INFANCY, *2*(1), 7–28

RESEARCH ARTICLES

Modeling Infant Speech Sound Discrimination Using Simple Associative Networks

Graham Schafer
Department of Psychology
University of Reading

Denis Mareschal
Centre for Brain and Cognitive Development
Birkbeck College

Infants' responses in speech sound discrimination tasks can be nonmonotonic over time. Stager and Werker (1997) reported such data in a bimodal habituation task. In this task, 8-month-old infants were capable of discriminations that involved minimal contrast pairs, whereas 14-month-old infants were not. It was argued that the older infants' attenuated performance was linked to their processing of the stimuli for meaning. The authors suggested that these data are diagnostic of a qualitative shift in infant cognition. We describe an associative connectionist model showing a similar decrement in discrimination without any qualitative shift in processing. The model suggests that responses to phonemic contrasts may be a nonmonotonic function of experience with language. The implications of this idea are discussed. The model also provides a formal framework for studying habituation–dishabituation behaviors in infancy.

In the course of the 20th century, scholars of cognitive and linguistic development amassed a wealth of knowledge about the young child's evolving repertoire of communicative behaviors (e.g., Jakobson, 1941/1968; Stern & Stern, 1907; see

Requests for reprints should be sent to Graham Schafer, Department of Psychology, University of Reading, Earley Gate, Reading RG6 6AL, UK. E-mail: g.w.schafer@reading.ac.uk

Bar-Adon & Leopold, 1971, for a collection of historic readings and Fletcher & MacWhinney, 1995, and Barrett, 1999, for more recent developments). Young children's cognitive and linguistic development often appears to proceed in stages. Performance in a given task is rarely a monotonic slope from outright failure to complete success. For example, from early in life until the age of about 8 months, infants respond in a categorical fashion to phonemic contrasts—even those not appearing in their native language. On the other hand, older infants and adults find such non-native contrasts difficult to detect (Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995; Trehub, 1976; Werker & Lalonde, 1988; Werker & Tees, 1983, 1984a; but see Best, McRoberts, & Sithole, 1988).

It is tempting to interpret such qualitative shifts in behavior as evidence of qualitative shifts in the way in which the developing infant is processing the information in a given task. For example, Werker and Pegg (1992) argued that the changes in infants' performance in speech-sound discrimination tasks are diagnostic of at least three, and possibly four, distinct stages in infants' speech processing. Such stages are hypothesized to reveal a functional reorganization over time: A stimulus arriving at Time 1 will be processed in a particular fashion, and the same stimulus arriving at Time 2 will be processed in a different fashion. A variant of this argument was presented by Stager and Werker (1997) in their report of a series of experiments conducted with 8- and 14-month-old infants (see also Werker, Cohen, Lloyd, Casasola, & Stager, 1998).

In this article, we begin by describing Stager and Werker's (1997) findings. We then suggest an alternative account of these data, in which a change in infants' behavior does not require a change in processing between Time 1 and Time 2. We consider whether changes in infants' ability to discriminate between particular stimuli necessarily signals a change in the underlying processing of those stimuli; and we introduce a method for the qualitative evaluation of infant habituation data using neural networks.

Stager and Werker (1997) used a bimodal habituation task to investigate the possible relation between word learning and speech-sound discrimination. In this task, infants are habituated to images and sounds presented at the same time. The general outline of the task is as follows: An infant is habituated to a bimodal stimulus pair presented simultaneously in auditory and visual modalities. Stager and Werker argued that such a task invokes mechanisms that subserve the learning of words; that is, learning that a given label (sound) goes with a given object (image). Following habituation, a change is made to the sound but not to the image. The extent to which the infant now dishabituates to this new sound–image combination is an index of the specificity of the binding between the previously habituated sound and the (unchanged) image. Infants who have habituated to a given sound–image combination will dishabituate only if they perceive the difference between the sound heard during the habituation phase and the sound heard during the subsequent testing phase. This difference in habituation is operationalized as a difference in looking times between trials in which the stimulus has changed (*switch* trials) and trials in which it has stayed the same (*same* trials).

Stager and Werker's (1997) data are shown in Figure 1. Within the context of label–object associative learning, 8-month-olds discriminate the label [bɪh] from the label [dɪh] (Figure 1, Experiment 2, left-hand pair). Somewhat surprisingly, 14-month-olds appear not to do so (Figure 1, Experiment 2, right-hand pair); however, the older infants can discriminate between a more distinct pair of labels such as [lɪf] and [nim] (Figure 1, Experiment 3); these 14-month-old infants can also discriminate [bɪh] from [dɪh] in a simple auditory discrimination task (Figure 1, Experiment 4). Furthermore, the 14-month-olds were not capable of discriminating [bɪh] from [dɪh] when the task involved learning about two label–object tokens (i.e., [bɪh] + object 1 and [dɪh] + object 2; Figure 1, Experiment 1). What are we to conclude from such data? Stager and Werker argued that, taken together, these data suggest that a functional reorganization of the language system occurs between the ages of 8 and 14 months. As a consequence of this reorganization, infants of different ages react differently to identical stimuli. Younger infants have not, it is argued, been provoked into a word-learning stance by these bimodal stimuli. Older infants, on the other hand, respond to the same input by trying to map the sounds they hear onto meaning (i.e., they treat the new sounds as to-be-learned words). This different stance to the same stimuli results in the two groups performing different processes, and different behaviors are therefore observed in the two

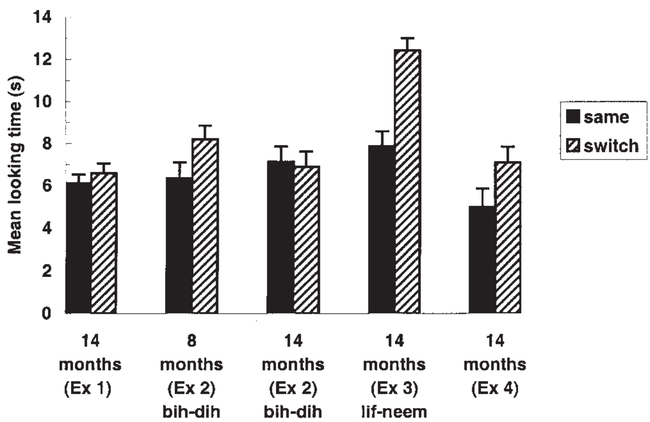

FIGURE 1 Stager and Werker (1997) infant data. Experiment 1: A habituation–dishabituation task with two label–object pairs. Experiment 2: A habituation–dishabituation task with a single label–object pair. Experiment 3: As Experiment 2, but with dissimilar labels. Experiment 4: Simple auditory discrimination. Data adapted with permission from *Nature, 388*: 381–382 (1997). Copyright 1997, Macmillan Magazines, Ltd.

