changed the Dog representation. In addition, there should be a significant increase in looking towards the novel cat exemplars between T1 and T2 in the Cat-first group (because of the disrupted category representation), but no significant increase in looking towards the novel dog exemplar between T1 and T2 in the Dog-first group. In fact, the pattern of looking time results is entirely consistent with the predictions from the autoencoder networks. At Trial T2, the two-way ANOVA revealed no significant effects (all F values <2.2). There are no significant differences in looking times across stimuli and conditions at T2 as predicted by the network models. In addition, in the Cat-first group, there was a significant increase in looking towards the novel cat exemplars from T1 to T2, ($t(23)=1.97$, $p=.03$, one-tailed) but no significant increase in looking towards the novel dog exemplar from T1 to T2 in the Dog-first group ($t(23)=1.32$, $p>.10$, one-tailed).

In summary, at the initial test T1, the infants familiarized with 12 cats showed a significant preference for looking at a novel dog over a novel cat, whereas the infants familiarized with 12 dogs did not show a significant preference for looking at a novel cat over a novel dog[5]. Both of these results are consistent with the previous finding that infants familiarized with cats will form a perceptual category representation that excludes dogs, whereas infants that are familiarized with dogs will form a perceptual category representation that does not exclude cats (Quinn et al. 1993). In contrast, at T2, the infants showed no significant preferences for the test stimuli in either of the familiarization conditions. In other words, for infants in Cat-first group, the subsequent learning of dog exemplars disrupted their previously acquired Cat perceptual category representation. In contrast, for infants in the Dog-first group, the subsequent learning of cat exemplars has not affected their previously acquired perceptual category representation of Dog[6].

General Discussion

We presented a connectionist autoencoder model of sequential category learning. The model predicted the presence of asymmetric retroactive interference when young infants learn cat and dog perceptual categories sequentially. The subsequent learning of dogs will disrupt the prior learning of cats, whereas the subsequent learning of cats will not disrupt the prior learning of dogs. This is a strong model prediction since retroactive interference in sequential category learning in infants has not been investigated before. The prediction of asymmetric retroactive interference was found to hold true for 3- to 4-month-olds required to learn the cat and dog perceptual categories sequentially.

The asymmetric interference can be traced to an inclusion relation in the distribution of feature values for cat and dog exemplars used to familiarize infants in the experimental procedure, and the fact that the hidden unit representations in the network reflect this distribution. An analysis of the data explains why asymmetric categorization is observed in infant behavior while the connectionist model explains how that data gets translated into behavior. In other words, the autoencoder model embodies a specific process account of how feature distribution information in the environment gets translated, through learning, into observable behavioral asymmetries.[7]

Knowing that there is an inclusion relation in the data is not enough to predict an asymmetry in the behavior. Many computational systems could process the same data and not produce an asymmetry in categorization. It is because the connectionist network develops internal representations that reflect the distribution of features in the data that this behavior is

observed. This analysis is not dependent on our use of backpropagation to train the networks. We chose to use backpropagation as a means of implementing gradient descent learning in a distributed artificial neural network. Many other gradient descent network learning algorithms would result in the same behavioral results (e.g., Grossberg, 1982; Ackley, Hinton, & Sejnowski, 1985).

An implication of this model is that much of early infant perceptual categorization is a bottom-up process. In contrast to adults seeing photographs of cats and dogs, the infants in these studies are responding to the stimuli solely on the basis of low level statistics (i.e., appearance of surface features and their frequency) and not the semantics of the representation (French, Mermillod, Quinn, & Mareschal, D., 2001). This process is analogous to distribution sensitive category abstraction in adults (e.g., Posner & Keele, 1970; Reed, 1972; Fried & Holyoak, 1984). Category learning by young infants (even in natural kind domains) reflects a bottom-up data driven process rather than the acquisition of "theories" or the unfolding of innate taxonomic structures.

