An Interacting Systems Model of Infant Habituation

Sylvain Sirois[1] and Denis Mareschal[2]

Abstract

■ Habituation and related procedures are the primary behavioral tools used to assess perceptual and cognitive competence in early infancy. This article introduces a neurally constrained computational model of infant habituation. The model combines the two leading process theories of infant habituation into a single functional system that is grounded in functional brain circuitry. The HAB model (for Habituation, Autoassociation, and Brain) proposes that habituation behaviors emerge from the opponent, complementary processes of hippocampal selective inhibition and cortical long-term potentiation. Simulations of a seminal experiment by Fantz [Visual experience in infants: Decreased attention familiar patterns relative to novel ones. *Science, 146,* 668–670, 1964] are reported. The ability of the model to capture the fine detail of infant data (especially age-related changes in performance) underlines the useful contribution of neurocomputational models to our understanding of behavior in general, and of early cognition in particular. ■

INTRODUCTION

Infancy has long been viewed as a crucial testing ground for the multiple incarnations of the recurring nature–nurture debate. In the last 20 years or so, a large number of infancy studies have explored early infant cognitive abilities with experiments based on habituation. The results of these experiments have been presented as evidence that infants younger than 6 months possess a conceptual understanding of the world (Spelke, 1998, 1999, 2000; Wynn, 1995; Spelke, Breinlinger, Macomber, & Jacobson, 1992; Baillargeon, 1987, 1999). An ongoing debate ensued when suggestions of complex cognitive abilities were questioned on the grounds of simpler perceptual explanations (Sirois, Buckingham, & Shultz, 2000; Haith, 1998), statistical properties of the experiments not related to the concepts under study (Bogartz, Shinskey, & Speaker, 1997), or methodological issues (Cohen & Marks, 2002). But this use of habituation methods to assess infant conceptual development is a relatively recent turn of events. In fact, habituation methods continue to play a central role in the investigation of learning, memory, and a range of perceptual and cognitive phenomena throughout infancy.

Whatever the paradigm or purpose of research, we believe that a better understanding of what kind of representations and processes are required to produce the behaviors observed in young infants can help clarify habituation findings. In this article, we will argue that computational modeling of infant habituation with due consideration to the relevant brain structures that support habituation can shed some much needed light in this respect. Indeed, neural computational models that draw on functional brain systems as sources of constraints have provided explanatory frameworks for a range of infant behaviors (e.g., Westermann & Mareschal, 2004; Mareschal & Johnson, 2002; Mareschal, Plunkett, & Harris, 1999; Dehaene & Changeux, 1993). Models are especially useful because they force researchers to be explicit about representations and processes. Moreover, giving the brain due consideration usefully constrains the range of plausible interpretations. A model that excels at simulating infant performance but that contradicts neuroscientific evidence would be, in our view, of limited interest.

Infant Habituation

Habituation behavior has been observed in a broad range of species, from simple mollusks to amphibians to mammals, including humans. This has led researchers to dub such a pervasive process as the ''simplest form of learning'' (Thorpe, 1956). However, infancy researchers are not typically interested in the same processes as observed in mollusks or worms, whereby habituation reflects processes local to the sensory receptors (Sirois, 2004). Rather, infant habituation is normally studied within a framework derived from the pioneering work of Soviet physiologist Evgeni Sokolov and his colleagues (Vinogradova, 1975; Sokolov, 1963, 1975).

Within this framework, infants are assumed to build (through learning) a neural or mental model of stimuli or stimulus events as these are repeatedly presented. When a stimulus is presented, it is compared to this neural model and discrepancies provide the basis for learning. This is often referred to as the comparator theory (Gilmore & Thomas, 2002). As learning progresses, discrepancies between the model and external

---

[1]University of Manchester, [2]Birkbeck University of London

events decrease, which result in a decrease of attention (i.e., there is a progressively lesser need to process the information as the internal model approximates such information). Crucially, it is assumed that stimuli that deviate from this internal model would elicit a relative increase in responding (a behavior referred to as dishabituation, release from habituation, or renewed responding). Indeed, a stimulus perceived as novel would require more processing than one perceived as familiar. Early work by Frances Graham and her colleagues provides a thorough introduction to comparator theory, and how physiological measures such as heart rate decelerations provided crucial refinements to the model (see also Graham, 1992; Graham et al., 1970; Clifton, Graham, & Hatton, 1968; Graham, Clifton, & Hatton, 1968; Graham & Clifton, 1966).

Renewed responding thus provides researchers with a unique opportunity to investigate internal representations in preverbal infants. What infants perceive as novel (inferred from renewed responding) given a set of habitual stimuli or events allows for inferences as to how they represent information. By carefully designing stimuli sets, researchers can systematically examine in which ways stimuli may be represented as distinct.[1]

Another influential account of habituation is proposed by Groves and Thompson (1970). Dual-process theory suggests that habituation and sensitization (i.e., increased responding to repeated stimulation) represent two distinct outcomes of the combination of independent excitatory and inhibitory processes. Sensitization is viewed as brief and not stimulus specific (i.e., it is a state arousal phenomenon), and thus should be reflected most in early trials, whereas habituation is stimulus specific and involves learning. Both processes combine to produce overt behavior, whereby the relative contribution of each is affected by stimulus complexity and arousal state. For instance, more complex stimuli can elicit more arousal and also be harder to learn, resulting in increased responding, relative to simpler stimuli. Moreover, increased arousal (e.g., using loud background noise) can augment sensitization relative to habituation. Although dual-process and comparator theories pertain to explain habituation, there have been few attempts to integrate them in a coherent whole (e.g., Graham, 1992).

We believe that computational modeling aimed at capturing behavioral features of habituation while being consistent with the neuroscience of habituation would provide solid basis on which to discuss infant behavior. A successful neurocomputational model of habituation would provide a causal framework for understanding how infants behave, and in particular, whether both leading theoretical accounts can be integrated. To this end, we have identified seven behavioral and neural markers of infant habituation that proper models of habituation need to accommodate (Sirois & Mareschal, 2002c). Although we certainly do not believe that these models need to be models of neural circuits per se, we

do believe that they should be consistent with known functional properties of neural circuits involved in habituation. Moreover, we are also aware that these neural and behavioral markers are but a subset of a larger list of habituation features (e.g., Thompson & Spencer, 1966). Nevertheless, these seven markers were chosen because they are directly relevant for current discussions about infant habituation, and thus as those by which to evaluate the value of models that attempt to explain infant habituation. Of the seven markers we identify, the first five are behavioral, and the last two are derived from a need for consistency with neural systems.