groups.[1] Thus, 14-month-olds listen for less phonetic detail during the bimodal habituation task than do 8-month-olds, presumably because the older infants are processing the stimuli differently (Stager & Werker, 1997, p. 382). In terms of Werker and Pegg's (1992) four-stage model of phonological development, the younger infants in the Stager and Werker study are in the process of refining their phonetic code to that of the ambient language, presumably by establishing that certain sounds are much more commonly heard than others; in contrast, the older infants are performing phonemic processing; that is, learning about the sounds that designate meanings. According to this account, the ability to discriminate a minimal pair of speech sounds ([bɪh]–[dɪh]) that is observed at 8 months is not subsequently observed at 14 months because of a reorganization of the infant's perceptual system. During this time, the infant has changed his or her stance to such bimodal stimuli. This reorganization, according to Werker and colleagues, serves to bias the perceptual system so that in two identical tasks, younger infants will detect a fine phonetic distinction, whereas older infants, being biased toward learning words, fail to make the same fine phonetic distinction. According to this view, the processing system of older infants differs radically from that of younger infants.

In making their argument for a change from speech-sound discrimination processing to word-learning processing, Stager and Werker (1997) were careful to exclude an additional possible explanation of the findings from Experiments 2 and 3. It is logically possible that the 14-month-olds failed the fine ([bɪh]–[dɪh]) discrimination in the bimodal habituation task not because they were processing for word meaning, but simply because they were incapable of discriminating the auditory stimuli. To control for this possibility, Stager and Werker performed a final study (Experiment 4) in which 14-month-olds were presented with a [bɪh]–[dɪh] discrimination in the presence of a black-and-white checkerboard rather than an object. Such an approach has commonly been used to study speech-sound discrimination; it is usually argued that the checkerboard does not constitute an "object." In this final study, the 14-month-olds were able to perform the discrimination: They showed increased looking during switch trials, even with the fine [bɪh]–[dɪh] discrimination. The authors argued that this further strengthened the possibility that infants at 14 months were processing the stimuli in Experiments 2 and 3 in a qualitatively different fashion from the way in which the 8-month-olds were processing the same stimuli.

[1]Throughout this article, the word *task* is confined to the stimuli presented to (and responses recorded from) our subjects (human or artificial). *Processing* and *process* refer to any algorithm that may be inferred to be carried out within the learner (human or artificial). Thus, if a learner decides that a particular stimulus pair should be treated at Time 1 as a piece of evidence about sound patterns in English and at Time 2 as a piece of evidence about word meaning, this would constitute a shift in processing. It would not, in our terminology, constitute a shift in task, because stimuli and dependent variable are fixed between Time 1 and Time 2 by decision of the experimenter.

The relation between a change in behavior and a change in mechanism, however, is not straightforward (Brainerd, 1978; Ingram, 1989; Kemler-Nelson, 1984). In particular, it is incorrect to assume that a change in observed behavior signals a change in underlying processing or that qualitative shifts in performance imply qualitative shifts in processing (e.g., Elman et al., 1996; Thelen & Smith, 1993). For example, connectionist networks are simple associative systems that can show abrupt shifts in behavior as a consequence of the continuous adaptation of a single processing system (Elman et al., 1996; Mareschal & Shultz, 1996; McClelland, 1995; Plunkett, Sinha, Møller, & Strandsby, 1992). Moreover, connectionist networks have recently been used to model early infant behavior across a range of different domains (Mareschal & French, 1997, 2000; Mareschal, Plunkett, & Harris, 1999; Munakata, 1998; Munakata, McClelland, Johnson, & Siegler, 1997; Quinn & Johnson, 1997). This article extends such work by exploring whether simple associative systems, whose adaptive properties do not change over time, can account for the apparent discontinuity in processing reported by Stager and Werker (1997).

In this article, we show how a simple learning device, an associative neural network, changes in its response to external stimuli as a result of experience. Our goals are threefold. First, we wish to suggest that behavior that changes over time does not necessarily signal a shift in underlying mechanism (or strategy) on the part of a learner. Second, we present a method for the modeling of habituation phenomena in infants. Third, we make some predictions based on the behavior of the model in the tasks we have modeled. We begin by describing a framework for modeling habituation–dishabituation studies using connectionist networks. This is followed by a description of an autoencoder model of infant performance in a homologue of the bimodal habituation–dishabituation task used by Stager and Werker (1997). Finally, the model's performance is evaluated and implications for understanding infant behavior are discussed.

MODELING RELEASE FROM HABITUATION

Habituation techniques are based on the assumption that infants direct more attention to unfamiliar or unexpected stimuli than to familiar or expected stimuli. The standard interpretation of this behavior is that infants are comparing an input stimulus to an internal representation of that 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 his or her internal representation. When the information in the internal representation is no longer discrepant with the visual input, attention is directed elsewhere. Thus, 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 an elevation in the amount of attending, because a new internal representation has to be constructed or an existing representation adjusted. The degree to which a novel object differs from the set of existing internal representations determines the amount of adjustment and, therefore, the duration of attention.

We used a connectionist autoencoder to model the relation between attention and representation construction (Mareschal & French, 1997, 2000). 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 (hence the word *autoencoder*). If there are fewer hidden units than either input or output units, there is a bottleneck in the flow of information through the network. In such an autoencoder, learning consists in developing, at the hidden unit level, a compacted internal representation of the input that reliably contains relevant information from the original input. Information is compressed into an internal representation and then expanded to reproduce the original input. The successive cycles of training in the autoencoder are the iterative process by which a reliable internal representation of the input is developed. The reliability of such a representation is tested by expanding it and comparing the resulting predictions to the actual stimulus presented. The difference between the observed output and the actual stimulus presented is termed *network error*.