Finally, the results of this model suggest that catastrophic interference (a well documented phenomenon in connectionist networks; French, 1999) may play an important role in early memory and categorization. Some evidence of catastrophic interference in early infant visual memory already exists. During the late to mid-seventies there was a debate surrounding the robustness of infant visual memory. A number of labs (e.g., Deloache, 1976; Fagan, 1973; McCall, Kennedy, & Dodds, 1977) suggested that infants suffer from substantial forgetting if presented with new material during the retention interval. These studies relied on a habituation procedure. Infants were habituated to a first image (A). After habituation to this image, they were habituated to a second image (B). After this second habituation phase, infants were presented with A again. A release in habituation to A (as measured by a renewed interest in A) was interpreted as suggesting that the intervening habituation to B had caused the memory of A to disappear. The puzzling thing was that retroactive interference did not occur with all stimuli. In some cases interference occurred whereas in other cases it did not (Cohen, Deloache, & Pearle, 1977; Fagan 1977). The only conclusions from these studies were (1) that it was necessary for the infants to encode B for interference to occur, and (2) that interference was related to the similarity between the images A and B. We believe that performance on the perceptual categorization and memory tasks reflects the operation of the same information processing mechanisms. Namely, it reflects the way in which information is stored in an associative system with distributed representations and, therefore, performance on the two classes of tasks is subject to the same interference effects.

Of course, this does not preclude the fact that infants have robust long-term memory. Indeed there is ample evidence of long term retention in early infancy using a range of testing methodologies (Nelson, 1995, Rovee-Collier, 1997; Bauer & Mandler, 1990). However, note that even though adults have robust long term memory, their short-term visual memory is also susceptible to interference (Dempster & Brainerd, 1995).

In summary, this paper has reported on a connectionist model of infant short-term visual memory and perceptual category abstraction. The model predicted asymmetric retroactive interference in the sequential learning of perceptual categories by young infants. An empirical study with 3- to 4-month-olds confirmed this prediction with cat and dog categories. Finally, the model highlighted how constructing computational models help further our understanding of

cognitive development by providing a tool for synthesis across multiple domains and a bridge between processing in infancy and adulthood.

References

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzman machines. Cognitive Science, 9, 147-169.

Baldi, P., & Hornik, K. (1989). Neural networks and principal components analysis: Learning from examples without local minima. Neural Networks, 2, 52-58.

Bauer, P. J., & Mandler, J. M. (1990). Remembering what happened next: Very young children's recall of event sequences. In R. Fivush & J. Hudson (Eds.), Knowing and remembering in young children (pp. 9-29). New York: Cambridge University Press.

Bomba, P. C., & Siqueland, E. R. (1983). The nature and structure of infant form categories. Journal of Experimental Child Psychology, 35, 294-328.

Charlesworth, W. R. (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, UK: Oxford University Press.

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

Cohen, L. B., Deloache, J. S., & Pearl, R. A. (1977). An examination of interference effects in infants' memory for faces. Child Development, 48, 88-96.

Cohen, L. B., & Strauss, M. S. (1979). Concept acquisition in the human infant. Child Development, 50, 419-424.

Cottrell, G. W., Munro, P., & Zipser, D. (1988). Image compression by backpropagation: An example of extensional programming. In N. E. Sharkey (Ed.), Advances in cognitive science, Vol. 3 (pp. 208-240). Norwood, NJ: Ablex.

Deloache, J. S. (1976). Rate of habituation and visual memory in infants. Child Development, 47, 145-154.

Dempster F. N., & Brainerd C. J. (1995). Interference and inhibition in cognition. San Diego, CA: Academic Press.

Eimas, P. D., Quinn, P. C., & Cowan, P. (1994). Development of exclusivity in perceptually based categories of young infants. Journal of Experimental Child Psychology, 58, 418-431.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). Rethinking innateness: A connectionist perspective on development. Cambridge, MA: MIT Press.

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

Fagan, J. F. III (1973). Infant delayed recognition memory and forgetting. Journal of Experimental Child Psychology, 16, 424-450.

Fagan, J. F. III (1977). Infant recognition memory: Studies in forgetting. Child Development, 48, 68-78.

Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. Science, 164, 668-670.

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.

French, R. M. (1992). Semi-distributed representations and catastrophic forgetting in connectionist networks, Connection Science, 4, 365-377.

French, R. M. (1997). Pseudo-recurrent connectionist networks: An approach to the "sensitivity–stability" dilemma. Connection Science, 9, 353-379.

