ARTICLE WITH PEER COMMENTARIES AND RESPONSE

Infant perseveration and implications for object permanence theories: A PDP model of the AB̄ task

Yuko Munakata

University of Denver

Abstract

From the earliest ages at which infants search for hidden objects, they make the AB̄ error, searching perseveratively at previous rather than current hiding locations (Piaget, 1954). This paper presents a parallel distributed processing (PDP) model that instantiates an explicit set of processing mechanisms to account for a large and diverse set of data on infants' AB̄ errors. The model demonstrates how basic processes – the formation of latent memory traces and their interaction with developing active memory traces – can provide a unifying framework for understanding why and when infants perseverate. Novel predictions from the model are discussed, together with its challenges for theories that posit a concept of object permanence in the first year of life.

In the AB̄ task, devised by Piaget (1954), infants watch an experimenter hide an object in one location (*A*). They are typically allowed to search for the object after a short delay, and this procedure is repeated some number of times. Infants then watch the experimenter hide the object in a new location (*B*). Following a short delay, infants often search perseveratively at *A*. This striking AB̄ error occurs even when infants merely observe hidings at *A*, and are only permitted to search after the toy is hidden at *B* (Diamond, 1983; Butterworth, 1974; Evans, 1973).

Why do infants perseverate, and what do their perseverative errors reveal about their representations of the world? In response to these questions, researchers have established a rich empirical database documenting influences upon the AB̄ error and proposed numerous theories about infants' understanding of object permanence (e.g., Piaget, 1954; Gratch, Appel, Evans, LeCompte, & Wright, 1974; Butterworth, 1977; Cummings & Bjork, 1983b; Diamond, 1985; Horobin & Acredolo, 1986; Wellman, Cross, & Bartsch, 1986; Harris, 1989). Table 1 summarizes a diverse collection of the empirical findings (described in more detail

below). No existing theory can account for them within a unified framework. In this paper, I propose such a theory within the adaptive process framework (Munakata, McClelland, Johnson, & Siegler, 1997) and introduce a parallel distributed processing (PDP) simulation based upon this theory. I show how the model simulates infant performance in the standard AB̄ task and in each of the task variants in Table 1, and present the model's novel empirical predictions. In each case, I discuss the mechanisms underlying the network's performance and their hypothesized analogs in the infant.

Wellman *et al.* (1986) conducted a meta-analysis that demonstrated the robustness of the AB̄ phenomenon in the standard search paradigm and revealed four factors that can reduce the production of AB̄ errors: older participants and shorter delays (e.g., Diamond, 1985), *more* hiding locations (e.g., Cummings & Bjork, 1983a; 1983b), and distinct hiding locations (e.g., Bremner, 1978b). Wellman *et al.* (1986) argued that no existing theory at the time of their meta-analysis could account for these findings. They proposed an alternate account in which infants understand the permanence of objects, but

Table 1 *A$\overline{\text{B}}$ tasks: The network simulates infant performance on all of these task versions and makes novel predictions about behavior. Behavior indicated as 'Better Performance' is relative to that in the standard 2-location task.*

A$\overline{\text{B}}$ Version	Behavior
2-location	Perseverative reaching (A$\overline{\text{B}}$ errors) (Piaget, 1954).
Gaze/Expectation	Better performance (Hofstadter & Reznick, 1996; Ahmed & Ruffman, 1997).
A-Observation	A$\overline{\text{B}}$ errors (Diamond, 1983; Butterworth, 1974; Evans, 1973).
At various ages	Better performance as age increases (Diamond, 1985; Wellman *et al.*, 1986).
At various delays	Better performance as delay decreases (Diamond, 1985; Wellman *et al.*, 1986).
Multiple-location	Better performance (Cummings & Bjork, 1983a, 1983b; Wellman *et al.*, 1986).
Distinctive covers	Better (Bremner, 1978b; Wellman *et al.*, 1986).
Covers only	A$\overline{\text{B}}$ errors (Smith *et al.*, 1995; Munakata, 1997).
A-cover, B-toy	Better performance (Munakata, 1997).
A-toy, B-cover	A$\overline{\text{B}}$ errors (Munakata, 1997).

search perseveratively due to the availability of two search strategies: *direct-finding* (searching directly where an object vanished) and *inferred-location* finding (searching based on the inferred position of an object, based on its prior movements). Inferred-location finding can lead to perseverative search and becomes more likely with longer delays (which inhibit direct-finding), fewer hiding locations (which lead infants to attempt to integrate *A* and *B* in their inferences), and identical hiding locations (the confusability of which inhibits direct-finding). Perseverative search decreases with age because infants get better at making accurate inferences.