This modeling approach has three implications:

1. Provided that the representations encoded by the network bear some relation to those encoded by the infant, looking times for infants will be positively correlated with error values for networks: The greater the error, the longer the looking time.
2. Prolonged exposure after looking (error) has reached an asymptote will not further improve memory of the stimulus in an infant (or network).
3. The degree to which looking time (error) increases on presentation of a novel stimulus will be governed by the similarity between the novel stimulus and the familiar stimulus.

On the basis of these predictions, presenting a series of similar stimuli will lead to a progressive reduction in looking time (error), both for the particular stimuli presented and for stimuli that are novel but bear some perceptual resemblance to those already presented. This holds both for infants (where looking time is the measured variable) and for autoencoders (where output error is the measured variable). Thus, throughout the treatment offered in what follows, network error can be thought of as analogous to looking time in the behavioral experiments.

THE MODEL[2]

To model infant performance, connectionist networks were given experience of an artificial "language" and then tested on their ability to learn specific label–object associations. This approach gives rise to a two-stage procedure, in which a language-exposure phase is followed by an experimental phase. The experimental phase is further subdivided into two stages: habituation and subsequent testing.

Networks were taught to autoencode labels and objects, in a homologue of looking and listening by the infant. Three-layer networks were trained to reproduce on their output units the label–object pairs that had been presented at the input (Figure 2). This task requires the networks to develop an internal representation across the hidden unit layer, merging the information from these two sources of information (e.g., Chauvin, 1988; Plunkett et al., 1992). Networks had 36 linear input units, 18 hidden units, and 36 output units. The activation function of the hidden and output units was the commonly used logistic ("squashing") function.

Language exposure was modeled as follows: All the networks were trained to autoassociate the same randomly generated bank of 240 label–object pairs. Networks were presented with successive label–object pairs, randomly selected from the bank of 240 available pairs. After each label–object pair presentation, connection weights were updated using the backpropagation learning algorithm (Rumelhart, Hinton, & Williams, 1986). After each language exposure trial, the label–object pair was returned to the bank and another pair was selected at random. To reflect the differential language exposure of the 8- and 14-month-old infants, "older" networks received more language exposure trials before testing than did "younger" networks.

The Coding Scheme

The artificial language to which networks were exposed was created in the following manner: Labels were represented as consonant–vowel–consonant (CVC) strings, with each phoneme represented as six binary bits (cf. Plunkett & Marchman, 1991). The six bits in each phoneme represented the following features: consonantal (one bit), voiced (one bit), manner (two bits), place (two bits). Label input vectors were generated by randomly selecting a consonant, then a vowel, and then a consonant, from a list of 20 consonants and 12 vowels.[3] There are

[2]The networks were simulated using the *tlearn*© program (Plunkett & Elman, 1997), available at http://crl.ucsd.edu/innate. Data files and more details about the simulations are available at http://www.rdg.ac.uk/~sxs97gws/discrim.html.

[3]It should be emphasized that although this coding scheme is based on phonemes, it can be thought of as representing any nonarbitrary feature of words, for example, phones. Thus, our approach is not limited to modeling processes explicitly confined to learning about phonemes.

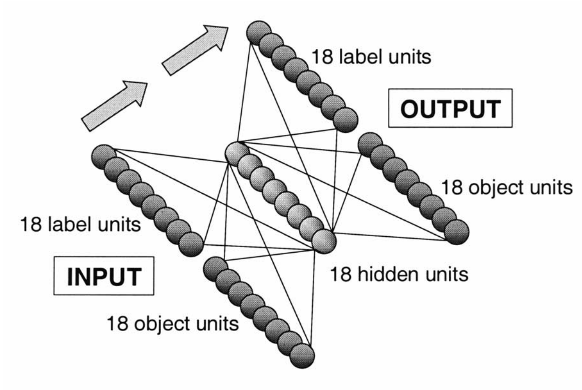

FIGURE 2 Network architecture.

4,800 ways to make this selection (20 × 12 × 20 = 4,800). Our artificial language comprised 240 labels: The input vocabulary therefore consisted of a 5% random sampling of the available CVC space. Each label was then associated with a different object (image, or referent). Within the context of modeling Stager and Werker's (1997) task, the visual stimulus (object) is of no intrinsic interest except inasmuch as it provides a certain amount of information to be bound with specific labels. (The critical manipulation made by Stager & Werker concerned posthabituation shifts in the auditory stimulus only.) Objects were therefore also coded as 18-bit binary vectors, of the same degree of complexity as—and equivalent similarity structure to—the labels. Object input vectors were generated by duplicating the list of 240 18-bit label vectors, shuffling this list, and assigning each of the resultant randomly ordered object vectors to a label vector.

The Training Regime

The networks were trained[4] according to a two-stage procedure. An initial language-exposure phase was followed by an experimental phase. Networks were initially exposed to a linguistic environment in which label–object pairs were successively presented to the network on a predetermined, fixed number of occasions, reflecting the "age" (language exposure) of the network at testing. Younger networks received 1,000 trials; older networks received 10,000 trials.

Following this language-exposure process, the experimental phase began, and networks were habituated to a label–object pair. Finally, after habituation, the net-

[4]Networks were trained using the backpropagation learning algorithm (Rumelhart et al., 1986), with the following parameter values: learning rate = 0.3, momentum = 0.15. Similar results are obtained with a wide range of parameter values.

works were tested by presenting them with a dishabituation stimulus and observing the resultant error. As outlined earlier, the network error in the face of a posthabituation stimulus is directly analogous to the duration of looking observed when an infant is confronted with the equivalent combination of stimuli.

Networks were tested on homologues of each of Stager and Werker's (1997) four experiments. Stager and Werker's Experiments 2 and 3 provide a sort of canonical form for this procedure and are therefore described first. During the habituation phase, a network was presented with 100 habituation trials. Each habituation trial used the same label–object pair (e.g., [bɪh] plus a corresponding object). During the dishabituation phase, the label segment of the input vector was replaced by the to-be-tested label (e.g., [dɪh]). Thus, in the dishabituation phase the network was presented with a familiar object but a novel label, as had been the case with the infants. Following Stager and Werker, we refer to this as a *switch trial*. In contrast, during a *same trial*, the label–object pair presented was the same as that used during habituation.