French, R. M. (1999). Catastrophic forgetting in connectionist networks. Trends in Cognitive Science, 3, 128-135.

French, R. M., Mermillod, M. & Quinn, P. C., Mareschal, D. (2001). Reversing category exclusivities in infant perceptual categorization: Simulation and data. In J. D. Moore & K. Stenning (Eds.) Proceedings of the twenty-third annual conference of the Cognitive Science Society (pp. 307-312). London: LEA

Grossberg, S. (1982). How does a brain build a cognitive code? Psychological Review, 87, 1-51.

Hunter, M. A., Ames, E. W., & Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. Developmental Psychology, 19, 338-352.

Japkowicz, N. (2001) Supervised and unsupervised binary learning by feedforward neural networks. Machine Learning, 42, 97-122.

Japkowicz, N., Hanson, S. J., & Gluck, M. A. (2000). Nonlinear autoassociation is not equivalent to PCA. Neural Computation, 12, 531-545.

Knapp, A. G., & Anderson, J. A. (1984). Theory of categorization based on distributed memory storage. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 616-637.

Luce, D. (1995). Four tensions concerning mathematical modeling in psychology. Annual Review of Psychology, 46, 1-26.

Mareschal, D. & Quinn, P. C. (2001) Categorisation in Infancy. Trends in Cognitive Science, 5, 443-450.

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.

McCall, R. B., Kennedy, C. B., & Dodds, C. (1977). The interfering effect of distracting stimuli on infant's memory. Child Development, 48, 79-87.

McCloskey, M. & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), The psychology of learning and motivation, Vol. 23 (pp. 109-164). New York: Academic Press.

Nelson, C. A. (1995). The ontogeny of human memory: A cognitive neuroscience perspective. Developmental Psychology, 31, 723-738.

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. Advances in Infancy Research, 10, 1-36.

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

Ratcliff, R. (1990). Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions. Psychological Review, 97, 285-308.

Reed, S. K. (1972) Pattern recognition and categorization. Cognitive Psychology, 3, 382-407.

Rovee-Collier, C. (1997). Dissociations in infant memory: Rethinking the development of implicit and explicit memory. Psychological Review, 104, 467-498.

Rumelhart, D., & McClelland, J. (1986). Parallel distributed processing, Vol. 1. Cambridge, MA: MIT Press.

Slater, A. (1995). Visual perception and memory at birth. Advances in Infancy Research, 9, 107-125.

Schafer, G., & Mareschal, D. (2001). Modeling infant speech sound discrimination using simple associative networks. Infancy, 2, 7-28.

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

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

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

Younger, B. A. (1985). The segregation of items into categories by ten-month-old infants. Child Development, 56, 1574-1583.

Younger, B. A. & Fearing, D. D. (1999). Parsing items into separate categories: Developmental change in infant categorization. Child Development, 70, 291-303.

Footnotes

1 Although the current model can explain looking times to stimuli when they are presented one at a time, it does not capture the pattern of shared looks that occurs when two stimuli are presented in pairs as is the case in preferential looking tasks. However, the current model can easily be extended to account for this case by assuming that competition occurs for attention to either one of the two stimuli. A system that looks first at the stimulus with the least output error, and continues to look at that stimulus until error has dropped below some threshold, then shifts to looking at the second stimulus (with higher initial error), will capture the classic pattern of looks in preferential looking tasks. Indeed, infants initially look at the familiar stimulus, followed by a longer look at the novel stimulus, resulting in greater total looking towards the novel stimulus (Hunter, Ames, & Koopman, 1983).

2 A full description of the raw data, including details of the frequency distributions can be found in Mareschal et al. (2000).

