A Connectionist Account of Perceptual Category-Learning in Infants[1]

Denis Mareschal
Centre for Brain and Cognitive Development
Department of Psychology
Birkbeck College
London, WC1E 7HX, UK
d.mareschal@bbk.ac.uk

Robert M. French
Psychology Department, B32
Université de Liège
4000 Liège, Belgium
rfrench@ulg.ac.be

Abstract

This paper presents a connectionist model of correlation-based categorization by 10-month-old infants (Younger, 1985). Simple autoencoder networks were exposed to the same stimuli used to test 10-month-olds. Both infants and networks used co-variation information (when available) to segregate items into separate categories. The model provides a mechanistic account of category learning within a test session. It shows how distinct categories are developed and demonstrates how categorization arises as the product of an inextricable interaction between the subject (the infant) and the environment (the stimuli).

Introduction

The ability to categorize underlies much of cognition. It is a way of reducing the load on memory and other cognitive processes (Rosch, 1975). Because of its fundamental role, any developmental changes in the ability of infants to categorize is likely have a significant impact on subsequent cognitive development as a whole. As a result, categorization is one of the most fertile areas of research in infant cognitive development.

Many studies of infant categorization have relied on visually presented material. The basic idea of these studies is to show infants a series of images that could be construed as forming a category. The infant's subsequent response to a previously unseen image is used to gauge whether the infant has formed a category based on his or her experience with the familiarization exemplars. Generalization to a novel exemplar from the familiar category, coupled with a preference or heightened responsiveness to a novel exemplar from a novel category is taken as evidence of category formation. There is considerable evidence that young infants can form categorical representations of shapes, animals, furniture, faces, etc. (see Quinn & Eimas, 1996, for a recent review).

At first, the categories developed by infants may appear similar to those developed by adults. However, occasionally, infant categories differ dramatically from those of adults. Quinn, Eimas, and Rosenkrantz (1993) report one striking example. These authors found that when 3.5-month-olds were shown a series of cat photographs, the infants would develop a category of CAT that included novel cats and excluded novel dogs (in accordance with the adult category of CAT). However, when 3.5-month-olds were shown a series of dog photographs, they would develop a category of DOG that included novel dogs but also included novel cats (in contrast to the adult category of DOG). There is an asymmetry in the exclusivity of the CAT and DOG categories developed by 3.5-month-olds.

To understand the source of this asymmetry, one needs to explore the basis on which infants categorize items. While there have been many studies describing infant categorization competence at various ages, there have been few mechanistic accounts of how the underlying categorical representations might emerge. One partial exception is the work by Quinn and Johnson (1997). These authors used a connectionist model to explore the order in which basic and super-ordinate level categories are acquired. Because the model was implemented as a working computer simulation, it is one of the first studies to ask how the mechanisms of learning constrain the nature of the categories that are acquired. Although this work explored how the characteristics of exemplars at different levels might dictate the order in which categories are acquired by infants as a whole, it did not directly address the issue of how categories are learned within the short-term testing sessions characteristic of many published categorization studies.

We believe that the way to a comprehensive synthesis of the numerous competence studies that abound in the infancy literature is to shift the debate to a mechanistic level. If the different studies are tapping into a common categorization ability, then there must exist a common set of mechanisms that can account for the observed behaviors. The search for a common set of mechanisms underlying performance on different tasks has already been successfully applied to explaining the causes of the exclusivity asymmetry mentioned above and a catastrophic interference effect in infant memory studies (Mareschal & French, 1997; Mareschal, French, & Quinn, submitted).

Mareschal & French (1997) and Mareschal *et al* (submitted) presented connectionist networks with the same cat and dog exemplars used to familiarize infants in the original Quinn *et al.* (1993) study. The networks developed the same exclusivity asymmetries as had the infants (i.e., the category of CAT excluded novel dogs, whereas the category of DOG did not exclude novel cats). This was accounted for in terms of the distribution of feature values in the familiarization stimuli and the fact that the connectionist networks developed internal representations reflecting the variability of the inputs they experienced. For almost all features, the distribution of

[1] A longer version of this paper will appear in Infancy.