Minor modifications allowed this procedure to be used for modeling Stager and Werker's (1997) Experiments 1 and 4. In Experiment 1, two different label–object associations had been presented to each infant during the habituation phase, in contrast with Experiments 2 and 3, in which habituation was confined to the repeated presentation of a single label–object pair. In Experiment 4, infants of 14 months had been presented with the [bɪh]–[dɪh] auditory discrimination in the presence of a checkerboard (i.e., a nonmeaningful referent), rather than in the presence of an image of an object (i.e., a meaningful referent). To model these experiments, we adapted the procedure as follows: To model Stager and Werker's Experiment 1, two label–object pairs were used in the habituation phase. In each habituation trial, one of these two pairs was selected at random to be presented to the network. To model Stager and Werker's Experiment 4, all input bits coding image information were set to 0.5 (midway between the 0 and 1 binary values used to encode object information), thereby conveying no information in object feature space. A vector comprising a string of 0.5 values can be likened to an average of all the features of objects the network has previously seen (cf. Shultz, 1998). These modifications correspond to analogous modifications in the procedure used by Stager and Werker for testing infants in Experiments 1 and 4.

There were 20 networks in each experimental group, all with different initial connection weights. These were randomly set at the outset to values between –0.5 and 0.5, using a homogenous distribution.

RESULTS

Figure 3 illustrates the network performance. A comparison of Figures 1 and 3 reveals a striking similarity between the pattern of performance obtained with networks and that obtained with infants. In the data from Experiments 2 and 3, older

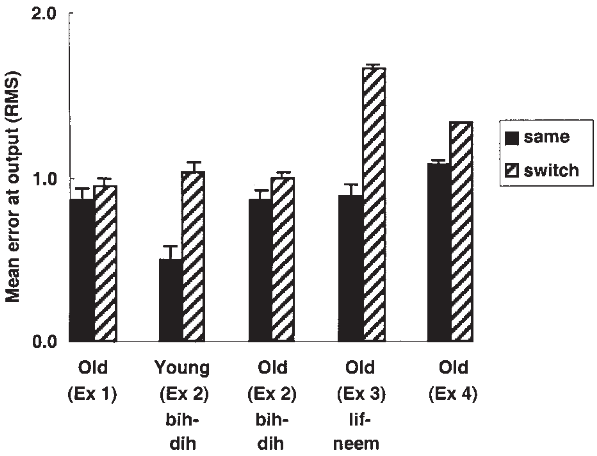

FIGURE 3 Results of network simulations of Experiments 1 to 4.

networks showed poorer discrimination of the similar pair ([bɪh]–[dɪh]) than did the younger networks; the older networks were nonetheless able to discriminate the more distinct pair ([lɪf]–[nim]). It is also important that the model captures an additional aspect of the Stager and Werker (1997) data. In Experiment 4, Stager and Werker presented 14-month-olds with a checkerboard as the visual stimulus, rather than with an object as in Experiments 2 and 3. The purpose of this experiment was to rule out the possibility that the older infants were failing the behavioral task because they were incapable of discriminating [bɪh] from [dɪh] in any circumstances. Infants did indeed discriminate the two stimuli [bɪh] + checkerboard and [dɪh] + checkerboard. To model this, networks were given a habituation task in which the object vector bits were all set to 0.5. Figure 3 shows that the older networks were able to discriminate [bɪh] + checkerboard from [dɪh] + checkerboard, as were the infants.

To further investigate the role of experience in the responses of the networks, we compared networks of a range of “ages” (i.e., differing degrees of language exposure) for their relative release from habituation. An index of relative novelty preference was computed for each network as follows:

$$\text{Novelty} = (\text{Error}_T - \text{Error}_H)/(\text{Error}_T + \text{Error}_H) \tag{1}$$

where Error_T is output error when the network is presented with the test stimulus and Error_H is output error when the network is presented with the habituated stimulus. This function can take values between 0% and 100%. The measure (shown in Figure 4) is a normalized version of the difference in pairs of error scores shown in

Figure 3, and it constitutes the Weber fraction for the given discrimination. That is, if, as argued earlier, error score is analogous to looking time, or attention, then the index of novelty in Equation 1 constitutes a measure of the amount of looking devoted to the difference between a pair of stimuli, normalized for the amount of looking engendered by the stimuli themselves.

Each point in Figure 4 represents an average of the scores of 20 different networks. In total, 400 networks were investigated to address this developmental issue. No initial network configuration (i.e., initial random setting of connection weights) was used more than once. Two aspects of the data are especially worth noting, because they constitute predictions of infant behavior.

First, novelty preference is nonmonotonic with age. For both similar ([bɪh]–dɪh]) and dissimilar ([lɪf]–[nim]) pairs, novelty preference exhibits two minimums in the range of language exposure evaluated. One minimum occurs at around 1,000 language exposure trials. Overall, novelty preference reaches a minimum at around 10,000 language exposure trials then increases again with further language exposure. This sort of nonmonotonicity is reminiscent of human behavior in the detection of non-native speech contrasts. Young infants are initially able to make these distinctions but lose this ability at some point before their first birthday (Werker & Tees, 1983, 1984a); nonetheless, adults are, in certain circumstances, able to make these distinctions (Werker & Tees, 1984b).