Although Wellman *et al.*'s (1986) account was consistent with the findings from their meta-analysis, Smith, McLin, Titzer, and Thelen (1995) have since demonstrated that hidden toys are unnecessary for the A$\overline{\text{B}}$ error; 8- and 10-month-old infants perseverate when simply reaching to visible covers over empty containers. In Smith *et al.*'s (1995) studies, an experimenter directed infants' attention to one of two identical covers over empty containers and allowed infants to reach following a short delay. Nothing was ever placed inside the containers, and nothing was ever hidden. On *A* trials, infants tended to reach to the cover to which the experimenter had directed their attention. On *B* trials, in which the experimenter directed attention to the other cover, infants tended to perseverate in reaching to *A*. That is, infants made the A$\overline{\text{B}}$ error with completely visible objects. As Smith *et al.* (1995) argued, these provocative results pose a challenge to almost all theories of the A$\overline{\text{B}}$ error because they are based upon infants'

knowledge and inferences about or memory for hidden toys, but hidden toys are unnecessary for the error.

Smith *et al.* (1995) also made the stronger claim that hidden toys are irrelevant to infant behavior in the A$\overline{\text{B}}$ task. They argued that A$\overline{\text{B}}$ errors may be understood in terms of infants' motor histories, not their representations of hidden objects. However, their studies demonstrated only that repeated exposure to a stimulus – hidden toy under cover or cover alone – at *A* led to perseveration when the same stimulus was presented at *B*; infants may have nonetheless represented covers with and without toys underneath in different ways. Munakata (1997) tested this possibility in two experiments. In the first, the experimenter presented *A* trials as in Smith *et al.* (1995), with only a visible cover at *A*. On *B* trials, the experimenter presented at *B* either a visible cover (cover condition) or a toy that was then hidden under the cover (toy condition). Infants reached perseveratively to *A* in the cover condition (replicating Smith *et al.*, 1995), but not in the toy condition, indicating that infants can distinguish covers with and without toys underneath on *B* trials. The second study was identical except that on all *A* trials the experimenter presented a toy and then hid it under the *A* cover. Infants made similar A$\overline{\text{B}}$ errors on *B* trials in the cover and toy conditions. The difference in *B*-trial performance between the two experiments indicated that infants distinguished covers with and without toys underneath on *A* trials, although their motor histories were the same in the two cases. Thus, although Smith *et al.* (1995) showed that infants perseverate with and without hidden toys, Munakata's (1997) studies showed that infants nonetheless represent hidden toys differently from visible covers alone, on both *A* and *B* trials. Any complete model of the A$\overline{\text{B}}$ error must account for influences of both motor history and hidden objects.

Lastly, many researchers have criticized Piagetian tests of object permanence because they require the use of 'performance' factors (such as reaching) to demonstrate an underlying 'competence' (such as a concept of object permanence); infants may have the underlying competence but fail Piagetian tasks due to performance limitations. Under this assumption, researchers have designed clever experiments that demonstrate earlier signs of competence. For example, a reduced production of perseverative errors has been observed in gaze and expectation variants of the A$\overline{\text{B}}$ task. Diamond (1985) first noted that infants occasionally reach to the wrong location while looking at the correct location, and often reach to the wrong location without looking in it and then immediately reach to the correct location. Hofstadter and Reznick (1996) confirmed that when infants' looking and reaching behaviors differ in this task, the

looking response is more accurate. And, infants make fewer errors in gaze variants of the A$\overline{\text{B}}$ task in which they observe hidings at *A* and *B* without ever reaching (Hofstadter & Reznick, 1996; Matthews, 1992; Lecuyer, Abgueguen, & Lemarie, 1992). Finally, in violation-of-expectation variants of the A$\overline{\text{B}}$ task, 8–12 month-old infants look longer when a toy hidden at *B* is revealed at *A* than when it is revealed at *B*, following delays at which they would nonetheless search perseveratively at *A* (Ahmed & Ruffman, 1997; Baillargeon & Graber, 1988; Baillargeon, DeVos, & Graber, 1989).