First, habituation responses generally show an exponential decrease over trials (Thompson & Spencer, 1966). Whether one measures looking times or heart rate changes, larger values will be recorded on the first few trials and these will drop rapidly to a low asymptote.[2] Secondly, responses have a temporal dimension. Although this may seem obvious, as we are discussing "durations of looks" or "heart rate changes," the temporal aspect of responses poses problems for several models of habituation (Sirois & Mareschal, 2002c). Third, there is a shift from a familiarity preference to a novelty preference over the course of habituation (Roder, Bushnell, & Sasseville, 2000; Hunter, Ames, & Koopman, 1983; Rose, Gottfried, Melloy-Carminar, & Bridger, 1982; Fantz, 1964). All other things being equal, infants will initially look longer at familiar items, and it is only later, when habituation is established, that they look longer at novel items (Jankowski, Rose, & Feldman, 2001; see also Jankowski & Rose, 1997; Rose & Feldman, 1987). Fourth, novelty responses disappear over repeated testing (Thompson & Spencer, 1966). In other words, infants habituate to test stimuli. Finally, infants do not represent habituated items indiscriminately (Quinn, 1987). Although they may show reduced interest to a class of stimuli at the end of habituation training, this appears to be the outcome of a learning process that spares (at least in part) the identity of individual items.

We also identify two specific neural functions involved in habituation that models should accommodate. Thus, the sixth point is that habituation involves a process of selective inhibition in the hippocampus (Nelson, 2002). The hippocampus has been identified as the locus of habituation in a variety of comparative studies (Wang, 1995; Vinogradova, 1975). Hippocampal neurons, over the course of habituation, show decreased responding to specific perceptual features or to combinations of such features (Vinogradova, 1975). The specificity of decreased responding explains dishabituation, whereby new stimulus features or combinations of features would not have been inhibited. The process is fast and momentary, and is likely sustained by the γ-aminobutyric acid (GABA) neurotransmitter (Disney & Calford, 2001). Importantly, the hippocampus is largely mature and functional from birth (Seress, 2001), which explains neonatal and even in utero habituation.

The seventh marker is that hippocampal-related cortical areas (especially the entorhinal cortex) are involved in both habituation and memory formation (Nelson, 2002; Mishkin, Vargha-Khadem, & Gadian, 1998). This is especially relevant, as the sort of short-term potentiation of the hippocampus cannot account for the well-documented long-term habituation effects (Zelazo, Weiss, & Tarquinio, 1991), nor for a potential contribution of prior experience to habituation performance (Bachevalier, 2001). Cortical areas allow for an interface between long-term potentiation (LTP) and short-term potentiation (STP). Indeed, recent imaging work with adults suggests that item familiarity memory is mediated by the perirhinal cortex and the dorsomedial thalamus, and not the hippocampus (Montaldi, Spencer, Alvarez, Roberts, & Mayes, 2003).

Data from both nonhuman primates and from humans suggest that novelty preferences observed in visual comparison tasks (such as those typically used with infants) are mediated by the hippocampus and surrounding structures (Bachevalier, Brickson, & Hagger, 1993; McKee & Squire, 1993). In addition, electrophysiological studies of monkeys engaged in delayed matching to sample tasks, highlight that neurons in TE, and in the perirhinal and entorhinal cortices exhibit different types of responses. A large proportion of neurons exhibit repetition suppression, whereby the neural responses are suppressed following repeated exposure to a stimulus (Miller & Desimone, 1994); this has also been observed in humans with fMRI (Jiang, Haxby, Martin, Ungerleider, & Parasuraman, 2000). However, several neurons in the entorhinal, perirhinal, and prefrontal cortices show enhanced responses to test stimuli that match the sample, a so-called response (or match) enhancement (Suzuki, Miller, & Desimone, 1997; Miller & Desimone, 1994). Finally, Miller and Desimone (1994) have shown that responses in the perirhinal cortex are enhanced with repetition of the sample stimulus; there appears to be no such enhancement when repeating nonmatching stimuli. Of course, these findings come from adult and comparative studies, and must therefore be interpreted with caution when considering their implications for infant processing.

To our knowledge, no existing model successfully captures all seven of the markers above. We therefore developed the model discussed in the next section.

The HAB Model of Habituation

The HAB model (for Habituation, Autoassociation, and Brain) maps autoassociator neural networks to a functional, updated version of Sokolov's (1963) Comparator framework. Autoassociation is known to exist both in the neocortex and the hippocampus (Rolls & Treves, 1998). The model's functional circuits are depicted schematically in Figure 1. External input relayed through the thalamus is forwarded to both the hippocampus and cortical areas. While the hippocampus learns to selectively inhibit the input, the cortical area learns to reproduce its input. These two functional areas have reciprocal connections, such that the hippocampus dampens previously processed input to the cortex, and the cortex amplifies known input to the hippocampus. This opponent process is related to Groves and Thompson's (1970) dual-process theory. Finally, both hippocampal and cortical areas modulate the response systems. Thus, the HAB model provides a synthesis of the dual-process (Groves & Thompson, 1970) and the comparator (Sokolov, 1963) theories of habituation that is grounded in the interactions of functional neural systems.

Figure 2 shows a full connectivity diagram for a small (i.e., three inputs) HAB model. An external input is processed by both hippocampal and cortical networks. The output of each network serves to dynamically modulate the input of the other network, as well as provide to the overall response of the system on a bank of output units. Further details, including equations and parameters, are provided in the Methods section.

We compared the model's performance against Robert Fantz's (1964) seminal experiment on human infant habituation. In his experiment, Fantz showed infants aged between 1 and 6 months 10 pairs of various pictures clipped from magazines. For each child, one picture was common to all pairs (the constant pattern), and it was paired with 10 different pictures (the variable pattern). Looking times to each picture when a pair was shown