CAT values was subsumed within the distribution of DOG values. The same mechanism was used to account for the fact that sometimes (but not always) material presented to infants during a retention interval leads to the catastrophic forgetting of the initial material. The model made the prediction that the subsequent learning of the DOG category would disrupt the prior learning of the CAT category, but that the subsequent learning of the CAT category would not disrupt the prior learning of the DOG category. This prediction was tested and found to be true for 3.5-month-olds (Mareschal & French, 1997; Mareschal, French, & Quinn, submitted). In short, the model demonstrated how the previously unrelated exclusivity asymmetry and interference effects were two sides of the same mechanistic coin.

This previous work establishes that autoassociators provide a good model of how categories overlap. However, one purpose of categorization is to parse the world into distinct units that are then acted on differently. Ultimately infants learn to separate out categories. In this paper, we will extend the previous work by exploring the basis on which distinct categories are developed by infants and connectionist networks given a series of exemplars. Younger (1985) showed that 10-month-olds could use the correlation between feature values to segregate items into separate categories. Although these results are based on presenting infants with line drawings of artificial animals, Younger (1990) found that infants could still use correlation information with natural kind images similar to those used in the Quinn et al. studies. We will explore whether the autoencoder connectionist architecture used to model the Quinn et al. data (Mareschal & French, 1997; Mareschal, French, Quinn, submitted) also responds to correlation information in the same way as infants.

The rest of this paper unfolds as follows. First we will describe in detail Younger's (1985) categorization studies with 10-month-olds. Next we will present connectionist simulations of categorization using the same stimuli as Younger used with her infants. We will present an illustration of the internal representations developed by the networks.

Category formation in 10-month-olds

The two simulations described below are attempts to model the behavior of 10-month-olds reported by Younger (1985). The network training regime is kept as close as possible to the infant familiarization conditions. Younger examined 10-month-olds' abilities to use the correlation between the variation of attributes to segregate items into categories. In the real world certain ranges of attribute values tend to co-occur. Thus, animals with long necks tend to have long legs whereas animals with short necks tend to have short legs. Younger examined whether infants could use these co-variation cues to segment artificial animal line drawings into separate categories.

In a first experiment, infants were familiarized with a set of exemplars. They were then tested with either: (a) an exemplar whose attribute values were the average of all the previously experienced values along each dimension, or (b) an exemplar containing the modal attribute values (i.e., the most frequently experienced values) along each dimension. Based on the finding that infants direct more attention to novel or unfamiliar stimuli, preference for a modal versus the average stimulus was interpreted as evidence that the infants had formed a single category from all the exemplars (as evidenced by the greater familiarity of the average stimulus). Preference for the average stimulus was interpreted as evidence that the infants had formed two categories (as indicated by the lesser familiarity of the average stimulus) since the boundary between correlated clusters lay on the average values. Younger found that 10-month-olds looked more at the modal stimuli when the familiarization set was unconstrained (i.e., all attribute values occurred with every other attribute value) suggesting that they had formed a single representation of the complete set of exemplars. However, the 10-month-olds looked more at the average stimuli when the familiarization set was constrained such that ranges of feature values were correlated suggesting that they had formed two distinct categories.

In a second experiment, Younger (1985) provided a more stringent test of category formation in infancy. In this experiment, the infants were presented with a constrained familiarization set (i.e., ranges of feature values were correlated across dimensions). However, the familiarization set was designed such that the modal stimulus was identical to the average stimulus. Infants were then tested with the modal/average stimulus and two stimuli with previously unseen attribute values but which were prototypical of the two possible categories contained within the familiarization set. Preference for the average/modal stimulus was interpreted as evidence that the infants had formed two categories (as indicated by the greater familiarity of the previously unseen stimuli) since the boundary between correlated clusters lay on the average/modal values. Preference for the previously unseen stimuli was interpreted as evidence that the infants had formed a single category from all the exemplars (as evidenced by the greater familiarity of the average/modal stimulus). Younger found that, under these conditions, 10-month-old infants looked longer at the average/modal stimuli suggesting that they had formed two distinct categories.

To model performance on these two experiments (in simulations 1 and 2 below respectively), the same artificial animal stimuli used by Younger were encoded for presentation to the networks. These animals were defined by their values along 4 dimensions: Leg length (ranging from 1.5 to 3.5 in intervals of 1.0), Neck length (ranging in value from 1.2 to 5.2 in intervals of 1.0), Tail length (ranging in value from 0.5 to 2.3 in intervals of 0.45), and Ear separation (ranging in values from 0.3 to 2.7 in intervals of 0.6). Because none of the attributes are intended to be more salient than any other attribute, each attribute was scaled to range between 0.0 to 1.0. This transformation ensures that the greater magnitude of one dimension (e.g., Ear separation) does not bias the networks

to attend preferentially to that dimension. Normalization was achieved by dividing each attribute value by the maximum value along that dimension.