These demonstrations of sensitivity to an object's new hiding location are commonly treated as evidence that infants 'know' where toys are hidden in the A$\overline{\text{B}}$ task, but search perseveratively due to deficits external to their knowledge representations. For example, Diamond (1991) argued from infants' correct gazing that they 'know where the toy is hidden, even though they reach back to *A* anyway….They seem to understand the concept of object permanence' (p. 85). Diamond (1985) proposed that infants' inability to inhibit a conditioned reaching response to *A* causes the A$\overline{\text{B}}$ error, and this interpretation has been widely accepted (e.g., Hofstadter & Reznick, 1996; Ahmed & Ruffman, 1997).[1] Similarly, Baillargeon and colleagues argued that infants' success in violation-of-expectation variants of the A$\overline{\text{B}}$ task showed that infants remembered a hidden object's new location and 'cast serious doubts on accounts that attribute infants' search errors to inadequate memory mechanisms' (Baillargeon *et al.*, 1989, p. 345). They attributed the A$\overline{\text{B}}$ error to deficits in problem-solving abilities (Baillargeon & Graber, 1988; Baillargeon *et al.*, 1989). They distinguished *reactive* problem-solving (responding immediately and without conscious reasoning) from *planful* problem-solving (based on active reasoning or computation); younger infants tend to be more reactive, which leads them to 'run off' previous solutions and respond perseveratively. Such accounts tend to treat early signs of sensitivity as evidence for knowledge taking the form of all-or-none principles, while infants' simultaneous (and subsequent) failures are considered to be outside the theoretical domain of interest (see discussions in Munakata *et al.*, 1997; Siegler, 1996; Thelen & Smith, 1994; Smith & Thelen, 1993; McClelland, 1994; Karmiloff-Smith, 1992; Fischer & Bidell, 1991; Johnson & Morton, 1991; Flavell, 1984).

While some findings appear to be consistent with these accounts, most of the evidence is equivocal, and there are some contrary findings. For example, Diamond's (1985) inhibition theory seems to be supported by the finding that infants occasionally search at *A* even when the object is visible at *B* (Butterworth, 1977; Harris, 1974; Bremner & Knowles, 1984). However, Sophian and Yengo (1985) showed that such errors reflected random incorrect responses rather than true A$\overline{\text{B}}$ errors: in an experiment including a third, control location, infants were as likely to err when the toy was visible at *B* by searching at the control location as at the *A* location. Many of the previously reviewed findings also challenge the inhibition and problem-solving accounts. For example, if infants make A$\overline{\text{B}}$ errors because they are conditioned or reactive, they should not respond perseveratively in *A*-observation conditions, when they have not established reaching to *A* as a solution, but they do. And, infants should become less reactive and more planful given more time, and so should make fewer perseverative errors with longer delays, but instead they make more.

Available ancillary deficit theories thus fail to account for a range of infant behaviors, challenging the tenet that infants' early successes indicate the existence of knowledge in the form of principles, with explanations of failures relegated to uninteresting ancillary deficits. Instead, it seems likely that infants' successes *and* failures provide insight into the nature of the underlying representations guiding their behavior. The adaptive process approach provides a framework for exploring such possibilities.

The Active-Latent Account

To account for infants' successes and failures in object permanence tasks, Munakata *et al.* (1997) proposed an adaptive process framework, in which the knowledge underlying infants' behaviors is viewed as graded in nature, evolving with experience, and embedded in specific processes underlying overt behavior. Within this framework, we argued that infants may use relatively weak internal representations to show sensitivity to hidden objects in certain paradigms (e.g., visual habituation) early in life, while other tasks (e.g., search) require stronger representations that develop later in life. Using PDP simulations, we demonstrated how a system could learn to form internal representations of objects based on experience with the world, how these representations could be graded in nature, and how such gradedness could support early signs of sensitivity in visual habituation measures but not in reaching.

Within this adaptive process framework, I propose a theory in which the continued strengthening of represen-

---

[1] Although Diamond's (1991) account also stressed the importance of memory demands in the standard A$\overline{\text{B}}$ task, it focused on inhibition to explain the discrepancies between infants' looking and reaching.

tations leads to eventual success in the AB̄ task. According to this theory:

- The AB̄ error arises based on a competition between 'latent' memory traces for *A* and 'active' memory traces for *B*.
- Latent memory traces result when infants change their biases toward a stimulus after processing it, so that they may respond differently to the stimulus on subsequent presentations.
- Active memory traces result when infants actively maintain representations of a stimulus. Unlike latent traces, such active representations may be accessible in the absence of subsequent presentations of the stimulus.
- Infants' performance on the AB̄ task and its variants can be understood in terms of the relative strengths of latent and active memory traces, and the effects of task variations on the strength of these traces.

According to this account, infants succeed in search tasks involving a single hiding location before the AB̄ task because the former have no competing latent traces and so do not require as strong an ability to maintain active representations. The increasing ability to maintain active traces leads to improvements in performance in the AB̄ task.