Figure 1. Functional organization of the HAB model. Dark arrows represent excitatory pathways, and the light arrow represents an inhibitory pathway [adapted from Sirois, 2004].
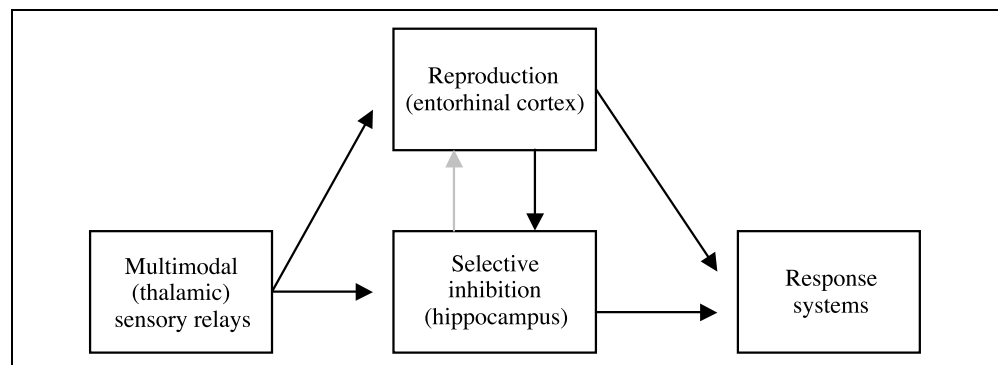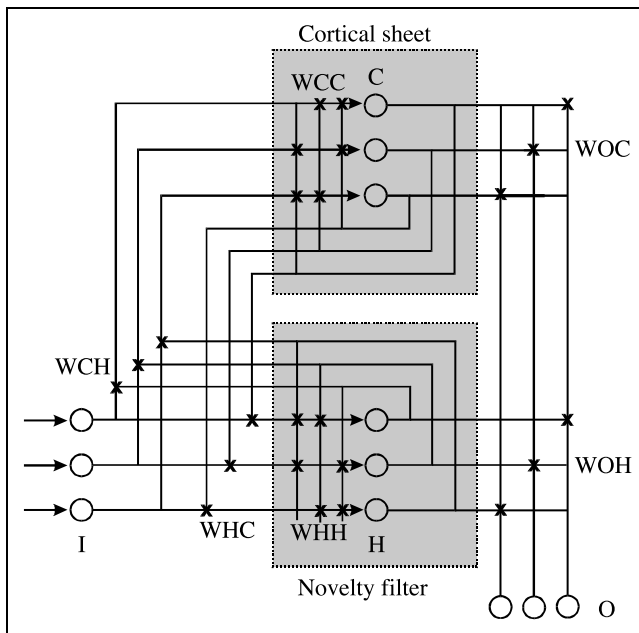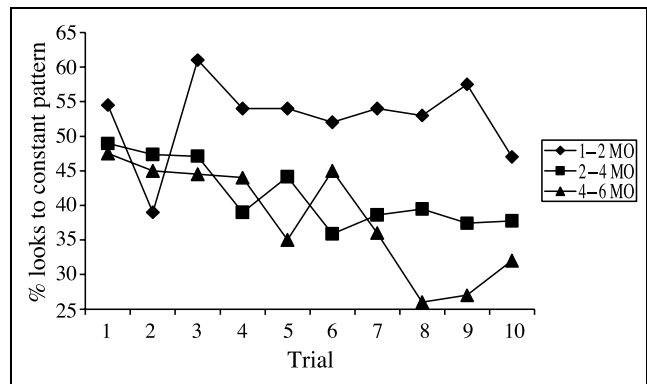
Figure 2. Connectivity diagram of a simple HAB model. External input would be applied to a bank of input units (I). This external input propagates to both hippocampal units (H) and cortical units (C). In the model, selective inhibition in the hippocampus is implemented by a novelty filter network (Kohonen, 1988), highlighted in the lower part of the figure. Local processing and learning takes place in WHH connections. Also highlighted is the cortical system, whose LTP function is implemented by a standard autoassociator network. Learning and local processing involves connections in WCC. Hippocampal units gate input to the cortex through connections WCH, and cortical units amplify hippocampal input through connections WHC. The contribution of each system to the output of the model (O) is weighed by connections WOC and WOH. (Adapted from Sirois & Mareschal, 2002b.)

Figure 3. Percentage of time spent looking at constant stimulus in three different age groups (after Fantz, 1964).

were recorded. The performance of the infants (divided into three age groups[3]) is shown in Figure 3. For the younger infants, there was no reliable preference for either constant or variable patterns. For older infants, however, there was an increasing preference for the variable pattern over trials. This preference increased with age. It is worth noting that Fantz's paradigm has been used recently for replication (Roder et al., 2000) and clinical (Rose, Feldman, & Jankowski, 2002) purposes.

Fantz (1964) identified age-related changes in habituation patterns that the model can explain through systematic changes in the functional neural circuits underpinning habituation. This idea builds in the evidence marshaled in support of a shift from subcortical-to-cortical control of visual attention (Nelson, 2002; Johnson, 1990, 1997; Atkinson, 1984; Bronson, 1974). In this view, early behaviors are largely under the control of subcortical processes. Cortical control progressively takes over as development unfolds. With respect to structures involved in habituation, the hippocampus and the entorhinal cortex appear relatively mature early in postnatal life (Seress, 2001; Nelson, 1995). However, the volume of the limbic cortex reaches adult-like levels

only in the second half of the first year (Nelson, 2002), if not later (Utsunomiya, Takano, Okazaki, & Mitsudome, 1999). The period between 3 and 12 months of age is characterized by excessive synaptogenesis (Seress, 2001). In addition, metabolic activity in the temporal lobe increases during the first 3 months of life, and there appears to be a growth difference between the hippocampus and other structures in the temporal lobe (Utsunomiya et al., 1999). Moreover, the connections from the entorhinal cortex to the hippocampus follow at least two developmentally distinct paths (van Groen, Miettinen, & Kadish, 2003). Thus, the cortical contributions to habituation change in many different ways over time, with adult-like function not fully emerging before the fifth year postpartum (Seress, 2001). Indeed, Graham et al. (1968) have also proposed that a possible subcortical–cortical shift in the first year postpartum might explain age-related changes in heart rate deceleration during habituation.

In the current model, we simplify this process by proposing that the strength of the cortical input to both hippocampal and output systems (weights WOC and WHC) increases over the first year of life. This has the effect of increasing the impact on habituation of cortical computations. We predict that increased cortical contribution to habituation will reproduce age-related effects observed by Fantz (1964). In reporting the results, we begin by describing the performance of the model with default connectivity parameters leading to a performance level comparable to that of 2- to 4-month-olds. We then describe how relative decreases and increases of connectivity between subsystems, mapping maturational changes, lead to performance levels of 1- and 6-month-olds, respectively.
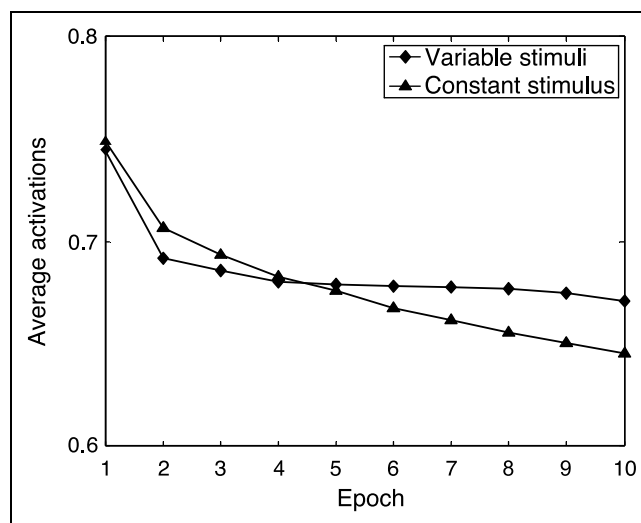
RESULTS

The performance of the HAB model with default parameters captures the typical behavioral profile of habituation training, as depicted in Figure 4. Average output