Networks were trained in batch mode. That is, all 8 familiarization items were presented as a batch to the network and the cumulative error was used to update the weights (to drive learning). This ensures that all the items in the familiarization set are weighted equally by the networks and is intended to reflect the fact that there were no significant changes in infant looking times across all familiarization trials. Batch learning also ensures that all order effects are averaged out.

Modeling habituation-dishabituation

Infant categorization tasks rely on preferential looking or habituation techniques based on the finding that infants direct more attention to unfamiliar or unexpected stimuli. The standard interpretation of this behavior is that infants are comparing an input stimulus to an internal representation of the same stimulus (e.g., Solokov, 1963; Charlseworth, 1969; Cohen, 1973). 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 the visual input, attention is directed 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 a 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. An autoencoder is a feedforward connectionist network with a single layer of hidden units. The network learns to reproduce on the output units the pattern of activation across the input units. Thus, the input signal also serves as the training signal for the output units. The number of hidden units must be smaller than the number of input or output units. This produces a bottleneck in the flow of information through the network. Learning in an autoencoder consists in developing a more compact internal representation of the input (at the hidden unit level) that is sufficiently reliable to reproduce all the information in the original input. 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 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 the visual input into an internal representation and then assessing that representation against the 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 representations.

This modeling approach has several implications. It suggests that infant looking times are positively correlated with the network error. The greater the error, the longer the looking time. Stimuli presented for a very short time will be encoded less well than those presented for a longer period. However, prolonged exposure after error (attention) has fallen off will not improve memory of the stimulus. The degree to which error (looking time) increases on presentation of a novel object depends on the similarity between the novel object and the familiar object. Presenting a series of similar objects leads to a progressive error drop on future similar objects. All of this is true of both autoassociators (where output error is the measurable quantity) and infants (where looking time is the measurable quantity).

The modeling results reported below are based on the performance of a standard 4-3-4 (4 input units, 3 hidden units, and 4 output units) feedforward backpropagation network. The learning rate was set to 0.1 and momentum to 0.9. Networks were trained for a maximum of 200 epochs or until all output bits were within 0.2 of their targets. This was done to reflect the fact that in the Younger (1985) studies infants were shown pictures for a fixed duration of time rather than using a proportional looking time criterion.

Simulation 1

In this simulation 24 networks were presented with 8 stimuli in which the full range of values in one dimension occurred with the full range of values in the other dimension (the Broad condition). Another 24 networks were presented with the 8 stimuli in which restricted ranges of values were correlated (the Narrow condition). The networks in both conditions were then tested with stimuli made up of the average feature values or the modal feature values. Table 1 shows the normalized values defining the stimuli in the Broad and Narrow familiarization conditions, and the three test stimuli. Figure 1 shows the networks' response to the average and modal test stimuli when familiarized in either the Narrow or Broad conditions. As with the 10-month-olds, networks familiarized in the Narrow condition showed more error (preferred to look) when presented with the average test stimulus than the modal test stimuli. Similarly, as with the 10-month-olds, networks familiarized in the Broad condition showed more error (preferred to look) when presented with the modal test stimuli than the average test stimuli.

Table 1. Normalized familiarization and test stimuli (Exp. 1)

Familiarization Stimuli							
Broad Condition				Narrow Condition			
Legs	Neck	Tail	Ears	Legs	Neck	Tail	Ears
0.27	1.0	0.22	1.0	0.27	1.0	0.8	0.33
0.27	0.23	1.0	1.0	0.27	0.81	1.0	0.33
0.45	0.81	0.41	0.78	0.45	0.81	1.0	0.11
0.45	0.42	0.8	0.78	0.45	1.0	0.8	0.11
0.82	0.42	0.8	0.33	0.82	0.42	0.22	1.0
0.82	0.81	0.41	0.33	0.82	0.23	0.41	1.0
1.0	0.23	1.0	0.11	1.0	0.23	0.41	0.78
1.0	1.0	0.22	0.11	1.0	0.42	0.22	0.78

Test Stimuli				
Average	0.64	0.62	0.61	0.56
Modal1	0.27	1.0	1.0	0.11
Modal2	1.0	0.23	0.22	1.0

Note: Values are scaled to range from 0.0 to 1.0.