This active-latent account shares several features with existing accounts of the AB̄ error. Many existing accounts similarly 1) posit the AB̄ error to arise based on a competition between two kinds of information (e.g., Butterworth, 1977; Harris, 1986; Wellman *et al.*, 1986); 2) describe something akin to the latent element of the active-latent account (e.g., a behavioral tendency toward *A* (Diamond, 1985) or a deep attractor for *A* (Thelen & Smith, 1994)); or 3) describe something akin to the active element of the active-latent account (e.g., short-term memory for *B* (Harris, 1986)). The active-latent distinction also maps onto cognitive distinctions proposed outside of the AB̄ realm. Active memories are related to notions of recall, working, and short-term memory, whereas latent memories are related to notions of recognition, habit, and long-term memory. The active-latent distinction is also related to the verbatim-gist distinction in the fuzzy-trace theory of Brainerd and colleagues (Brainerd & Reyna, 1990; Brainerd & Kingma, 1985; Brainerd & Reyna, 1993), in which verbatim memories decay more rapidly such that gist memories form the basis of much of behavior. Thus, the active-latent account of the AB̄ error shares much with existing cognitive distinctions and with other accounts of the AB̄ error.

However, the active-latent account differs from previous accounts of the AB̄ error in three fundamental

ways – in its treatment of infant knowledge, its specificity about underlying mechanisms, and its neural motivations. First, infants' knowledge of an object's location is not treated as a reified entity, disembodied from underlying processing mechanisms. Instead, these processing mechanisms – active and latent – are viewed as giving rise to the knowledge representations governing behavior. Second, the active-latent theory provides an explicit mechanistic account of performance in the AB̄ task. The distinction between active and latent traces and the dynamic between them is specified and implemented in the form of a PDP network. Third, the active-latent distinction at the heart of the proposed account is motivated by a distinction at the neural level, described below. These three distinguishing features of the active-latent account contribute to its ability to provide a unified framework for understanding infants' perseverative behavior.

A collection of behavioral and physiological data motivates the active-latent distinction central to the proposed theory of the AB̄ error, and indicates that active traces require the prefrontal cortex whereas more posterior areas of the brain may suffice for latent traces. Miller and Desimone (1994) distinguished two ways in which neurons can 'remember': through sustained firing for a stimulus, or through changes in firing thresholds or synapses that affect neurons' subsequent firing to a stimulus. Miller and Desimone (1994) showed that when monkeys performed a task that required memory for a specific stimulus item, neurons in the prefrontal cortex showed sustained firing for the stimulus, across intervening stimuli. This active memory is consistent with a number of neural recording and imaging experiments in the prefrontal cortex (e.g., Fuster, 1989; Goldman-Rakic, 1987; Cohen, Perlstein, Braver, Nystrom, Noll, Jonides, Smith, 1997). In contrast, on an easier task that required memory for any familiar stimulus, monkeys appeared to solve the task on the basis of some form of latent memory in neurons in the inferotemporal cortex; these neurons showed no maintained firing signal, but showed reduced firing when familiar stimuli were presented again. Monkeys appeared to simply process stimuli and as a result, laid down latent memory traces for them, resulting in facilitated processing (i.e., reduced firing) when they were repeated. Steinmetz, Connor, Constantinidis, and McLaughlin (1994) demonstrated a similar active-latent distinction for memories of spatial locations in neurons in prefrontal and posterior parietal cortex. Finally, humans with frontal lobe damage show deficits in working memory, but are unimpaired in discriminating novel and familiar stimuli (see Petrides, 1989 for review). Such recognition memory might depend on

latent memory traces that do not require the prefrontal cortex.

According to the active-latent account, infants' increasing ability to maintain active traces – dependent on the development of prefrontal cortex – leads to success in the A$\overline{\text{B}}$ task. Consistent with this, lesions of the prefrontal cortex (but not of parietal cortex) impair infant and adult rhesus monkeys' performance on the A$\overline{\text{B}}$ task (Diamond & Goldman-Rakic, 1989, 1986).

Simulation Modelling

In the PDP framework, latent traces take the form of changes to connection weights between network processing units, and active traces take the form of sustained activations of units. For example, the connection between an input unit and an internal representation unit for location A might be strengthened each time a stimulus is presented at A and subsequently represented. This change in connection strength would serve as a latent memory trace for the A location. In contrast, an active memory trace for A would take the form of sustained activity on the units for location A. These active memory traces can be graded in nature, and can become stronger as changes are made to the underlying connection weights. PDP simulations can thus provide a useful tool for exploring the competition between active and latent traces (for contributions of PDP simulations to understanding other developmental phenomena, see Elman, Bates, Karmiloff-Smith, Johnson, Parisi, & Plunkett, 1996).