Figure 4. Average unit activations over all patterns for 20 networks. Activations are shown for output, hippocampal, and cortical units.
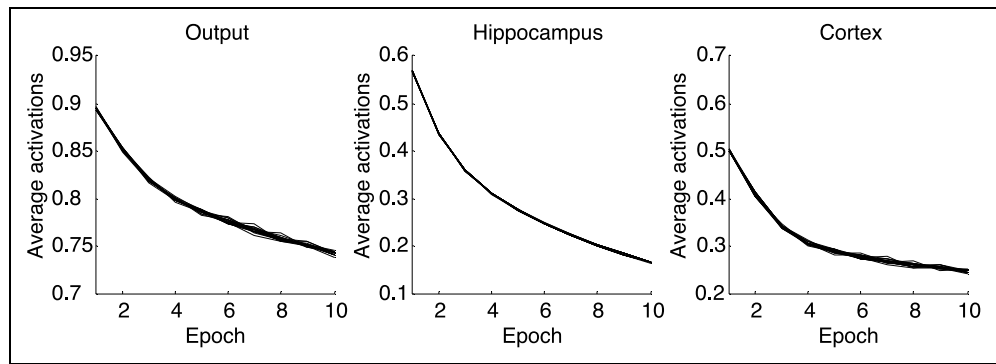
activations[4] over processing cycles for 20 networks exhibit exponential decrease over trials. In Figure 5, we average output across the networks and separate out activations to constant and variable portions of stimuli. Although activations are initially larger for the constant pattern, there is a shift midway through training when activations become larger for the variable pattern (thus showing a familiar-to-novel shift). In Figure 6, we plot average activation to the constant pattern relative to average overall activation, as Fantz (1964) did (see Figure 3). The average output to constant stimulus is divided by overall average output, thereby providing a relative or proportional activation score for the constant stimulus with respect to the total activation elicited by both stimuli. The figure highlights the familiarity–novelty shift typical of such habituation experiments with infants (Roder et al., 2000).

It is also interesting to examine the behavior of the model as it unfolds within a single trial, as shown in Figure 7. Even by the end of training, when the model exhibits a stronger response to the variable stimulus on average, the initial response of a HAB model on the first cycle of a trial is always stronger for the constant or familiar stimulus. As training progresses, this initial familiarity response on a given trial produces stronger inhibition in the hippocampal system, which results in progressively more novelty preference.

What of the age-related changes in habituation profiles observed by Fantz (1964)? Figure 8 shows relative activation to the constant pattern for immature, default, and mature versions of the HAB model. The immature model shows an initial preference for the constant stimulus that progressively disappears but does not turn into a preference for the variable pattern. The default model, discussed above, exhibits a smooth familiarity-to-novelty shift. Finally, the mature model never quite reaches an initial familiarity preference despite increased activations to the constant pattern over the first three epochs. After the third epoch, the mature model exhibits a progressively more pronounced novelty preference than the default model. Plotting the data over processing cycles, as in Figure 9, shows how for each version of the model the initial response on a trial is always larger for the constant or familiar pattern. The inhibitory effect of this

Figure 5. Average output activations by epoch for constant and variable stimuli.
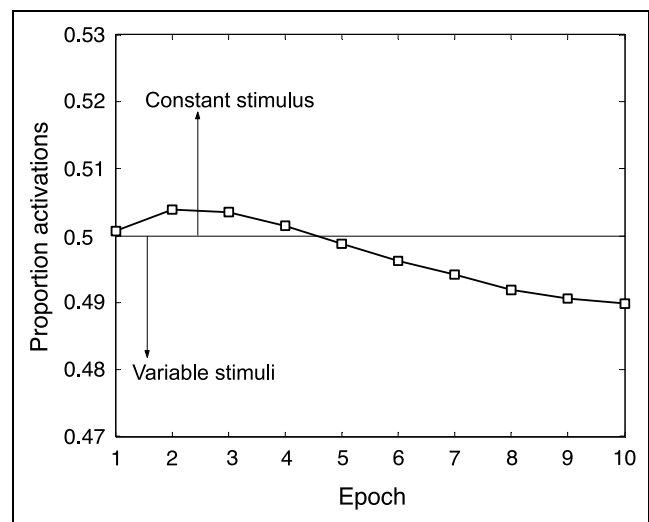
Figure 6. Fantz-type plot of the data from Figure 4. The proportion of activation values for the constant stimulus over the variable stimuli is plotted by epoch. Values above 0.5 reflect orienting to the constant stimulus, whereas values below 0.5 reflect orienting response to the variable stimulus.
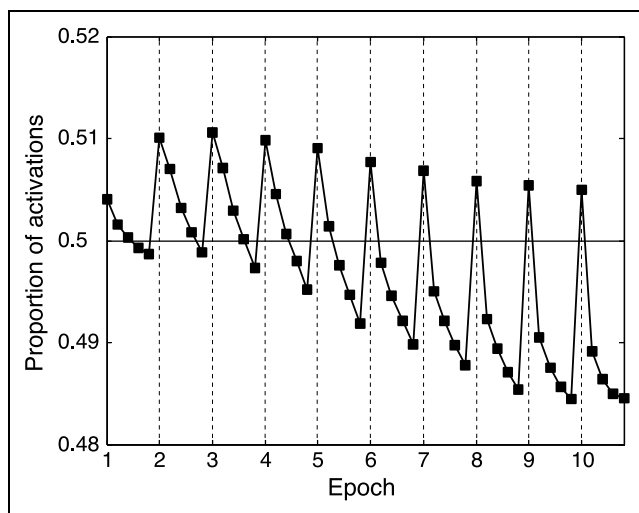
initial activity for the constant stimulus, discussed for the default version previously, is slight for the immature model but pronounced in the mature version.
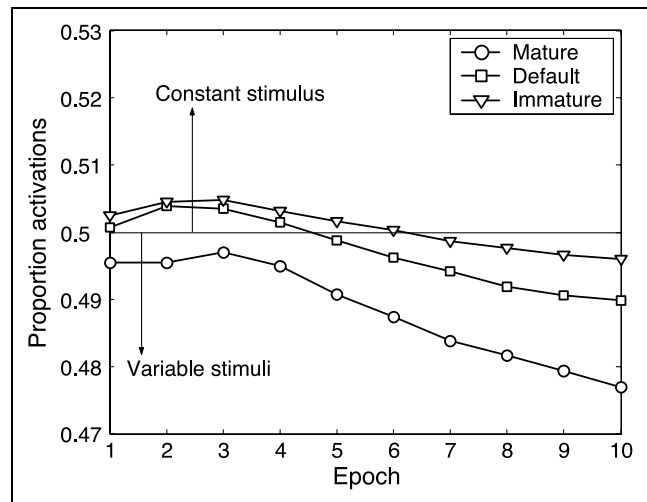
Conclusion

The HAB model is, by design, consistent with the relevant functional neuroscience of habituation. This, and the natural emergence of the key behavioral markers of habituation in human infants, sets it apart from prior models of habituation (Sirois & Mareschal, 2002c). Indeed, in light of the HAB model's performance, the contribution of interacting, opponent processes appear as necessary conditions to an account of infant habituation.

The default parameters of the HAB model produce a behavioral profile consistent with both the coarse and fine detail of the 2- to 4-month-olds in Fantz's (1964) study. In addition, by varying the effective strength of the connections from the cortical system, the model captures the behavioral profiles of infants from 1 to 6 months. In effect, further consideration of the neural basis of habituation leads to a deeper understanding of habituation as it unfolds within a trial, as well as it unfolds across development.