This was confirmed by an analysis of variance with one between-subject factor (Conditions: narrow or broad) and one within-subject factor (Stimulus: average or modal) which revealed a significant interaction of Condition x Stimulus ($F(1,46)=752$, $p<.0001$).

Internal category representation

This section describes the internal representations developed by the networks in Simulation 1and discusses how they lead to the observed preferential looking behaviors described above.

From a behavioral perspective, categorization can be said to have occurred when identifiably different exemplars are treated in the same way. In hidden unit space, members

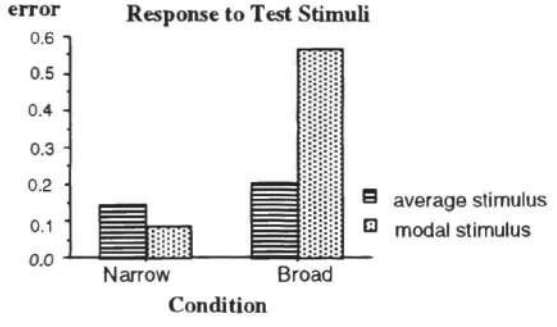

Figure 1. Responses to the average and modal test stimuli for networks familiarized in Broad and Narrow conditions.

of the same category will be mapped to points close together; they will elicit similar activation patterns across the hidden units. Members of different categories will be mapped to points further apart; they will elicit different activation patterns across the hidden units. Because members of a category produce similar hidden unit activation patterns, they will be responded to in a similar fashion by the output units. In contrast, members of a different category that produce different hidden unit activation patterns will be responded to differently by the output units.

Figure 2 shows the distribution of exemplars within the hidden unit space for a representative network trained in the Narrow and Broad conditions of Simulation 1. In the Narrow condition (Figure 2a), exemplars are grouped together in two distinct clusters. One cluster corresponds to those exemplars forming one category and the other cluster correspond to those exemplars forming the second category. The test exemplars are also plotted. Note that the two modal exemplars each fall within (or very close to) one of the category clusters whereas the average exemplar falls between the two clusters. This explains why there is more error (longer looking) to the average exemplar than to either of the modal exemplars. The modal patterns fall within areas that are well covered by the category representations, and hence, for which the network has already learned to make accurate responses. In contrast, the average pattern falls in an area that is not well covered, and hence, for which the network has no experience of making accurate responses.

Figure 2b shows the exemplars within hidden unit space for networks trained in the Broad condition. The internal representations are spread throughout the hidden unit space, reflecting the fact that the exemplars are maximally spread out. Remember that in this condition any feature value can occur with any other feature value. All three of the test stimuli (the average and modal stimuli) project to a similar location at the center of the space. This is because all three have comparable similarities (in terms of feature values) to all of the familiarization exemplars considered individually. There isn't the space in this article to discuss the different ways that similarity can be measured, but by referring to Table 1 we can see intuitively why the test stimuli have comparable similarities to all the familiarization exemplars. Because of the systematic structure of the familiarization set, the average stimulus has feature values that lie mid-way within the range of all possible values. Thus, it is about "half as similar" to any exemplar along any dimension. The modal stimuli have 2 out of the 4 feature values that tend to match the feature values of any particular exemplar. In some cases the match is exact and in others the match is approximate (i.e., both values are high or both values are low). The remaining two values always go in the opposite direction (i.e., the modal value is high when the exemplar value is low or *vice versa*). In short, the three test stimuli are comparably related to the exemplars in the familiarization set: the average stimulus because it has feature values mid-way between the possible range of feature values, and the modal stimuli because they share (approximately) 2 out of 4 feature values with every exemplar.

