Mechanisms of Categorization in Infancy

Denis Mareschal
Centre for Brain and Cognitive Development
Birkbeck College

Robert French
Psychology Dept.,
Université de Liege

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. The familiarisation 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 provides a mechanistic account of category learning with a test session. It demonstrates how categorization arises as the product of an inextricable interaction between the subject (the infant) and the environment (the stimuli). The computational characteristics of both subject and environment must be considered <u>in conjunction</u> to understand the observed behaviors.

Mechanisms of Categorization in Infancy

The ability to categorize underlies much of cognition. It is a way of reducing the load on memory and other cognitive processes (Roch, 1975). Because of its fundamental role, any developmental changes in the abilities 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 (e.g., Reznick & Kagan, 1983). 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 familiarisation exemplars. Generalization of familiarisation 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. Evidence that young infants can form categorical representations of shapes, animals, furniture, faces, etc. is discussed throughout this special section of the current issue (see also Quinn and Eimas, 1996 for a current review).

At first, the categories developed by infants appear similar to those developed by adults. However, occasionally, the 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 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 different level exemplars might dictate the order in which categories are acquired across infancy as a whole, it did not directly address the issue of how categories are set up <u>within</u> a short term testing session characteristic of many published categorization studies.

We believe that the way to a comprehensive synthesis of the numerous competence studies that abound in the 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 an elusive catastrophic interference effect in infant memory studies (Mareschal & French, 1997; Mareschal, French, Quinn, submitted).

Mareschal <u>et al</u>. presented connectionist networks with the same cat and dog exemplars used to familiarise infants in the original Quinn <u>et al.</u> (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 familiarisation 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 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 (Fagan, 1973; Deloache, 1976; McCall, Kennedy, & Dodds, 1977, Fagan 1977; Cohen, Deloache, & Pearle, 1977). 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, & Quinn, submitted). In short, the model demonstrated how the previously unrelated exclusivity asymmetry and elusive interference effects were two sides of the same mechanistic coin.

In this paper, we will extend that work by exploring the basis on which 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 begins by describing the connectionist modeling paradigm with particular attention to the autoencoder network used to model infant categorization. Network performance is then described. Next, the networks' internal category representations are described to help explain the networks' behavior. Finally, implications for understanding infant behaviors are discussed.

Connectionist models

Connectionist models are computer models based loosely on the principles of neural information processing (Rumelhart & McClelland, 1986; Hertz, Krogh, & Palmer, 1991). A connectionist network is made of simple processing units connected together via weighted communication lines. Each unit performs a very simple computation. The unit sums the activation arriving into it and takes on some activation level determined by its own activation response function. In general, that response function is non-linear; i.e., the unit's resulting activation level is not just some proportion of the total input.

Units that receive activation from outside the network are called input units, units that send information out of the network are called output units, and all units inside the network are called hidden units. Figure 1b shows one such network that will be discussed in more detail below. Information is encoded as a pattern of activation across some set of units. As information comes into the network, is processed by the network, and leaves the network, the input units first become active, then the hidden units, then the output units respectively. The pattern of activation produced across the hidden units constitutes an internal representation of the information first encoded across the input units.

The behavior of the network is determined by the connection weights between all the units. As the weights change, the behavior changes. Hence, learning consists in adjusting the connection weights in the network. Usually those weights are adjusted gradually (with exposure to an environment) such that the network learns to produce some desired response across the output units when presented with some particular input.

One implication of this process is that connectionist networks develop their own task appropriate internal representations as part of the learning process. This is what makes them ideal systems for modeling development (e.g., Elman, Bates, Johnson, Karmiloff-Smith, & Plunkett, 1997; Mareschal & Shultz, 1996; Plunkett & Sinha, 1992; McClelland, 1989). Initially, a network is constructed with random connection weight values. As the network encounters task exemplars, the weights are slowly tuned to produce meaningful (task appropriate) internal representations across the hidden units.

Building the model

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. This process is illustrated in Figure 1a. During the period of sustained attention, the infants encode the stimulus. That encoding is then compared to an existing internal representation. As long as a discrepancy is found between the contents of the internal representation and the new encoding, the internal representation is adjusted and the cycle repeated.

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 sustained attention, encoding, and representation construction. An autoencoder is a feedforward connectionist network with a single layer of hidden units (Figure 1b). 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. This process is illustrated in Figure 1b. 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 autoencoders (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. A Fahlman offset of 0.1 was also used (Fahlman, 1988). 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.

The data and familiarisation regime.