The HAB model, overall, is a promising framework to study how infants process information during habituation given its unique ability to explain behavior while being consistent with the underlying neuroscience. The model does not implement abstract, conceptual knowledge and as such would be a parsimonious, low-level account of infant performance. Infant habituation is viewed as the functional combination of short-term inhibition and long-term potentiation, whereby the reciprocal connections between these opponent functions

Figure 8. Fantz-type plot of data from mature (6-month-old), default (2- to 4-month-old), and immature (1-month-old) networks. Relative activations for constant and variable stimuli by epoch are shown for each type of network.

mature in the first year of life. Because it is implemented as an artificial neural network, the model can easily generate novel, testable predictions.

As Figures 7 and 9 have shown, the model's initial familiarity response at the onset of any trial produces stronger inhibition in the hippocampal system, resulting in progressively more novelty preference over processing cycles (increasingly so with maturity). In some ways, this feature of the model is analogous to the inhibition of return effect observed in visual attention experiments (albeit involving different neural structures; see Klein,
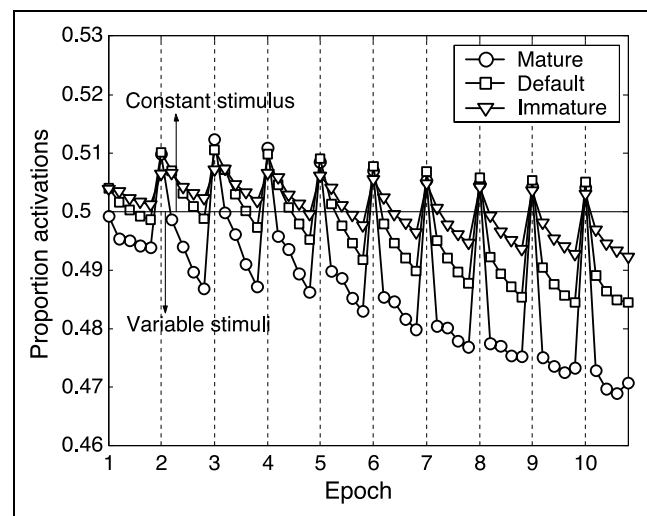

Figure 7. Plot of data from Figure 5, broken down by processing cycle. The initial response of the model is always largest for the familiar or constant stimulus even when the average response (within a trial) indicates orienting to the novel or variable stimulus.

Figure 9. As shown in Figure 7, but broken down by processing cycle. Relative activations for constant and variable stimuli by epoch and by cycle are shown for each type of network. For all networks, the initial response at the onset of a trial is largest for the constant stimulus, throughout the simulation.

2000). The model predicts that in pairwise habituation experiments, looking times to novel stimuli should be relatively longer when infants first look at a familiar stimulus. In other words, the relative novelty of a stimulus is directly enhanced through inhibition of familiar information.

An obvious next research step would be to "situate" the model within an environment in which the model's current output affects subsequent input. As a preliminary step, we could use the output to directly modulate the input through corresponding connections. A more satisfying solution may eventually involve output control of a "foveal" area that constrains input.

Two other models of habituation have recently been proposed. Gilmore and Thomas (2002) describe a statistical tool to fit habituation curves to individual infants' behavior. This tool is an important methodological advance because it can be used online to assess when infants reach habituation criterion, taking into account the effects of noise. However, it is not a process model of habituation (nor does it claim to be) and as such serves a different purpose than a model such as HAB.

Schoener and Thelen (in press), however, have proposed a dynamic field model of habituation which has many similarities, in spirit, to HAB. In their model, two coupled and interacting fields produce looking: one field that drives looking and one that drives looking away. Like HAB, their model is consistent with the dual-process theory. However, although this model does a good job of explaining how an individual habituation trial unfolds through time, it is unclear where LTP (long-term learning) takes place in the model. Nor is it clear by what mechanisms the dynamics of the system are affected by prior knowledge. Although the temporal dynamics of the HAB model may be similar to those of the Schoener and Thelen model, the close mapping between the HAB model and a functional neural substrate allows us to make explicit statements about how LTP in either the cortical or the hippocampal sheets affects the current habituation behaviors.

By providing a domain general account of habituation, the HAB model opens many doors for future research. Just as habituation (as a method) is ubiquitous in infancy research, so too could this model (or one based on the same processing principles) be used to model a very broad range of behaviors across a number of domains. For example, there are two important areas to which the HAB model could be applied. First, the issue of individual differences in habituation, which has received substantial attention (Jankowski et al., 2001; Jankowski & Rose, 1997; Bronson, 1991; Mayes & Kessen, 1989; Colombo, Mitchell, Obrien, & Horowitz, 1987a, 1987b; Bornstein & Benasich, 1986; Ritz, Rose & Feldman, 1984; Woodruff, & Fagen, 1984; Rose et al., 1982, 2002; McCall, 1979). Gilmore and Thomas (2002) view habituation data as noise around a monotonic decreasing exponential function. The crucial question is to what extent the noise is generated by the process of habituation (variability of endogenous origin), as opposed to external sources (variability of exogenous origin). There are many candidate causes of endogenous variability in neural networks (see Thomas & Karmiloff-Smith, 2003, for a full discussion of potential factors giving rise to individual differences in connectionist network models). These include noisy connection weights, restricted connectivity, prior experience stored in the cortical network, and learning rates that vary between networks. Exogenous sources of variability, including differences in arousal, attentional biases, task demands, and distractions would merely result in noisy input to the model. However, whether a definitive answer is even possible remains an open question. Because many combinations of different parameter or state changes could lead to the same observable behavioral variation, it may be almost impossible to isolate single causes of individual variation. However, computer simulations can play an important role in this debate by mapping out the relation between parameter states and observed behaviors. By referring back to this "map," one can then work back probabilistically (e.g., via maximum likelihood estimates) from individual infant behaviors to likely causes of the individual variability in that infant.

A second domain that we have already begun to investigate is that of perceptual categorization in young infants because it involves the interaction between habituating to the immediate stimulus and some representation of previously encountered stimuli retrieved from memory (Mareschal & Quinn, 2001). During the development of the HAB model, we derived default parameter values using a generic habituation experiment. We have used these parameters in pilot simulations that successfully capture Quinn's (1987) habituation data in a shape categorization tasks with 3- to 4-month-olds (Sirois & Mareschal, 2002a). Further work along these lines is currently ongoing.

Combined with the ability of the HAB model to capture Fantz's (1964) data as well, and especially the age-related changes, this breadth of coverage highlights the utility of neurocomputational models in bridging the brain–behavior gap.

In summary, the HAB model illustrates how computational models in developmental cognitive neuroscience can contribute to our understanding of behavior. With habituation, we show how opponent, complementary brain functions give rise to the crucial behavioral markers of interest, and how consideration about the maturation of the contribution of one of these functions allows for additional data coverage. The model, being explicit about representations and processes, allows for a detailed examination of how information is processed. This, in turns, leads to testable predictions that push the research agenda forward.