For example, Cohen, Dunbar, and McClelland (1990) proposed a model of the Stroop task based on the distinction between active and latent traces (which they distinguished as controlled and automatic processing). In the Stroop task, subjects name the ink color or read the word for color word stimuli (e.g., the word 'green') printed in a certain color of ink (e.g., red ink). Subjects are faster and more accurate at reading than at color-naming, and conflicting words interfere with color-naming but conflicting ink color does not interfere with reading. In Cohen *et al.*'s (1990) PDP model of performance in the Stroop task, latent traces take the form of color-naming and reading pathways, and active traces take the form of activity on attention units corresponding to stimulus features (ink color or word). Weaker pathways are more dependent on modulation from the attention units, such that less practiced tasks are more dependent on active traces and more susceptible to interference. As I will show, the same principle applies to the A$\overline{\text{B}}$ model's performance.

In addition, Dehaene and Changeux (1989) proposed

a PDP model of A$\overline{\text{B}}$ behavior based on the distinction between active and latent traces (which they distinguished as 'memory' and 'synaptic efficacies'). A more detailed description of that model is provided later in this paper, together with a comparison of it and the current model. Briefly, I note here that the current model avoids three limitations of Dehaene and Changeux's (1989) model: 1) an abrupt switching on of memory units, 2) poor matches to the data (their model either fails to show sensitivity to delay or fails to show true perseverative responding), and 3) the use of a questionable reinforcement signal for learning. I now turn to consideration of my model of the A$\overline{\text{B}}$ task.

Architecture

The network is comprised of six layers: three input, one hidden, and two output (Figure 1). The three input layers encode information about: 1) location (three units corresponding to locations A, B, and C), 2) cover type (two units corresponding to C1, the default cover type, and C2, a different cover type), and 3) toy type (two units corresponding to T1, the default toy, and T2, a different toy type). The hidden and output layers each contain three units corresponding to locations A, B, and C. One output layer indicates the network's reaching response; the other indicates the network's gaze response or expectation. The single difference between the two output layers is the frequency of their responses (i.e., the updating of the unit activity) during the A$\overline{\text{B}}$ task. The gaze/expectation layer responds to every input, while the reaching layer responds only to inputs corresponding to a stimulus within 'reaching distance'.

This updating constraint on the output layers is meant to capture the different frequencies of infants' reaching
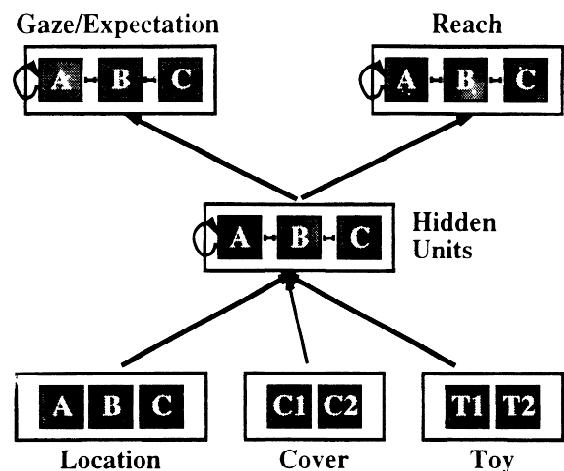

Figure 1 *The A$\overline{\text{B}}$ network.*

and gaze/expectation during the $A\overline{B}$ task. Reaching is permitted at only one point during each trial – when the apparatus is moved to within the infant's reach. In contrast, nothing prevents infants from forming expectations (which may underlie longer looking to impossible events) throughout each trial. Similarly, although infants' gaze is sometimes restricted during $A\overline{B}$ experiments, infants nonetheless have more opportunities to gaze than to reach. As I will show in the simulations, more frequent responses can change the dynamic between active and latent traces, resulting in dissociations between looking and reaching.[2]

Connectivity

Each of the three input layers is fully connected to the hidden layer, and the hidden layer is fully connected to each of the two output layers. The hidden and output layers have inhibitory connections between units and self-recurrent excitatory connections back to each unit. The recurrent connections are largely responsible for the network's ability to maintain active representations; units that are active tend to remain active when they receive their own activity as input through sufficiently large weights. I have proposed that a critical aspect of development is the gradual improvement in this ability to maintain active representations. Therefore, to test networks of different ages in the $A\overline{B}$ task, I manipulate the strength of the network's recurrent connections, with 'older' networks having higher recurrence.[3] Such increases in the strength of recurrent weights might correspond to the strengthening of connections within

the prefrontal cortex and between the prefrontal cortex and more posterior areas of the brain.

The network's initial connectivity includes a bias to respond appropriately to location information, for example, to look to location *A* if something is presented there. Infants appear to enter $A\overline{B}$ experiments with such biases. The bias is implemented in the network through relatively large weights (0.7) from input units to their corresponding hidden units and from hidden units to their corresponding output units, and relatively small weights (0.3) from input and hidden units to their non-corresponding hidden and output units, respectively. So, if the input unit for location *A* is activated, the *A* hidden unit will become active and then the *A* output units will become active. The network's initial connectivity includes no bias about the locations of cover types and toy types; each of these input units has equal weight (0.3) to each hidden unit. Inhibitory weights were fixed at -0.6.