3 This asymmetric interference could be due to unequal initial learning of the Dog and Cat categories by the networks. To explore this possibility, 50 new networks were trained to a fixed error criterion rather than a fixed epoch criterion. For practical reasons, a maximum criterion of 2500 epochs (10 times the 250 epoch criterion) was used to terminate any simulations that failed to reach the 0.2 error criterion. This is analogous to the fact that, in practice, any study with infants has a fixed maximum duration. The new training procedure was identical to that described above except that all training continued until all output units were within 0.2 of their target values. This ensures that the networks have learned to autoencode each input to the same minimum criterion. Under these conditions, the networks initially trained with Cats, and subsequently trained with Dogs, showed an average error of 0.28 (SD=0.11) and 0.46 (SD=0.12) to a novel cat and a novel dog at T1 respectively, and an average error of 0.38 (SD=0.11) and 0.38 (SD=0.13) to a novel cat and novel dog at T2 respectively. In contrast, networks initially trained with Dogs, and subsequently trained with Cats, showed an average error of 0.42 (SD=0.15) and 0.35 (SD=0.12) to a novel cat and a novel dog at T1 respectively, and an average error of 0.31 (SD=0.16) and 0.38 (SD=0.12) to a novel cat and novel dog at T2, respectively. Asymmetric interference is thus found even in networks trained to a fixed error criterion.

4 Each infant was brought to the second author's laboratory by a parent and placed in a reclining position on the seated parent's lap. An experimenter wheeled the apparatus over the infant keeping the infant's head centered with respect to the middle of the display stage. As soon as the infant was properly aligned, a trial was begun. The experimenter loaded the stimuli from a nearby table into the stimulus compartments, elicited the infant's attention and closed the stage, thereby exposing the stimuli to the infant. The center of the display stage was approximately 30.5 cm in front of the infant while the stimuli were in view. During a trial, the experimenter observed the infant through the peephole and recorded fixations to the left and right stimuli using a 605 XE Accusplit stopwatch held in each hand. The criterion for fixation was observing the corneal reflection of the stimulus over the infant's pupil. Interobserver reliability, as 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, was high (Pearson $r = 0.97$). Between trials, the experimenter opened the stage, recorded the looking time data on a data sheet, changed the stimuli (or their position), recentered the infant's gaze, and closed the stage, thereby beginning the next trial.

Two experimenters were used to record fixations, one during familiarization and interference trials, and another during test trials. Both were trained research assistants who were naive to the hypotheses of the studies. The experimenter recording during test trials was also naive to the stimulus information that the infant was shown during the familiarization and interference trials.

5 Developmental psychologists have traditionally reported proportional looking times when reporting novelty preferences. In this article, we have reported the raw looking times to highlight the match between the model and the infants behaviors. However, the proportional looking times revealed the same pattern of results. In the Cat-first group, the novel category preference scores at T1 and T2 were 56.98% (SD=13.16) and 46.63% (SD =19.87) respectively, whereas in the Dog first group, the novel category preferences were 51.75% (SD =19.63) and 44.73% (SD =17.89). The only novelty preference significantly different from chance was the preference for a novel dog exemplar shown by the Cat-first group at T1, $t(23)=2.60$, $p<.02$, two-tailed. The results at T1 replicate Quinn et al.'s (1993) findings and are consistent with their conclusions that infants have formed a perceptual category representation for cats that excludes dogs and a perceptual category representation for dogs that includes cats. This conclusion is also informed by the fact that these infants do not show a prior preference for cats or dogs, and that they do show a novelty preference for a novel bird exemplar.

6 It could be argued that the pattern of responses seen in the Cat-first group simply reflects shifts in novelty of a category due to the recency of exposure to exemplars. The change in the infants' novelty preferences at T1 and T2 reflects the fact that the contrasting category in T1 is the most recently experienced category at T2. More specifically, infants in the Cat-first group would look longer at a Dog at T1 because, having just seen a series of Cats, the Cats are less novel. This preference would then disappear at T2 (leading to no preference at T2) because the relative novelty of Dog diminishes as a result of the subsequent exposure to Dog exemplars. This interpretation is consistent with the behavior observed in the Cat-first group. However, it is not consistent with the behavior observed in the Dog-first group. The shifting novelty argument would predict an initial preference for Cats at T1. However, infants in the Dog-first group show no preference for a novel dog or a novel cat at T1. Furthermore, even if one were to argue that the lack of preference for cats at T1 was due to some inherent difficulty at learning dogs, the shifting novelty hypothesis would predict an incorrect preference response at T2. The lack of preference for a cat or a dog at T1 suggests that exemplar from these categories are equally novel at this point in familiarization. The shifting novelty hypothesis would predict that the subsequent exposure to cats would decrease the relative novelty of cats, thereby resulting in a preference for a novel dog at T2. However, this is not the case. Infants in the Dog-first group continue to show no preference for cats or dogs at T2. Consideration of the behavior of the Cat-first group in conjunction with the behavior of the Dog-first group rules out the shifting novelty hypothesis. Thus, this pattern of preferential looking responses cannot be attributed simply to the shifting novelty of the exemplars as a result of sequential presentation.