Second, release from habituation follows a different time course for the two types of stimulus pair. There is a period during which release from habituation will occur for dissimilar pairs but not for similar pairs. Many accounts of habituation assume the presence of a response mechanism in which a behavioral response is

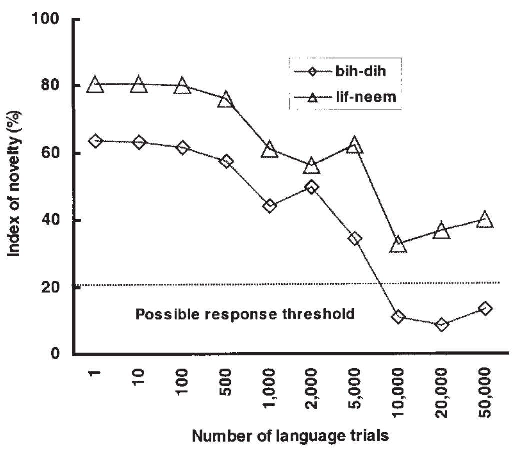

FIGURE 4 Development of novelty preference with increasing amounts of language exposure.

only observed once the novelty preference exceeds some threshold level (cf. Bornstein, 1985; Lamb & Bornstein, 1987). Within the context of the model, if we suppose that a behavioral response is observed only when novelty preference exceeds a 20% threshold, then younger, less experienced networks would manifest a sensitivity to differences between both pairs, whereas older networks would only appear to detect the difference in the dissimilar pair. We suggest that a similar mechanism may explain the difference between 8- and 14-month-olds' behavior in Stager and Werker's (1997) Experiments 2 and 3.

How can the developmental profiles of the novelty preference be explained? The pattern of behavior underlying network performance is instructive here. The behavior of the networks at test is governed by two factors: (a) the way in which the network encodes the habituation stimulus during the habituation phase, and (b) the inherent difference between the habituation and posthabituation stimuli. The first factor is mediated by the network's previous experience with language. Connectionist networks extract the statistical regularities of the environments they are placed in, so that the representations of linguistic knowledge in the networks (in the form of connection weights) are continuously evolving in response to increasing linguistic exposure. The second factor is a function of the (a priori) similarity between the particular stimuli used. Thus, [bɪh] and [dɪh] are minimally distinct, differing only in a single bit (which encodes place), whereas [lɪf] and [nim] differ in five bits (which encode place, manner, and voicing distinctions).

Table 1 shows the number of networks, out of 20, successfully habituating (global $RMS < 0.25$) to the stimuli during the habituation phase of 100 training trials. Like the novelty preference data in Figure 4, the number of networks ha-

TABLE 1
Number of Networks Successfully Habituating to the Training Stimulus in 100 Trials (Maximum = 20)

	Habituation Regime (Labels Used Pre- and Posthabituation)	
Number of Previous Language-Exposure Trials	*Bih–Dih*	*Lif–Neem*
0	20	20
10	20	20
100	20	20
500	18	18
1,000	13	13
2,000	16	12
5,000	14	16
10,000	4	3
20,000	2	3
50,000	1	4

bituating is nonmonotonic with time. Early in development, all networks—irrespective of their start state—learn the habituation stimulus. Later, only a proportion of networks successfully learn the habituation stimulus. This consequence of language exposure drives the shape of the curves in Figure 4 (particularly in the central portion of the figure, between about 500 and 10,000 language trials). These developmental profiles arise as an interaction between the differential language experience of the young and old networks and the computational requirements of the two test conditions (i.e., the [bɪh]–[dɪh] or [lɪf]–[nim] habituation tasks). The networks' responses to the habituation or posthabituation regime will evolve, and this evolution is in turn dependent on the inherent difference between the habituation and posthabituation stimuli. This issue is considered further in the Discussion.

DISCUSSION

In this article we have presented a simple connectionist autoencoder model of infant speech-sound discrimination. The model captures the infant habituation and dishabituation data reported by Stager and Werker (1997) while assuming only a single processing mechanism throughout development. Although the model provides a very close fit to actual infant behaviors, we would not wish to argue that infant cognition can be reduced to simple autoencoder mechanisms. This model is a first attempt to show how associative learning mechanisms, operating in conjunction with the construction of compressed internal representations that encode both visual and auditory information, may account for infant behaviors. As with all models, we have made a number of simplifying assumptions. Nonetheless, the model has explanatory value (in the sense that it provides a causal, mechanistic account of how structure in the input stimuli causes the observed behavior) and predictive value (in the sense that it can be used to make explicit predictions of novel infant behaviors). We discuss the simplifications and predictions of the model here.

Model Simplifications and Limitations

Qualitative changes in behavior can occur in an artificial learner (a simple associative neural network) without the need for qualitative changes in the algorithm instantiated in that learner. Of course, this does not necessarily mean that the same (or even a similar) process accounts for infant development. Nonetheless, it suggests that a cautious observer of infant behavior would be well advised to consider the possibility that this is how infants are doing what they are doing, given the simplicity of the algorithm implemented here. Such an approach necessarily entails the use of simplifying assumptions. We discuss these here.

Selection of arbitrary parameters. We selected a 20% criterion when comparing the responses of networks to [bɪh]–[dɪh] and [lɪf]–[nim] distinctions. Although this figure is not unrepresentative of human infants, it is nonetheless an arbitrary one. Similar observations might apply to the selection of learning and momentum parameters and to our decision to consider 1,000 language trials as representative of young networks and 10,000 trials as representative of older ones. In this last case, inspection of Figure 4 suggests that although we might have suggested other cross-sectional points, such as 500 and 50,000, rather than 1,000 and 10,000, and still been able to draw the same conclusions, there are, in fact, many comparisons (e.g., 1,000 and 10,000 language trials with a 45% criterion) which would not have worked.

Although these are valid caveats, two points need to be made. First, the model stands as an existence proof of an alternative explanation of a set of data. As such, the arbitrary selection of parameters is justified, so long as these are not obviously unrepresentative of the situation we wish to model. Second, the broad phenomenon we have outlined—the relative preservation of one contrast over another—is a robust effect found over a wide range of parameter values, especially for learning rate and momentum, but also for the ages of the young and old networks and for the threshold criterion adopted. In particular, we stress that the principles illustrated here—differential ages of acquisition for different contrasts, and nonmonotonicity in a single processing algorithm—are not dependent on the selection of any particular value for any of the free parameters in the model.