Finally, because the internal representations are located close to each other in hidden unit space, the network will tend to respond to them in a similar fashion. Since they are in sparsely populated region of the space, the network has little experience with decoding these types of internal

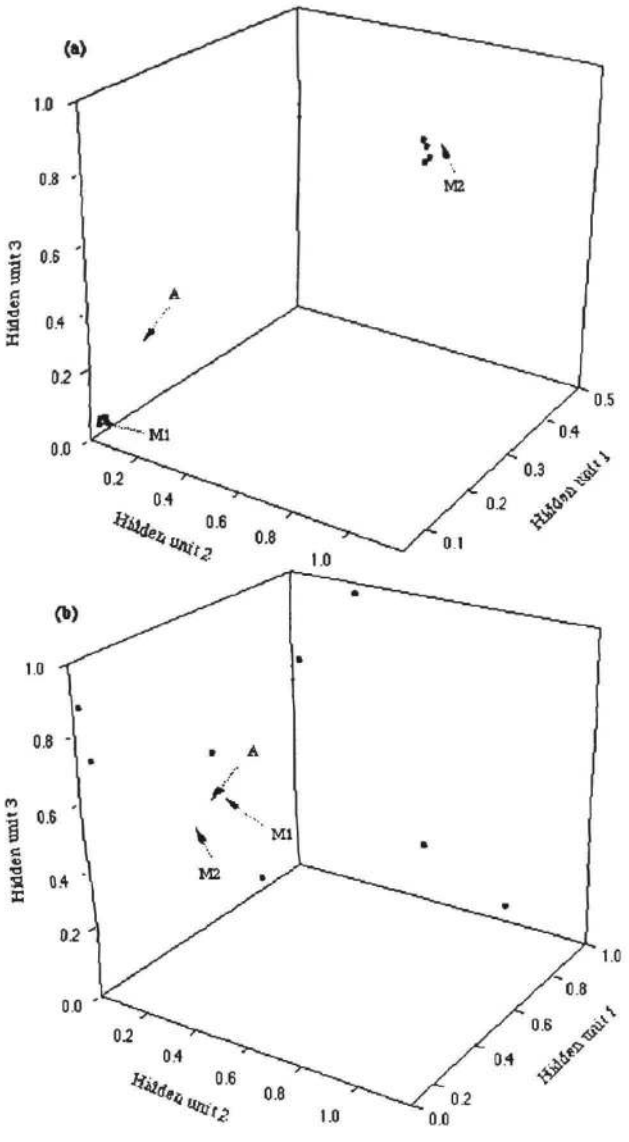

Figure 2. Locations of hidden-unit representations of members of items in the Narrow (2a) and Broad (2b) conditions. M1 and M2 are modal test items, A is an average test item.

representation. As a result, it will output an average of all the outputs it is familiar with. This is fine for the average stimulus since the correct response is precisely the average of all responses (remember that the autoassociation task requires the network to reproduce on the output units the original input values), but it is completely inappropriate for the modal stimuli whose feature values lie at the ends of the possible ranges. Hence, there is more error for the modal stimuli than the average stimulus.

Simulation 2

Younger's (1985) experiment 2 provides a stronger test of category segregation by equating the average and modal values for the full set of familiarization items. In this simulation 24 networks were familiarized with the 10 exemplars designed such that the modal and average values were the same. Under these conditions, the greater familiarity of a stimulus containing previously unseen values (but which are prototypes of two distinct categories) over the average/modal values, would provide strong evidence that the items had been segregated into two distinct categories. As in the Narrow condition of Experiment 1, familiarization stimuli were constructed such that restricted ranges of values were correlated. The networks were then tested with stimuli made up of the average/modal feature values or the previously unseen feature values. Table 2 shows the normalized values defining the stimuli in the Broad and Narrow familiarization phase, and the three test stimuli.

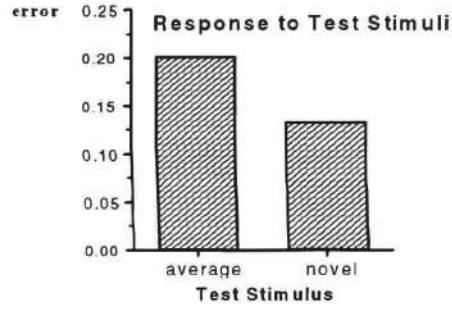

Figure 3. Network response to the average/modal and previously unseen test stimuli.

Figure 3 shows the networks response to the average/modal test stimulus and the previously unseen stimuli. As with 10-month-olds, networks showed more error (longer looking) when presented with the average/modal test stimulus than the stimuli with previously unseen values suggesting that they had formed two distinct categories. A two-way Student t-test revealed this difference was highly significant ($t(23)=6.59$, $p<.001$).

Discussion

This paper presented a model of correlation based categorization by 10-month-old infants. Simple autoencoder networks were exposed to the same stimuli used to test 10-month-olds. The familiarization regime was kept as close as possible to that used with the infants. The model's performance matched that of the infants. Both infants and networks used co-variation information (when available) to segregate items into separate categories.