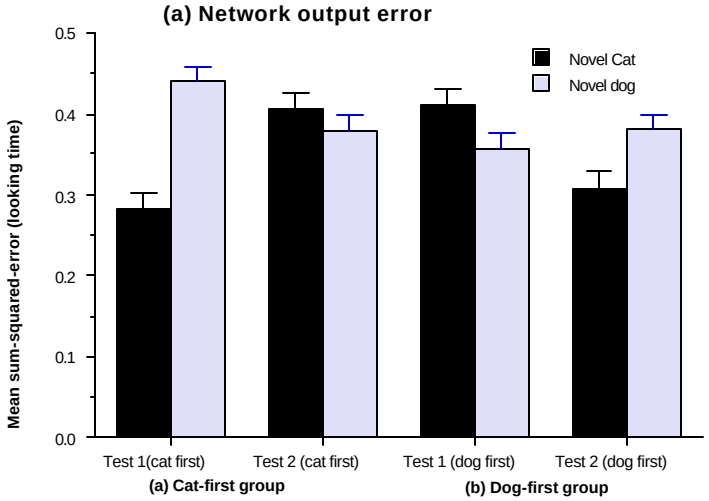
7 See Luce (1995) for a discussion of the differences between process models and normative or descriptive models.

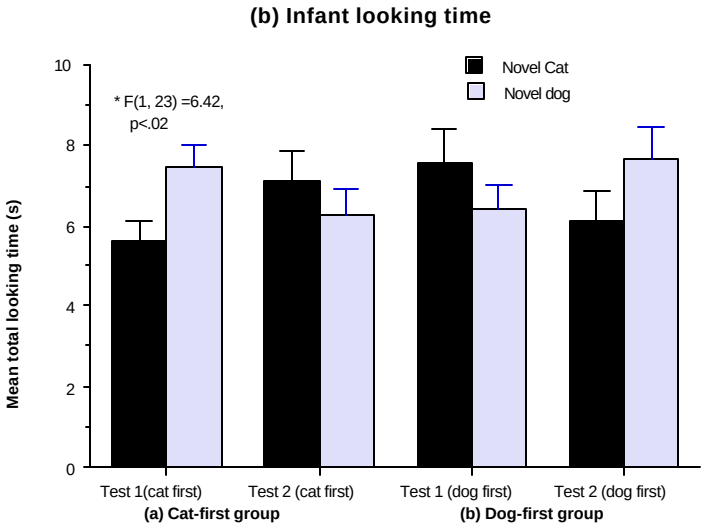
Table 1.Mean fixation times (in seconds) and standard deviations (in parentheses) during familiarization and interference trials.

Group	Trials 1-3	Trials 4-6	Trials 7-8 (Interference trials)
CAT-first	10.78 (1.92) (Cats)	9.44 (3.04) (Cats)	9.86 (3.19) (Dogs)
DOG-first	10.98 (2.92) (Dogs)	9.35 (2.98) (Dogs)	10.22 (3.47) (Cats)

Figure Captions

Figure 1. Response of (a) model and (b) infants to novel and familiar exemplars at tests T1 and T2. The differences in height between the solid and hashed bars represent the degree of preferential looking at one or the other novel stimulus under the different test conditions.




(b) Infant looking time
Novel Cat
Novel dog
* F(1, 23) =6.42,
p<.02
Mean total looking time (s)
10
8
6
4
2
0
Test 1(cat first)
Test 2 (cat first)
Test 1 (dog first)
Test 2 (dog first)
(a) Cat-first group
(b) Dog-first group