The two simulations described below are attempts to model the behavior of 10-month-olds as reported by Younger (1985). The network training regime is kept as close as possible to the infant familiarisation 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 (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). 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 the line drawings varied along 4 dimensions, in 5 equal discrete steps. These dimensions were: leg length (ranging from 1.5 to 5.5 cm), tail width (ranging from 2.3 to 0.5 cm), neck length (ranging from 5.2 to 1.2 cm), and eye separation (ranging from 0.3 to 2.7 cm). Readers should refer to Younger (1985) for a more detailed description of these stimuli.

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

In a first experiment, infants in were familiarised with a set of 8 exemplars. Table 1 shows the feature values for all stimuli used. Rather than showing the actual attribute values, Table 1 follows Younger (1985) and lists the rank values of the features making up the stimuli in order to highlight the correlational structure in the stimuli. In one condition (the Broad condition) there were no constraints on the co-occurrence of features; the full range of values on one dimension occurred with the full range of values on the other dimensions. In the second condition (the Narrow condition), feature values were constrained to co-vary; restricted ranges of values were correlated across dimensions. So, for example, values 1 and 2 always co-occurred in any animal, as did values 4 and 5.

Infants were then tested with two types of stimuli: (a) an exemplar whose attribute values were the average of all the previously experienced values along each dimension (i.e., 3 3 3 3; the average stimulus), or (b) an exemplar containing previously experienced feature values along each dimension (i.e., 1 1 1 1 or 5 5 5 5; the modal stimuli). Preference for a modal over 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 familiarisation set was unconstrained suggesting that they had formed a single representation of the complete set of exemplars in this condition. However, the 10-month-olds looked more at the average stimuli when the familiarisation set was constrained such that ranges of feature values were correlated suggesting that they had formed two distinct categories in this condition.

============= Insert Table2 about here==============

In a second experiment, Younger provided a more stringent test of category formation in infancy. In this experiment, the infants were presented with a constrained familiarisation set (i.e., ranges of feature values were correlated across dimensions). However, the familiarisation set was designed such that the modal stimulus was identical to the average stimulus. Infants were then tested with the modal/average stimulus (with familiar attribute values), and two novel stimuli (with previously unseen attribute values) that were prototypical of the two possible categories contained within the familiarisation set. Table 2 shows the feature values for the stimuli used this experiment. 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 stimuli with previously unseen attribute values 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. The actual attribute values were used as opposed to the rank values reported in Tables 1 and 2. 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. Normalisation was achieved by dividing each attribute value by the maximum value along that dimension.

Networks were trained in batch mode. That is, all 8 familiarisation 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 familiarisation 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 familiarisation trials. Batch learning also ensures that all order effects are averaged out.

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 3 shows the normalised values defining the stimuli in the Broad and Narrow familiarisation conditions, and the three test stimuli.

=================== Insert Table 3 about here   =============

Figure 2 shows the networks' response to the average and modal test stimuli when familiarised in either the Narrow or Broad conditions[1]. As with the 10-month-olds, networks familiarised 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 familiarised in the Broad condition showed more error (preferred to look) when presented with the modal test stimuli than the average test stimuli.

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 ($\underline{F}(1,46)=752$, $\underline{p}<.0001$). This interaction was accounted for by a significant effect of Stimulus within the narrow condition ($\underline{F}(1, 23) = 21$, $\underline{p}<.0001$), as well as within the broad condition ($\underline{F}(1, 23) = 1932$, $\underline{p}<.0001$).

=========== Insert Figure 2 about here ================

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 familiarisation items. In this simulation 24 networks were familiarised 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, familiarisation 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 4 shows the normalised values defining the stimuli in the Broad and Narrow familiarisation phase, and the three test stimuli.

=================== Insert table 4 about here =============

Figure 3 shows the networks response to the average/modal test stimulus and the previously unseen stimuli[2]. 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 that this difference was highly significant ($\underline{t}$(23)=4.00, $\underline{p}$<.005).

=========== Insert Figure 3 about here ================

Internal representations

One advantage of computer models is that they can be taken apart to help understand what produces the observed behaviors. This section describes the internal representations developed by the networks and discusses how they lead to the observed preferential looking behaviors described above.

When an exemplar is presented to the network, activation flows from the input units to the hidden units. The pattern of activation across the hidden units is an internal representation of that input. It is the internal representation that drives the response at the output. Every exemplar will produce a different activation pattern across the hidden units. One way to explore these representations is to plot them as points in a 3 dimensional space. For any given input, each of the three hidden units will have some activation value. These three values can be interpreted as co-ordinate values (e.g., x=0.1, y=0.38, z= 0.72) within a 3 dimensional space. Each internal representation (arising from each separate exemplar) corresponds to a point in that space.