Use of binary feature vectors, specified a priori. Our networks are not exposed to real speech sounds. They “hear” (experience) binary vectors representing a phonological feature description. We have built in the phonemic feature description as part of our input to the model. But how does the infant know what constitutes a phoneme (or phone)? At one level, this question represents an important general critique of the connectionist enterprise, which has been conducted extensively elsewhere (see Fodor & Pylyshyn, 1988). Note, however, that we are not presenting a complete model of phonemic development. Although the model has been expressed in terms of the representation of phonemic features, the coding system is largely arbitrary. Indeed, in the case of the object vectors, it is wholly arbitrary. In the case of the label vectors, as we have already stated, the input vector bits may be considered to be representing phones just as much as phonemes. Even as phonemes, the vectors selected represent a random selection of CVCs rather than an accurate sampling of early input to infants. The phonemic feature description (Plunkett & Marchman, 1991) is capable of improvement to better capture the similarity structure in English phonology (K. Plunkett, personal communication, July 1998); however, rather than reduce the power of the framework we have presented, we believe that adoption of this encoding scheme enhances the generality of our findings.

Identification of looking time with perceived novelty. In the earlier section "Modeling Release From Habituation," we argued that network error was analogous to infants' looking time. This argument is based on the assumption that the more novel the stimulus, the more infants will be inclined to look at it. This may be an oversimplification. Hunter and colleagues (Hunter & Ames, 1988; Hunter, Ames, & Koopman, 1983) have argued that infants progress from showing no preference, to a preference for familiar stimuli, to a declining level of interest in the familiar, and ultimately to a novelty preference of the sort described in the preceding paragraph. According to this account, whether an infant in a standard habituation–dishabituation task will prefer the novel or the familiar stimulus depends on three factors: (a) the amount of time the infant has had to learn about the habituation stimulus, (b) the age of the infant, and (c) the complexity of the stimulus presented. For example, older infants are known to habituate faster than younger infants, and infants of a given age will habituate faster to simpler stimuli (Cohen, 1969; Hunter et al., 1983); however, according to Hunter and Ames (1988), in the limit, infants of all ages will ultimately show a novelty preference to stimuli. As a consequence, given sufficient time to habituate, infants will behave according to Sokolov's (1963) model as discussed earlier. We therefore make the same simplifying assumption as Stager and Werker (1997) did and assume that infants are showing a novelty preference (in accordance with the Sokolov account of habituation) when interpreting the infant looking behaviors.

Departure of the model from the behavior of older infants. Although 14-month-olds are not capable of the fine [bɪh]–[dɪh] discrimination in a bimodal context (Stager & Werker, 1997, Experiment 2), it appears that 20-month-old infants are capable of performing this discrimination (Werker, Corcoran, & Stager, 1999). This apparent recovery in the ability to map minimally distinct novel words to referents is consistent with earlier findings, such as that of Garnica (1973). Indeed, minimally contrastive pairs are usually defined in terms of adults' ability to use such mappings. Yet by our account, fine discriminatory ability remains below the hypothetical 20% threshold. In other words, our account of the learning mechanism cannot account for more mature performance.[5] One possibility, therefore, is that our model may be applicable only to younger infants—say, to infants in the sensorimotor period (Piaget, 1954). Another, related, possibility is that other mechanisms are involved in determining the responses of 20-month-old infants and of adults.

Another, more interesting—but speculative—suggestion is that the hypothesized 20% novelty preference threshold does not remain constant with age. If the

[5]Although the networks do appear to be on the way back up at 50,000 trials, these values of novelty index represent asymptotes. The final position of the asymptote depends on the language experienced by the network up to that point and on the similarity of the discrimination stimuli.

horizontal line in Figure 4 were sloped, perhaps from a relatively high value to a relatively low one, it would be possible to encompass all the known data in a single process. Young infants and adults would evince “switch” effects that rose above threshold; older infants (at an intermediate period of exposure and whose category structure was consequently in a transitional state) would not detect the effects of a switch. Although hardly parsimonious, this explanation makes a sort of intuitive sense: Adults, or experts in a cognitive domain, tend to be more sensitive to small distinctions than do infants, or novices. A gradually reducing function for the critical value of the Weber fraction for detection of a categorical change is an attractive idea; it remains to be explored further in this context.

A second limitation in our account of habituation behavior is that, in general, infants tend to habituate faster as they get older (Bornstein, 1985) whereas the older the network, the less likely it is to habituate to the test stimulus (Table 1). Data on the development of infants’ habituation, however, relate predominantly to exclusively visual habituation; it is not clear that these data can be generalized to bimodal presentations. Indeed, it has been explicitly argued that infants do not habituate to such bimodal stimuli (Hirsh-Pasek & Golinkoff, 1996). Although this is clearly not the case in the experiments under discussion—our approach and Stager and Werker’s (1997) findings are both predicated on the notion that infants do habituate in such circumstances—there is clearly some doubt about the application of unimodal habituation to the bimodal situation. More research is required in this area; until such data are available, the performance of the networks in this study constitutes a prediction of infant behavior. We discuss this further in the next section.

A third issue might be the adoption of 100 learning trials as the duration of the habituation phase. Is this not rather a lot of trials? Would it not be more representative to have trained to a criterion anyway? In our view, the absolute number of trials is not important. Similar results can be obtained with fewer habituation trials, but this is not in itself the issue. We used networks to implement a particular theory of perceptual learning in infants (i.e., classical conditioning; Rescorla & Wagner, 1972). What is important is the outcome of such learning, not the trial-by-trial behavior en route to that outcome. As to the issue of fixed versus criterial training, both methods are used extensively in the cognitive development community (Bornstein, 1985; Kellman & Arterberry, 1998); however, when modified to habituate to a range of percentages of initial looking time (error), the key findings of this study (nonmonotonicities and differential release from habituation) remain robust.

Model Predictions

Models have two principal functions: The first is to provide a coherent explanatory framework within which to reason about some phenomenon; the second is as a source of novel predictions about behavior. We believe that the model presented here provides a useful framework for the parameterization of learning in infants. Does it lead to any predictions? A first prediction of our model arises from inspection of Figure 4, in which the index of novelty (Equation 1) can be seen to be nonmonotonic with respect to time. The model predicts that for some habituation–dishabituation tasks, groups of infants will exhibit nonmonotonic behavior. Over successive time periods, the elicitation of such contrasts will initially be relatively easy, then relatively hard, and then relatively easy again. In fact, such behavior has already been observed with respect to this task. When presented with a task that defeated the 14-month-olds ([bɪh]–[dɪh] discrimination in a one-object bimodal procedure; Stager & Werker, 1997, Experiment 2), 20-month-olds were able to make the fine discrimination (Werker et al., 1999). To treat such a result as confirmation of our prediction is premature for two reasons, however. First, perhaps 20-month-olds are responding in an adultlike fashion. But at the end of the track in Figure 4—that is, at 50,000 language trials—our networks have not recovered above the 20% criterion. Second, although the U-shaped function may well be a general feature of this type of learning (e.g., Plunkett et al., 1992), it would be more prudent to demonstrate the phenomenon in networks trained with a more realistic vocabulary and phonemic feature description.