METHODS

Stimuli

Whereas Fantz (1964) showed infants a series of photographs collected from magazines, we presented the networks with small black and white bitmap images to represent 10 pairs of stimuli, where one stimulus is constant between pairs. Counterbalancing for position, this required 20 images, shown in Figure 10. The constant stimulus was represented by a cross, and the 10 variable stimuli were represented by bands.[5] Each pixel in an image served as input to the model, thus requiring $5 \times 10 = 50$ input units. To avoid the issue of how complex visual images are preprocessed and because Fantz's results should generalize beyond the specific set of stimuli that he used (and they have, see Roder et al., 2000), we have used simple stimuli that allow us to replicate the experiment and implement salience controls. Stimuli had equal salience, as every stimulus was a unique combination of five "live" bits or features. Moreover, the constant stimulus had varying levels of perceptual overlap or similarity with the variable stimuli: no shared feature with four stimuli, one shared feature with four stimuli, and three shared features with two stimuli. Each of the variable stimuli shared features with five other variable stimuli.[6]

Simulation Details

An external stimulus is presented to a bank of thalamic input units (see Figure 2). Their activations are forwarded to corresponding units in both the hippocampal and the cortical networks. These are autoassociator neural networks, which circulate activations internally such that a unit stimulates or inhibits every other unit in the local network. For a given input pattern, the hippocampal network learns to inhibit active units (leading to selective inhibition), whereas the cortical network learns to activate active units (leading to response enhance-
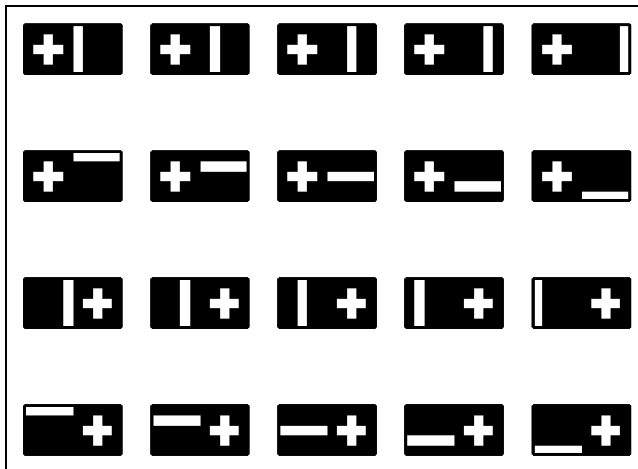
Figure 10. Stimuli used to simulate Fantz's (1964) experiment.

ment). Reciprocal connections between both networks allow hippocampal units to modulate the activity of corresponding units in the cortical network, and vice-versa. Both networks contribute to the response of the model to an external input. Activity on the output units is expected to direct the attention of the organism to specific aspects of the environment (see Sokolov, 1963). That is, the more active the response to a feature, the more processing is requested. Activity of output units serves as our measure of interest, which in a more complete system would drive what Sokolov called the orienting reflex.

When a pattern is initially presented to the model (i.e., the input units take the value of the corresponding features of the image, zero or one, as their activation value), the net input to both the hippocampal and cortical systems is equal to the input times an input scalar α, which determines the contribution of the external input relative to internal activity. Units in both systems compute their resulting activations as

$$a_i = \frac{1}{1 + \exp(-\text{net}_i)} \quad (1)$$

where a_i is the activation of unit i, and net_i is the net input reaching unit i. Equation 1 is a logistic function, which produces an s-shaped output (i.e., a smooth, threshold-like function with range $[0, 1]$).

Activations are then circulated within and between systems over a number of processing cycles. At any time $t + 1$, where t is the preceding processing step, the net input to the hippocampal system can be computed by

$$\text{net}_i(t+1) = \alpha I_i \cdot \text{WHC}_i \cdot C_i(t) + \sum_n \text{WHH}_{in} \cdot H_n(t) \quad (2)$$

where $\text{net}_i(t + 1)$ is the net input to unit i in the hippocampus at time $t + 1$, α is the input scalar, I_i is the activation of input unit i, WHC_i is the cortex to hippocampus weight for corresponding unit i, $C_i(t)$ is the activation of cortical unit i at time t, n is the number of inputs, WHH_{in} are the internal weights between hippocampal units i and n, and H_n is the activation of hippocampal unit n. Updated activations are obtained by using the results of Equation 2 with Equation 1.

Similarly, for cortical units, net input at time $t + 1$ is computed by

$$\text{net}_i(t+1) = \alpha I_i \cdot \text{WCH}_i \cdot H_i(t) + \sum_n \text{WCC}_{in} \cdot C_n(t) \quad (3)$$

where $\text{net}_i(t + 1)$ is the net input to unit i in the cortex at time $t + 1$, α is the input scalar, I_i is the activation of input unit i, WCH_i is the hippocampus to cortex weight for feature i, $H_i(t)$ is the activation of hippocampal unit i at time t, n is the number of inputs, WCC_{in} are the internal weights between cortical units i and n, and C_n is the activation of cortical unit n. Again, updated acti-

vations are obtained by using the results of Equation 3 with Equation 1.

Finally, net input to output units at any time t is computed with

$$\text{net}_i(t) = \text{WOH}_i \cdot H_i(t) + \text{WOC}_i \cdot C_i(t) \qquad (4)$$

where $\text{net}_i(t)$ is the net input to output unit i at time t, WOH_i is the hippocampus to output weight for feature i, $H_i(t)$ is the activation of hippocampal unit i at time t, WOC_i are the cortex to output weights for feature i, and C_i is the activation of cortical unit i. The updated activity of output units is obtained by using the results of Equation 4 in Equation 1.

The only modifiable weights in the model are the internal weights in both the hippocampal and cortical systems. These were modified on every processing cycle. For the hippocampal system, the learning rule was

$$\Delta w_{ij} = \lambda(\text{floor} - a_i)a_j \qquad (5)$$

where Δw_{ij} is the amount by which to change the connection weight between receiving unit i and sending unit j, λ is the learning rate, floor is the minimum value of the activation function (with Equation 1, it is 0), a_i is the activation of receiving unit i, and a_j is the activation of sending unit j. With such a rule, proposed by Kohonen (1988), the activity of a unit i becomes inhibited by any unit j when they share correlated activity.

For cortical connections, weights are changed using the delta rule, or Widrow–Hoff rule, expressed as

$$\Delta w_{ij} = \lambda(i_i - a_i)a_j \qquad (6)$$

where Δw_{ij} is the amount by which to change the weight between receiving unit i and sending unit j, λ is the learning rate, i_i the input value for feature i, a_i the activation of receiving unit i, and a_j the activation of sending unit j. With this rule, weight changes are progressively smaller as activations approach target values (i.e., the input).