From a behavioral perspective, categorization is diagnosed when identifiably different exemplars are treated in the same way. In hidden unit space, members 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 4 shows the distribution of exemplars within the hidden unit space for networks trained in the Narrow and Broad conditions of Simulation 1. In the Narrow condition (Figure 4a), 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.

=============== Insert Figure 4 about here ================

Figure 4b 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 centre of

the space. This is because all three have comparable similarities (in terms of feature values) to all of the familiarisation 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 3 we can see intuitively why the test stimuli have comparable similarities (of the order of 1/2) to all the familiarisation exemplars. Because of the systematic structure of the familiarisation 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 familiarisation 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 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.

A model prediction

One implication of the more diffuse pattern of points in the Broad condition is that the error (looking time) in this condition there will tend to be higher (on average) than in the Narrow condition, irrespective of the nature of the exemplar. There will be relatively more looking at the test stimuli in the Broad condition than in the Narrow condition. This is a concrete model prediction that remains to be tested against actual infant looking times.

Suggestive but inconclusive evidence supporting this prediction can already be gleaned from the original Younger (1985) data. Adding together the total looking times to both the modal and average stimuli (reported on page 1579 in Table 2, column 1, of Younger, 1985) reveals a total looking time of 4.94 sec and 4.36 sec for infants in the Broad and Narrow conditions respectively. Moreover, when compared with a novel stimulus, the looking times to both the average and modal stimuli are greater in the Broad condition than in the Narrow condition. Although there is insufficient variance information reported in the original paper to test the statistical significance of these findings, the trend towards a longer looking times in the Broad condition than in the Narrow condition is encouraging.

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 data presented to the networks was derived from the actual dimensions used to generate the stimuli presented to the infants. The familiarisation 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 familiarisation stimuli.

This model extends the work reported by Mareschal et al. (1997; submitted). It is a model of category learning <u>within</u> a single test session. It leaves open questions of how this categorization ability develops. In other words, how does the developmental time scale interact with the course of learning during a task? Younger and 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 familiarisation 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 autoencoder networks can capture the full richness of infant categorization. There is far more to an infant than 11 neurones! 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; Mareschal & Shultz, 1996). We chose to use autoencoder networks to model infant looking time behaviors because they are simple, well-understood systems whose functioning could be mapped onto the theories underlying representation construction during preferential looking in a straight forward fashion. Future work that attempts to capture other aspects of infant categorisation behaviors may choose to rely on other connectionist architectures.

Connectionism has inherited the Hebbian rather than the Hullian tradition of associative learning. What goes on 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. This model continues to 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 subject (the infant)

and the environment (the stimuli). The computational characteristics of both subject and environment must be considered <u>in conjunction</u> to explain the observed behaviors.

References

Charlesworth, W. R. (1969). The role of surprise in cognitive development. In D. Elkind & J. Flavell (Eds.), <u>Studies in cognitive development. Essays in honor of Jean Paiget</u>, pp. 257-314, Oxford, UK: Oxford University Press.

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

Cohen, L. B. (1979). Concept acquisition in the human infant. <u>Child Development, 50</u>, 419-424.

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

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

Cottrell, G. W., Munro, P., & Zipser, D. (1988). Image compression by backpropagation: an example of extensional programming. In N. E. Sharkey (Ed.), <u>Advances in cognitive science, Vol. 3</u>. Norwood, NJ: Ablex.

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

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

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

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

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

Fahlman, S. E. (1988). Faster-learning variations on back-propagation: An empirical study. In D. S. Touretzky, G. E. Hinton, and T. J. Sejnowski (Eds.), <u>Proceedings of the 1988 Connectionist Models Summer School</u>. Los Altos, CA: Morgan Kaufmann.

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

Hertz, J., Krogh, A. & Palmer, R. G. (1991). <u>Introduction to the theory of neural computation</u>. Redwood city, CA: Addison-Wesley.

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

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

Mareschal, D., & Shultz, T. R. (1996). Generative connectionist networks and constructivist cognitive development. <u>Cognitive Development, 11</u>, 571-603.

Mareschal, D. & French, R. M. (1997). A connectionist account of interference effects in early infant memory and categorization. In M. G. Shafto & P. Langley (Eds.),

Proceedings of the 19th annual conference of the Cognitive Science Society (pp 484-489). Mahwah, NJ: LEA.

Mareschal, D., French, R. M., & Quinn, P. C. (submitted). A connectionist account of asymmetric category learning in infancy.