Stimuli

The stimulus set is comprised of ten versions of the $A\overline{B}$ task (Table 1). Each version of the $A\overline{B}$ task consists of four pre-trials (corresponding to the 'practice' trials typically provided at the start of an experiment to induce infants to reach to *A*), two *A* trials, and one *B* trial.[4] Each trial consists of four segments, corresponding to the experimental segments of an $A\overline{B}$ trial as follows:

start: Covers sit in place on apparatus, before experimenter draws infant's attention to particular location (1 time step).
presentation: Experimenter draws infant's attention to one location in apparatus (3–4 time steps, depending on whether experimenter presents cover alone or toy and cover).
delay: Apparatus sits with covers in place (1 time step in pre-trials; 0–5 time steps in *A* and *B* trials).
choice: Experimenter presents apparatus with covers in place for infant's response (1 time step). Reaching permitted only during this segment.

During each segment, patterns of activity are presented to the input units corresponding to the visible aspects of the stimulus event; all other input units have no activity. The levels of input activity represent the salience of aspects of the stimulus, with more salient aspects (e.g. a

---

[2] In Munakata *et al.*'s (1997) simulations of 3–8 month-olds' performance in single hiding location tasks, the greater sensitivity of the looking system to graded representations resulted from delayed and slowed learning in the reaching system. Although the later mastery of the reaching system might also play a role in the looking-reaching task-dependency in the $A\overline{B}$ task, this factor may contribute less because infants tested in the $A\overline{B}$ paradigm are typically older (and thus have more experience reaching) than infants in the single-location looking and reaching tasks. One might ask whether different frequencies of reaching vs. expectation-formation during single-location tasks might also contribute to this looking-reaching task-dependency. I suspect that they could. In addition, the different frequencies of these behaviors in general (outside of experimental tasks) may contribute to their differential sensitivity, so the two explanations of different frequencies of behaviors in particular tasks and differential sensitivities of behaviors may be closely related.

[3] Other parameters may also influence the network's ability to maintain active representations. What is critical to the proposed developmental theory is the effect of representation strength on performance, not the particular parameter underlying changes in representations strength. Future simulation and neurophysiological work may help to further specify the nature of these underlying mechanisms.

[4] Pre-trials at *A* are not necessary for the occurrence of the $A\overline{B}$ error in infants (Diamond, 1985; Smith *et al.*, 1995) or in the model, though they do increase the frequency of $A\overline{B}$ errors in both infants (Smith *et al.*, 1995) and the model.

toy that the experimenter waves) producing more activity.

Figure 2 shows the input patterns for the segments of an *A* trial in the standard 2-location A$\overline{\text{B}}$ task. As in all trials, the *start*, *delay*, and *choice* segments consist of similar patterns of activity because the same aspects of the stimulus are visible (in this case, locations *A* and *B* and the default cover type, C1); the only difference is in the level of unit activity. The relevant input units are activated to 0.15 at *start*, 0.4 during *delay*, and 0.6 at *choice*. The low activity during start represents the low salience of the stimulus before the experimenter manipulates the stimulus on that trial; the slightly higher activity during delay represents the greater salience of the stimulus following the experimenter's presentation of certain aspects of the stimulus, and the even higher activity during choice represents the increased salience of the stimulus when it is presented for a response.

The *presentation* consists of relatively high levels of activity across presented aspects of the stimulus event, and relatively low levels of activity across visible but non-presented aspects. In trials in which a toy is hidden (as in the standard 2-location *A* trial shown in Figure 2), the *presentation* consists of a toy-presentation followed by a cover-presentation. During toy-presentation, input units T1 and *A* are activated to 1, and C1 and *B* are activated to 0.1. In the cover-presentation following toy-

presentation, C1 and *A* are activated to 0.5 (reflecting the lower salience of a cover relative to a toy), T1 is activated to 0.5 (reflecting the partial visibility of the toy as the cover is put into place), and *B* is activated to 0.1. The toy-presentation segment is presented for three time steps, and the subsequent cover-presentation for one time step.

Figure 3 shows the input patterns for the segments of an *A* trial in three other versions of the A$\overline{\text{B}}$ task: 3-location, distinctive covers, and covers-only. The input pattern for the cover-presentation in the covers-only version is identical to that following toy-presentation in the standard task (Figure 2); however, the cover-presentation in the covers-only version is presented for three time steps. In all versions of the A$\overline{\text{B}}$ task, pre-trials and A-trials have identical input patterns in each trial segment, but *A*-trials typically have longer delay segments. *B*-trial *delay* segments are always the same length as those of *A*-trials in the same experiment. The corresponding input patterns for *B* trial segments are similar to those in *A* trials, but the *B* unit (rather than the *A* unit) receives relatively high activity during presentation.