A second prediction arises from the way in which we have formalized our description of how the infant might be learning. We have made the minimal assumption that the infant learns, through time, which name goes with which image. That is, the infant is a covariation detector. Based on this assumption, the model makes a prediction about how infants might behave differentially in the bimodal habituation task depending on the auditory discrimination stimuli used. This prediction arises from the general characteristics of this class of learning algorithms (autoencoders). Such predictions form testable hypotheses of the theoretical assumptions on which the model is based. To explain this prediction more fully, however, it is necessary to review some of the formal properties of the type of networks implemented in this study. The backpropagation algorithm is a multilayer extension of the delta rule, itself formally equivalent to the Rescorla–Wagner model of classical conditioning (Elman et al., 1996; Rescorla & Wagner, 1972; Rumelhart, McClelland, & the PDP Research Group, 1986). During training, either in the language acquisition component or in the habituation component, the network adjusts its weights to generate a compressed description of the input patterns. The network processes the N-dimensional input (where $N = 36$ is the number of input units) to minimize the error on the output units. This is a "greedy" learning algorithm that alters weights to bring about the largest instantaneous error reduction. This results in a redescription of the input as a linear combination of k basis vectors, where k is the number of hidden units (18 in this case). At any given time, the input vectors are redescribed across the hidden units in such a way that the maximum variance is captured by the k basis vectors. Under some conditions, this corresponds to computing the k first principal components of the training set (e.g., Baldi & Hornik, 1989; but see also Japkowicz, Hanson, & Gluck, 2000); however, this process is constantly evolving as training proceeds. In other words, the principal component description of the training set that is generated across the hidden units is also evolving with time.

On encountering the 240 label–object pairs, the network begins to adjust its weights (which were initially set randomly between –0.5 and +0.5) to minimize the error between output and input (autoencoding). Because the backpropagation algorithm always takes the route of greatest error reduction when adjusting weights, the network initially deals preferentially with the first principal component of the input data.[6] The second principal component tends to come next, and so on, up to *k* principal components. Because backpropagation is a greedy algorithm, however, it always makes weight changes that reduce error maximally, at that instant in training. It may, for example, be more efficient to switch to learning about the second principal component before all the variability associated with the first principal component has been reduced; however, because we are interrupting this process at different intervals (to habituate and test the network), we are probing different, transitory states in the network's principal component description of the input matrix (the artificial 240-word language we have created). In effect, we are testing networks with different internal models of their common language environment. These transitory states are differentially sensitive to the two habituation stimuli, [bɪh] and [lɪf]. Although the networks' behavior is determined by the nature of the input (language exposure and habituation stimuli), this behavior is not readily predicted, because of the complex nature of the input. In the case of nonlinear units of the sort we have employed here, a further important source of complexity is that nonlinear networks with random initial weights are likely to find different paths to the solution (Japkowicz et al., 2000); that is, to show different transitory states. This fact contributes to the nonmonotonic progression seen in the data in Table 1.

Having briefly discussed how networks extract structure from their environment, we are now in a position to consider what predictions flow from this approach. In general, there will be pairs of auditory discrimination stimuli—let us call them [A1]–[A2] and [B1]–[B2]—which, although equally distinct from one another in terms of phonemic feature distance, give rise to differing habituation–dishabituation behaviors over time, because of the different relations that the two pairs have with the statistical properties of the input language. To illustrate this, suppose that in word-initial position, the phone [b] is more commonly encountered (in the input to infants) than is the phone [k]. Further suppose that in word-final position, these phones ([b] and [k]) are encountered equally often. The increased presence of [b] in the input, relative to [k], in initial position, will, all else remaining

[6]More formally, it reduces error preferentially in the direction of the eigenvector of the input covariance matrix associated with the largest eigenvalue (Baldi & Hornik, 1989).

equal, result in preferential learning about initial-[b] over initial-[k]. There will be no such preference in leaning about word-final position. As a result, discriminations of the form [bXk]–[kXk] (where X stands for any phone) will result in greater release from habituation than will discriminations of the form [kXb]–[kXk], even though the two discriminations are equally distant in phonemic feature space. Furthermore, this differential response in the habituation–dishabituation task will interact with the infant's developing state of knowledge of the input language in a manner that is rather hard to predict mathematically but can readily be modeled given a plausible description of the input vocabulary. Finding such pairs may thus provide us with a toehold for determining the similarity structure of infants' internal code for phonemic information.

Making more specific predictions requires a more accurate characterization of the phonological code used by infants. The language that the networks learned in this study is made up of a random collection of CVCs; thus, although the phonemic feature description of each of the three phonemes in each of the 240 labels bears some relation to the structure of English (i.e., at the level of phonemic features), the relations between given phonemes do not.[7] This limits the generality of the model, which can be expected to be reasonably accurate at depicting the relative distinctiveness of pairs of CVCs (so that, e.g., [bɪh] is more similar to [dɪh] than [lɪf] is to [nim]) but to be less accurate about, say, the relative order of emergence of knowledge about particular phonemes in actual English. For specific predictions of this nature, a model that learns a more realistic vocabulary is required.

In summary, this research had three primary objectives: The first was to model a pattern of behavior in which responses to a set of stimuli emerge from the interaction of a single algorithm acting on a complex data set. Our model provides an example of how a pattern of qualitative shift in behavior can be observed without requiring a corresponding shift in the underlying processing (i.e., algorithm). A second goal was to present a method for the modeling of habituation phenomena in infants. We believe that conceptualizing habituation as a process of representation construction within an associative network is very useful. The final goal was to illustrate how long-term experience (language learning) and short-term experience (within-task learning) can interact to produce what appear to be age-related developmental differences.