The model makes the explicit prediction that, in general, looking time to the test stimuli in the Broad condition will be higher than that in the Narrow condition. This can be related to the structure of the internal representations developed by the networks. Encouraging trends that support this prediction can be found in the original Younger (1985) data. Exploration of the model's internal representations suggests that in the Broad condition, looking times are determined by the similarity of the test stimuli to the familiarization stimuli.

This model extends work reported by Mareschal & French (1997) and Mareschal et al. (submitted). It is a

model of category learning within a single test session. It leaves open questions of how this categorization ability develops. In other words, how does

Table 2 Normalized familiarization and test stimuli (Exp 2)

Familiarization Stimuli				
	Legs	Neck	Tail	Ears
	0.27	0.62	1.0	0.56
	0.27	1.0	0.61	0.11
	0.27	1.0	0.61	0.56
	0.64	1.0	1.0	0.11
	0.64	0.62	1.0	0.11
	0.64	0.62	0.22	1.0
	0.64	0.23	0.22	1.0
	1.0	0.23	0.61	0.56
	1.0	0.23	0.61	1.0
	1.0	0.62	0.22	0.56
Test Stimuli				
Average/Modal	0.64	0.62	0.61	0.56
Novel1	0.45	0.81	0.80	0.33
Novel2	0.82	0.42	0.41	0.78

Note: Values are scaled to range from 0.0 to 1.0

the developmental time scale interact with the course of learning during a task? Younger & Cohen (1986) describe a sequence of development from no use of correlation information at 4 months of age to the use abstract invariant relations at 10 months. Future modeling needs to explore how the ability to use correlation information comes about.

The complex relationship between the similarity of test stimuli to familiarization stimuli, and relative looking times can be explored through the model before making further empirical predictions. This illustrates the function of a model as a tool for reasoning about untested contexts. In the same way that a model bridge can help engineers reason about a real bridge, a computer model can help experimental psychologists reason about categorization. However, it is also important to understand that in the same way that a model bridge is never meant to embody all the characteristics of the real bridge, the computer model is not meant to capture all the richness of infant behavior.

We do not wish to claim that simple autoassociator networks can capture the full richness of infant categorization. There is far more to an infant than 11 neurons! This model is intended as an illustration of the computational properties of an associative system with distributed representations. There are other such systems that share many of the same computational properties (e.g. Grossberg, 1980; Knapp & Anderson, 1984).

Connectionism has inherited the Hebbian rather than the Hullian tradition of associative learning. What goes in inside the head is crucial for understanding behavior. Connectionist models force us to think about internal representations, to ask how they interact with each other, and to ask how they determine observed behaviors. We argue that connectionist methods are fruitful tools for exploring perceptual and cognitive development.

Finally, we wish to suggest that the observed infant categorization behaviors are inextricably linked to both the categorization mechanisms internal to the infant, and the properties of the external stimuli shown to the infants during the study. Thus, categorization is the product of an inextricable interaction between the infant and its environment. The computational characteristics of both subject and environment must be considered *in conjunction* to explain the observed behaviors.

References

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 Paiget,* 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.

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.* Norwood, NJ: Ablex.

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

Knapp, A. G. & Anderson, J. A. (1984). Theory of categorization based on distributed memory storage. *JEP:LMC, 10,* 616-637.

Mareschal, D. & French, R. M. (1997). A connectionist account of interference effects in early infant memory and categorization. In *Proc. of the 19th annual conference of the Cognitive Science Society* NJ: LEA,484-489.

Mareschal, D., French, R. M., & Quinn, P. C. (submitted). Interference effects in early infant memory and categorization: A connectionist model.

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.

Quinn, P. C., & Johnson, M. H. (1997). The emergence of perceptual category representations in young infants, *J. of Exp. Child Psychology, 66,* 236-263.

Quinn, P. C., & Eimas, P. D. (1996). Perceptual organization and categorization in young infants. *Advances in infancy research, 10,* 1-36.

Rosch, E. (1975). Cognitive representations of semantic categories. *JEP:General 104,* 192-233.

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

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

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

Younger, B. A. (1990). Infants' detection of correlations among feature categories. *Child Dev., 61,* 614-620.

Younger, B. & Cohen, L. B. (1986). Developmental changes in infants' perception of correlation among attributes. *Child Development, 57,* 803-815.