Activation Function

In all of the simulations, the net input to units was computed as:

$$net_i = \sum_j a_j w_{ij} + \theta \qquad (1)$$

where w_{ij} is the weight from unit$_j$ to unit$_i$ and θ is the bias weight. θ was fixed at -0.6 so that units were biased to have low activation states. Units could take on continuous activation values between 0 and 1. Activation levels were updated according to the sigmoidal logistic activation function, using a gain (γ) of 5:

$$sig(net_i) = \frac{1}{1 + e^{-\gamma net_i}} \qquad (2)$$

Activation-based time-averaging was performed using a step size (λ) of 0.2, according to the following functions:

$$\Delta a(t) = \lambda(sig(net_i) - a(t-1)) \qquad (3)$$

$$a(t) = a(t-1) + \Delta a(t) \qquad (4)$$

where *a* is the time-averaged activity of the unit. Activations settled over ten cycles per time step. The activity of the hidden and output units was recorded at the end of each time step, and the output activity at *choice* was treated as the network's response for that trial. The
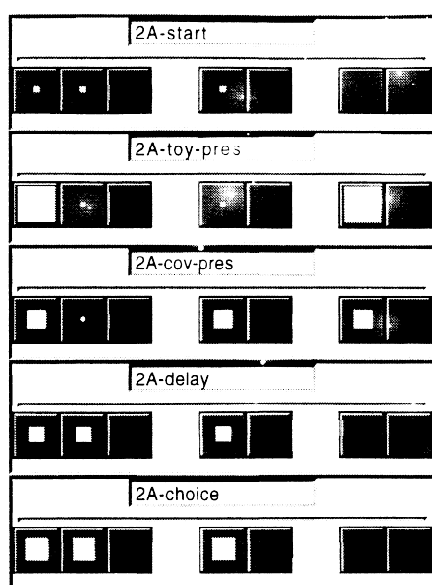

Figure 2 *Input patterns for five A-trial segments (*start, toy* presentation, cover presentation, delay, and choice) in the standard 2-location A$\overline{\text{B}}$ task: The amount of activity on a given input unit is represented by the size of the square in that unit. The three sets of units, from left to right, represent location (A, B, and C), cover type (C1 and C2), and toy type (T1 and T2).*
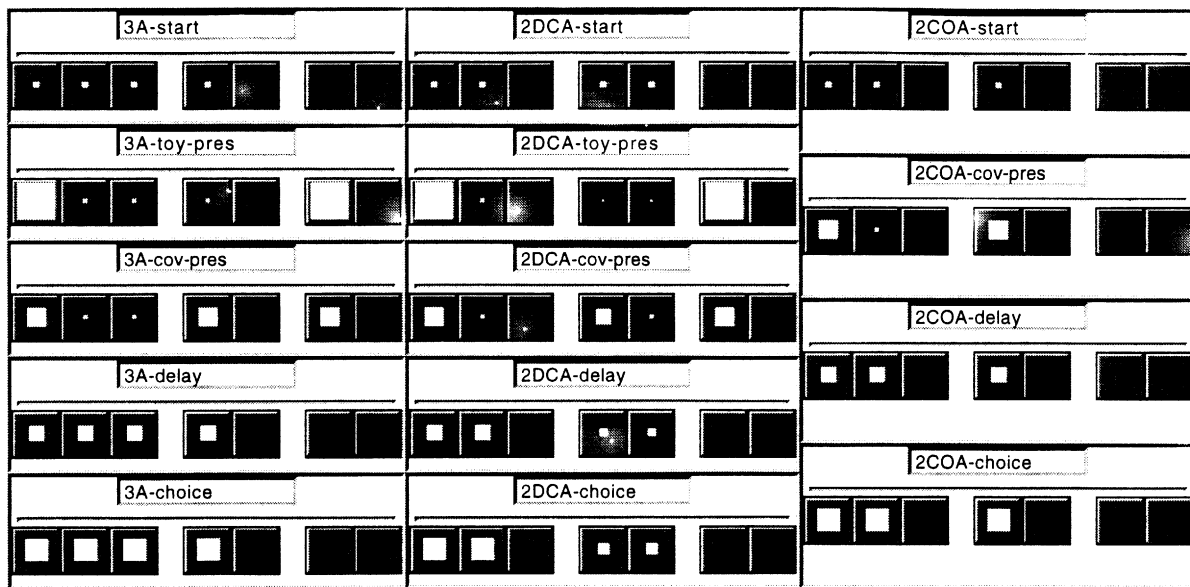
Figure 3   *Input patterns for segments of an A trial in the three variants of the AB̄ task: 3-location (left column), distinctive covers (middle column), and covers-only (right column). As in the 2-location task, the start, delay, and choice segments within each of these versions consist of similar patterns of activity, because the same aspects of the stimulus are visible. With distinctive covers, individual unit activity in the cover-type units is reduced (though the sum over the units in each segment is equal to that in the standard 2 location task), to reflect the greater difficulty of taking in featural information given two types of covers rather than one. During the B trial in the distinctive covers version, C2 is presented rather than C1. The covers-only version has only four segments because no toy is presented.*