We have seen that a novelty preference in an infant habituation–dishabituation paradigm may be modeled by a simple learning device, and further, that this preference is highly nonmonotonic over experience. We conclude that complex behavior is seen in simple systems. Although this observation has been made before (e.g., Elman et al., 1996; Thelen & Smith, 1993), we have shown that it may be ap-

[7]Intraphonemic regularities are captured; interphonemic regularities are not.

plied to infant habituation data in the rich field of speech-sound discrimination and word learning in infancy.

ACKNOWLEDGMENT

We are grateful to Janet Werker and three anonymous reviewers for comments on an earlier draft of this article.

REFERENCES

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

Bar-Adon, A., & Leopold, W. (Eds.). (1971). *Child language: A book of readings.* Englewood Cliffs, NJ: Prentice Hall.

Barrett, M. D. (Ed.). (1999). *The development of language*. Hove, England: Psychology Press.

Best, C. T., McRoberts, G. W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two nonnative consonant contrasts. *Infant Behavior and Development, 18,* 339–350.

Best, C. W., McRoberts, G. W., & Sithole, N. N. (1988). The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 345–360.

Bornstein, M. H. (1985). Habituation of attention as a measure of visual information processing in human infants: Summary, systematization, and synthesis. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview* (pp. 253–301). Norwood, NJ: Ablex.

Brainerd, C. (1978). The stage question in cognitive-developmental theory. *Behavioral and Brain Sciences, 2,* 173–213.

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, England: Oxford University Press.

Chauvin, Y. (1988). *Symbol acquisition in humans and neural (PDP) networks*. Unpublished doctoral dissertation, University of California at San Diego, San Diego, CA.

Cohen, L. B. (1969). Observing responses, visual preferences, and habitutation to visual stimuli in infants. *Journal of Experimental Child Psychology, 7,* 419–433.

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

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.

Fletcher, P., & MacWhinney, B. (Eds.). (1995). *The handbook of child language*. Oxford, England: Blackwell.

Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28,* 3–71.

Garnica, O. K. (1973). The development of phonemic speech perception. In T. Moore (Ed.), *Cognitive development and the acquisition of meaning* (pp. 214–222). New York: Academic.

Hirsh-Pasek, K., & Golinkoff, R. M. (1996). *The origins of grammar: Evidence from early language comprehension*. Cambridge, MA: MIT Press.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. In C. Rovee-Collier & L. Lipsitt (Eds.), *Advances in infancy research* (Vol. 5, pp. 69–95). Norwood, NJ: Ablex.

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.

Ingram, D. (1989). *First language acquisition: Method, description and explanation*. Cambridge, England: Cambridge University Press.

Jakobson, R. (1968). *Child language, aphasia, and phonological universals*. The Hague, Netherlands: Mouton. (Original work published 1941)

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

Kellman, P. J., & Arteberry, M. E. (1998). *The cradle of knowledge: Development of perception in infancy.* Cambridge, MA: MIT Press.

Kemler-Nelson, D. G. (1984). The effect of intention on what concepts are acquired. *Journal of Verbal Learning and Verbal Behavior, 23,* 734–759.

Lamb, M. E., & Bornstein, M. H. (1987). *Development in infancy* (2nd ed.). New York: Random House.

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: Lawrence Erlbaum Associates, Inc.

Mareschal, D., & French, R. M. (2000). Mechanisms of categorization in infancy. *Infancy, 1,* 59–76.

Mareschal, D., Plunkett, K., & Harris, P. L. (1999). A computational and neurophysiological account of object-oriented behaviours in infancy. *Developmental Science, 2,* 306–317.

Mareschal, D., & Shultz, T. R. (1996). Generative connectionist networks and constructivist cognitive development. *Cognitive Development, 11,* 571–603.

McClelland, J. L. (1995). A connectionist perspective on knowledge and development. In T. J. Simon & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 157–204). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Munakata, Y. (1998). Infant perseveration and implication for object permanence theories: A PDP model of the AB task. *Developmental Science, 1,* 161–211.

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S (1997). Rethinking infant knowledge: Towards an adaptive process account of successes and failures in object permanence tasks. *Psychological Review, 104*, 686–713.

Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.

Plunkett, K., & Elman, J. L. (1997). *Exercises in rethinking innateness: A handbook for connectionist simulations*. Cambridge, MA: MIT Press.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for language acquisition. *Cognition, 38,* 43–102.

Plunkett, K., Sinha, C., Møller, M. F., & Strandsby, O. (1992). Symbol grounding or the emergence of symbols? Vocabulary growth in children and a connectionist net. *Connection Science, 4,* 293–312.

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

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II* (pp. 64–69). New York: Appleton-Century-Crofts.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. *Nature, 323,* 533–536.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations*. Cambridge, MA: MIT Press.

Shultz, T. R. (1998). A computational analysis of conservation. *Developmental Science, 1,* 103–126.

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

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception tasks than in word-learning tasks. *Nature, 388,* 381–382.

Stern, C., & Stern, W. (1907). *Die Kindersprache: Eine psychologische und sprachtheoretische Untersuchung* [Child language: A psychological and linguistic investigation] (4th rev. ed.). Leipzig, Germany: Barth.

Thelen, E., & Smith, L. (1993). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Trehub, S. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development, 47,* 466–472.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word–object associations by 14-month-old infants. *Developmental Psychology, 34,* 1289–1309.

Werker, J. F., Corcoran, K. M., & Stager, C. L. (1999, June). *Infants of 20 months can distinguish similar sounding sounds.* Poster presented at the 9th annual meeting of the Canadian Society for Brain, Behavior, & Cognitive Science, Edmonton, Canada.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24,* 1–12.

Werker, J. F., & Pegg, J. E. (1992). Infant speech perception and phonological acquisition. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 285–311). Timonium, MD: York Press.

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology, 37,* 278–286.

Werker, J. F., & Tees, R. C. (1984a). Cross language speech perception: Evidence for perceptual reorganisation during the first year of life. *Infant Behavior and Development, 7,* 49–63.

Werker, J. F., & Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75,* 1866–1878.