McClelland, J. L. (1989). Parallel distributed processing: Implication for development and cognition. In R. G. M. Morris (Ed.) Parallel distributed processing: Implications for psychology and neurobiology (pp. 8-45). Oxford, UK: Oxford University Press.

Plunkett, K. & Sinha, C. (1992). Connectionism and developmental theory. British Journal of Developmental Psychology, 10, 209-254.

Pylyshyn, Z. W. (1984). Computation and cognition: Towards a foundation ofr cognitive sceince. Cambridge, MA: MIT press.

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, Journal of Experimental 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.

Reznick, J. S. & Kagan, J. (1983). Category detection in infancy. Advances in Infancy Research, 2, 79-111.

Rosch, E. (1975). Cognitive representations of semantic categories. Journal of Experimental Psychology: 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.

Rumelhart, D. & McClelland, J. (1986). Parallel Distributed Processing, Vol. 1. Cambridge, MA: The MIT Press.

Shultz, T. R. & Mareschal, D. (1997). Rethinking innateness, learning, and constructivism: Connectionist perspectives on development. Cognitive Development, 12, 563- 586.

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

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

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

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

Footnotes

Footnote 1: The modal value shown in Figure 2 is the average of the responses to both modal test stimuli. In the original study (Younger, 1985) infants were presented with only one of the two randomly selected modal test stimuli.

Footnote 2: The "previously unseen" value shown in Figure 3 is the average of the responses to both "previously unseen" modal test stimuli. In the original study (Younger, 1985) infants were presented with only one of the two randomly selected modal test stimuli.

Figures

Figure 1. Learning via iterative representation adjustment in (a) infants, and (b) connectionist autoencoder networks.

Figure 2. Responses to the average and modal test stimuli for networks familiarised in the Broad and Narrow conditions. Standard-error bars are also plotted.

Figure 3. Network response to the average/modal and previously unseen test stimuli. Standard-error bars are also plotted.

Figure 4. Exemplar distribution in hidden unit space for networks familiarised in the (a) Narrow, and (b) Broad conditions.

Table 1. Rank Order Values of Familiarization and Test Stimuli for Experiment 1

Familiarization Stimuli							
Broad Condition				Narrow Condition			
Legs	Neck	Tail	Ears	Legs	Neck	Tail	Ears
1	1	5	5	1	1	2	2
1	5	1	5	1	2	1	2
2	2	4	4	2	2	1	1
2	4	2	4	2	1	2	1
4	4	2	2	4	4	5	5
4	2	4	2	4	5	4	5
5	5	1	1	5	5	4	4
5	1	5	1	5	4	5	4
Test Stimuli							
Average	3	3	3	3			
Modal1	1	1	1	1			
Modal2	5	5	5	5			

Note: The integers 1 to 5 represent the incremental rank of feature values along each dimension. E.g., for Leg length, 1=1.5cm, 2=2.5cm, 3=3.5cm, 4=4.5cm, and 5=5.5cm.

Table 2. Rank Order Values of Familiarization and Test Stimuli for Experiment 2

	Familiarization Stimuli			
	Legs	Neck	Tail	Ears
	1	3	1	3
	1	1	3	1
	1	1	3	3
	3	1	1	1
	3	3	1	1
	3	3	5	5
	3	5	5	5
	5	5	3	3
	5	5	3	5
	5	3	5	3
Test Stimuli				
Average/Modal	3	3	3	3
Novel1	2	2	2	2
Novel2	4	4	4	4

Note: The integers 1 to 5 represent the incremental rank of feature values along each dimension. E.g., for Leg length, 1=1.5cm, 2=2.5cm, 3=3.5cm, 4=4.5cm, and 5=5.5cm.

Table 3. Normalised Familiarisation and Test stimuli for Experiment 1

Familiarisation 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	0.81	1.0	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.

Table 4. Normalised Familiarisation and Test stimuli for Experiment 2

	Familiarisation Stimuli			
	Legs	Neck	Tail	Ears
	0.27	0.62	1.0	0.56
	0.27	0.23	0.61	0.11
	0.27	0.23	0.61	0.56
	0.64	0.23	1.0	0.11
	0.64	0.62	1.0	0.11
	0.64	0.62	0.22	1.0
	0.64	1.0	0.22	1.0
	1.0	1.0	0.61	0.56
	1.0	1.0	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	0.42	0.80	0.33
Novel2	0.82	0.81	0.41	0.78

Note: Values are scaled to range from 0.0 to 1.0

Response to Test Stimuli

Response to Test Stimuli

(a)

(b)