The number of processing cycles on a given trial was set to 5. Networks were trained for 10 blocks of trials, where each block (called an epoch) consisted of a presentation of each of the input patterns in random order. An input scalar (α) of 1.5 was used for both cortical and hippocampal subsystems. The learning rates (λ) for the hippocampus and the cortex were 0.001 and 0.1, respectively. Reciprocal weights from hippocampus to cortex (WCH) and cortex to hippocampus (WHC) were set to 1. Output weights from the hippocampus (WOH) were set to 1, and output weights from the cortex (WOC) were set to 4. For the mature model (6-month-olds), WHC and WOC were set to 2 and 8, respectively. For the immature model (1-month-olds), these same parameters were set to 0.5 and 2.

These parameters were originally derived during pilot work (Sirois & Mareschal, 2002b). Learning rates were adapted to the specific size of input used in the current simulations. Maturational parameters were derived from a pilot simulation of this task, using orthogonal (as opposed to overlapping) stimuli (Sirois & Mareschal, 2003).

Notes

1. If researchers designed stimuli or events that differed solely on a conceptual basis (and, importantly, not at the perceptual level), then renewed responding to conceptually novel items would certainly support the suggestion that infants exhibit the underlying conceptual ability to distinguish such conceptually distinct events. Whether this is actually possible remains an open question (Sirois & Mareschal, 2002c).
2. We recognize that the actual shape of responses may vary from stimulus to stimulus, and between individuals. Thus, this point may refer only to the overall trend. McCall (1979) had raised the issue of whether monotonic, exponential decrease was a feature of responses for only a subset of infants. Recent work by Gilmore and Thomas (2002), using advanced mathematical models, has shown that McCall's results were statistical artifacts, and that infants who habituate have behavior that is best explained by a monotonic decreasing exponential function when noise is filtered out.
3. For clarity, we have combined the similar curves of 2–3 and 3–4 months in Fantz' report, weighing the original data points for unequal group size.
4. This is the average activation value of the 50 output units.
5. For our purposes, the different shapes are essentially illustrative. At this stage there is no boundary or shape detection built into the model. Each stimulus is merely a specific set of 5 "on" bits in a 5×10 field.
6. We have also replicated Fantz's findings using less realistic, orthogonal stimuli.

REFERENCES

Atkinson, J. (1984). Human visual development over the first 6 months of life: A review and a hypothesis. *Human Neurobiology, 3,* 61–74.

Bachevalier, J. (2001). Neural bases of memory development: Insights from neuropsychological studies in primates. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (pp. 365–370). Cambridge: MIT Press.

Bachevalier, J., Brickson, M., & Hagger, C. (1993). Limbic-dependent recognition memory in monkeys develops early in infancy. *NeuroReport, 4,* 77–80.

Baillargeon, R. (1987). Object permanence in 3 1/2- and 4 1/2-month-old infants. *Developmental Psychology, 23,* 655–664.

Baillargeon, R. (1999). Young infants' expectations about

hidden objects: A reply to three challenges. *Developmental Science, 2,* 115–132.

Bogartz, R. S., Shinskey, J. L., & Speaker, C. J. (1997). Interpreting infant looking: The event set * event set design. *Developmental Psychology, 33,* 408–422.

Bornstein, M. H., & Benasich, A. A. (1986). Infant habituation: Assessments of individual differences and short-term reliability at five months. *Child Development, 57,* 87–99.

Bronson, G. W. (1974). The postnatal growth of visual capacity. *Child Development, 45,* 873–890.

Bronson, G. W. (1991). Infant differences in rate of visual encoding. *Child Development, 62,* 44–54.

Clifton, R. K., Graham, F. K., & Hatton, H. M. (1968). Newborn heart-rate response and response habituation as a function of stimulus duration. *Journal of Experimental Child Psychology, 6,* 265–278.

Cohen, L. B., & Marks, K. S. (2002). How infants process addition and subtraction events. *Developmental Science, 5,* 186–201.

Colombo, J., Mitchell, D. W., Obrien, M., & Horowitz, F. D. (1987a). The stability of visual habituation during the 1st year of life. *Child Development, 58,* 474–487.

Colombo, J., Mitchell, D. W., Obrien, M., & Horowitz, F. D. (1987b). Stimulus and motoric influences on visual habituation to facial stimuli at 3 months. *Infant Behavior and Development, 10,* 173–181.

Dehaene, S., & Changeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neuroscience, 5,* 390–407.

Disney, A., & Calford, M. B. (2001). Neurosteroids mediate habituation and tonic inhibition in the auditory midbrain. *Journal of Neurophysiology, 86,* 1052–1056.

Fantz, R. L. (1964). Visual experience in infants: Decreased attention familiar patterns relative to novel ones. *Science, 146,* 668–670.

Gilmore, R. O., & Thomas, H. (2002). Examining individual differences in infants' habituation patterns using objective quantitative techniques. *Infant Behavior and Development, 25,* 399–412.

Graham, F. K. (1992). Attention: The heartbeat, the blink, and the brain. In B. A. Campbell, H. Hayne, & R. Richardson (Eds.), *Attention and information processing in infants and adults: Perspectives from human and animal research* (Vol. 8, pp. 3–29). Hillsdale, NJ: Lawrence Erlbaum Asociates.

Graham, F. K., Berg, K. M., Berg, K. W., Jackson, J. C., Hatton, H. M., & Kantowitz, S. R. (1970). Cardiac orienting responses as a function of age. *Psychonomic Science, 19,* 363–365.

Graham, F. K., & Clifton, R. K. (1966). Heart-rate change as a component of the orienting response. *Psychological Bulletin, 65,* 305–320.

Graham, F. K., Clifton, R. K., & Hatton, H. M. (1968). Habituation of heart rate response to repeated auditory stimulation during the first five days of life. *Child Development, 39,* 35–52.

Groves, P. M., & Thompson, R. F. (1970). Habituation: A dual-process theory. *Psychological Review, 77,* 419–450.

Haith, M. M. (1998). Who put the cog in infant cognition? Is rich interpretation too costly? *Infant Behavior and Development, 21,* 167–179.

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.

Jankowski, J. J., & Rose, S. A. (1997). The distribution of visual attention in infants. *Journal of Experimental Child Psychology, 65,* 127–140.

Jankowski, J. J., Rose, S. A., & Feldman, J. F. (2001). Modifying the distribution of attention in infants. *Child Development, 72,* 339–351.

Jiang, T., Haxby, J. V., Martin, A., Ungerleider, L. G., & Parasuraman, R. (2000). Complementary neural mechanisms for tracking items in human working memory. *Science, 287,* 643–646.

Johnson, M. H. (1990). Cortical maturation and the development of visual attention in early infancy. *Journal of Cognitive Neuroscience, 2,* 81–95.

Johnson, M. H. (1997). *Developmental cognitive neuroscience: An introduction.* Cambridge; Oxford: Blackwell Publishers.

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences, 4,* 138–147.

Kohonen, T. (1988). *Self-organization and associative memory.* Berlin: Springer-Verlag.

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

Mareschal, D., & Quinn, P. C. (2001). Categorization in infancy. *Trends in Cognitive Sciences, 5,* 443–450.

Mareschal, D., & Johnson, S. P. (2002). Learning to perceive object unity: A connectionist account. *Developmental Science, 5,* 151–172.

Mayes, L. C., & Kessen, W. (1989). Maturational changes in measures of habituation. *Infant Behavior and Development, 12,* 437–450.

McCall, R. B. (1979). Individual differences in the pattern of habituation at 5 and 10 months of age. *Developmental Psychology, 15,* 558–568.

McKee, R. D., & Squire, L. R. (1993). On the development of declarative memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 397–404.

Miller, E. K., & Desimone, R. (1994). Parallel neural mechanisms for short-term memory. *Science, 263,* 520–522.

Mishkin, M., Vargha-Khadem, F., & Gadian, D. G. (1998). Amnesia and the organization of the hippocampal system. *Hippocampus, 8,* 212–216.

Montaldi, D., Spencer, T., Alvarez, A., Roberts, N., & Mayes, A. R. (2003). Strength of item familiarity memory is mediated by the perirhinal cortex not the hippocampus. *Neuroimage, 15,* S195.

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

Nelson, C. A. (2002). The ontogeny of human memory: A cognitive neuroscience perspective. In M. H. Johnson, Y. Munakata, & R. O. Gilmore (Eds.), *Brain development and cognition: A reader* (2nd ed., pp. 151–178). Malden, MA: Blackwell Publishers.

Quinn, P. C. (1987). The categorical representation of visual pattern information by young infants. *Cognition, 27,* 145–179.

Ritz, E. G., Woodruff, A. B., & Fagen, J. W. (1984). Short-term and long-term stability of habituation rate and looking time in 4-month-old infants. *Journal of Genetic Psychology, 144,* 285–286.

Roder, B. J., Bushnell, E. W., & Sasseville, A. M. (2000). Infants' preferences for familiarity and novelty during the course of visual processing. *Infancy, 1,* 491–507.

Rolls, E. T., & Treves, A. (1998). *Neural networks and brain function.* Oxford: Oxford University Press.

Rose, S. A., & Feldman, J. F. (1987). Infant visual attention: Stability of individual differences from 6 to 8 months. *Developmental Psychology, 23,* 490–498.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2002). Processing speed in the 1st year of life: A longitudinal study of preterm and full-term infants. *Developmental Psychology, 38,* 895–902.

Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology, 18,* 704–713.

Schoener, G., & Thelen, E. (in press). Using dynamic field theory to rethink infant habituation. *Psychological Review.*

Seress, L. (2001). Morphological changes of the human hippocampal formation from midgestation to early childhood. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (pp. 45–58). Cambridge: MIT Press.

Sirois, S. (2004). Autoassociator networks and insights into infancy. *Developmental Science, 7,* 133–140.

Sirois, S., Buckingham, D., & Shultz, T. R. (2000). Artificial grammar learning by infants: An auto-associator perspective. *Developmental Science, 3,* 442–456.

Sirois, S., & Mareschal, D. (2002a). *A connectionist process model of infant habituation.* Paper presented at the International Conference on Infant Studies (ICIS2002), Toronto, Canada.

Sirois, S., & Mareschal, D. (2002b). Infant habituation: A review of current computational models and a new proposal. In J. A. Bullinaria & W. Lowe (Eds.), *Proceedings of the seventh neural computation and psychology workshop: Connectionist models of cognition and perception* (pp. 90–103). Singapore: World Scientific.

Sirois, S., & Mareschal, D. (2002c). Models of habituation in infancy. *Trends in Cognitive Sciences, 6,* 293–298.

Sirois, S., & Mareschal, D. (2003). *A neurocomputational model of infant habituation.* Paper presented at the 2003 Biennal Meeting of the Society for Research in Child Development, Tampa, Florida.

Sokolov, E. N. (1963). *Perception and the conditioned reflex.* Oxford: Pergamon.

Sokolov, E. N. (1975). The neuronal mechanisms of the orienting reflex. In E. N. Sokolov & O. S. Vinogradova (Eds.), *Neuronal mechanisms of the orienting reflex* (pp. 217–235). Hillsdale, NJ: Erlbaum.

Spelke, E. S. (1998). Nativism, empiricism, and the origins of knowledge. *Infant Behavior and Development, 21,* 181–200.

Spelke, E. S. (1999). Innateness, learning and the development of object representation. *Developmental Science, 2,* 145–148.

Spelke, E. S. (2000). Core knowledge. *American Psychologist, 55,* 1233–1243.

Spelke, E. S., Breinlinger, K., Macomber, J., & Jacobson, K. (1992). Origins of knowledge. *Psychological Review, 99,* 605–632.

Suzuki, W. A., Miller, E. K., & Desimone, R. (1997). Object and place memory in the macaque entorhinal cortex. *Journal of Neurophysiology, 78,* 1062–1081.

Thomas, M. S. C., & Karmiloff-Smith, A. (2003). Connectionist models of development, developmental disorders and individual differences. In R. J. Sternberg, J. Lautrey, & T. Lubart (Eds.), *Models of intelligence: International perspectives* (pp. 133–150). Washington, DC: American Psychological Association.

Thompson, R. F., & Spencer, W. A. (1966). Habituation: A model phenomenon for the study of neuronal substrates of behavior. *Psychological Review, 73,* 16–43.

Thorpe, W. H. (1956). *Learning and instinct in animals.* London: Methuen.

Utsunomiya, H., Takano, K., Okazaki, M., & Mitsudome, A. (1999). Development of the temporal lobe in infants and children: Analysis by MR-based volumetry. *American Journal of Neuroradiology, 20,* 717–723.

van Groen, T., Miettinen, P., & Kadish, I. (2003). The entorhinal cortex of the mouse: Organization of the projection to the hippocampal formation. *Hippocampus, 13,* 133–149.

Vinogradova, O. S. (1975). The hippocampus and the orienting reflex. In E. N. Sokolov & O. S. Vinogradova (Eds.), *Neuronal mechanisms of the orienting reflex* (pp. 128–154). Hillsdale, NJ: Erlbaum.

Wang, D. (1995). Habituation. In M. A. Arbib (Ed.), *Handbook of brain theory and neural networks* (Vol. 15, pp. 441–444). Cambridge: MIT Press.

Westermann, G., & Mareschal, D. (2004). From parts to wholes: Mechanisms of development in infant visual object processing. *Infancy, 5,* 131–151.

Wynn, K. (1995). Infants possess a system of numerical knowledge. *Current Directions in Psychological Science, 4,* 172–177.

Zelazo, P. R., Weiss, M. J. S., & Tarquinio, N. (1991). Habituation and recovery of neonatal orienting to auditory stimuli. In M. J. S. Weiss & P. R. Zelazo (Eds.), *Newborn attention: Biological constraints and the influence of experience* (pp. 120–141). Norwood, NJ: Ablex.