activity of the output units in the reaching layer was clamped to 0 during all segments other than *choice*, to reflect the fact that the system was only permitted to reach during *choice*. For all hidden and output unit analyses, activations were converted into percentages of responses made to a given location out of all possible locations. For example, in the 3-location task, the network's percent response to *A* was computed as the activation of unit *A* divided by the sum of activation over units *A*, *B*, and *C* in that layer. In the 2-location task, the network's percent response to *A* was computed as the activation of unit *A* divided by the sum of activation over units *A* and *B*.[5] Unit activations were reset to 0 between trials.

---

[5] The network models the average probabilities of responses through the use of graded continuous activation values. This prevents this network from modelling trial-by-trial responses of infants. However, a network using stochastic binary activation values would allow the modelling of discrete trial-by-trial responses. There is reason to believe that the asymptotic behavior of these discrete networks will closely approximate the behavior of the current network (Hopfield, 1982, 1984). Thus, the current model can be viewed as an abstraction of the more detailed stochastic network, sufficient for capturing the essential characteristics of AB̄ phenomena.

Learning Rule

Feedforward weights were adjusted after every time step in each segment according to a variant of the 'zero-sum' Hebbian learning algorithm (O'Reilly & McClelland, 1992):

$$\Delta w_{rs} = \varepsilon a_r (a_s - \mu(a_s)) \qquad (5)$$

where a_r and a_s are the activities of the receiving and sending units, respectively, ε is the learning rate (set to 0.025) and $\mu(a_s)$ is an approximation of the average activity of the sending units in the layer:

$$\mu(a_s) = \frac{\Sigma a_s}{N + \eta} \qquad (6)$$

η is a novel term that was introduced for reasons described below.

The tenet underlying the Hebbian algorithm states that the connection strength between two units increases when the units are synchronously active. The zero-sum variant of the Hebbian algorithm introduces the average sending unit activity term, so that a sending unit's activity is considered relative to the activity of other sending units in the layer. As a result, for a given active receiving unit, weight increases occur when a sending

unit stands out relative to other units in its layer, that is, when its activity is higher than the average activity of the layer. On the flip side, for a given active receiving unit, weight decreases occur when a sending unit's activity is lower than the average activity of the layer. The zero-sum Hebbian algorithm thus captures the assumption that more learning will occur when the input contains discriminable aspects (resulting in patterns of high and low activity) than when the stimulus is uniform. Thus, more learning will occur when an event occurs at one location but not another (e.g., a toy is hidden at A, and nothing happens at B) than when the two locations are similarly salient (e.g., during the *delay*, when the A and B location input units are identically active).

η was introduced to offset the small number of input units in the network, to allow some learning to occur when all of the input units in a layer were equally active. For example, during the *choice* in a three-location version of the A$\overline{\text{B}}$ task, all three location units are equally active, so that the activity of any given sending unit is equal to the average sending unit activity for the layer. Thus, with the original zero-sum Hebbian algorithm, no learning would occur in this segment. However, it is likely that infants do learn during *choice* periods (Smith *et al.*, 1995), presumably because the A$\overline{\text{B}}$ apparatus stands out from the environment. That is, even if all three locations are equally active, the activity of a single location would be greater than the average activity computed over the surrounding environment. Thus, with a larger number of input units to represent the relatively inactive surrounding environment, the network would show some learning when all inputs corresponding to the A$\overline{\text{B}}$ apparatus were equally active. However, in the current simulations, the number of hidden units was kept small in order to simplify the analysis of network behavior; the η term was thus introduced as a proxy for a larger number of input units. η was set to $\frac{1}{9}$N.

Feedforward weights could take any values between 0 and 1. 'Soft' weight bounding was implemented by multiplying the Δwt term by $(1 - wt)$ if the weight change was greater than 0, and by $(wt - 0)$ if the weight change was less than 0. Bounding of the weights prevents extreme weight values; soft weight bounding allows this to occur as part of the natural dynamics of the network's weight updating. Weights were initialized at the start of each variant of the A$\overline{\text{B}}$ task.

Performance on Standard Two-location A$\overline{\text{B}}$ Task

The results reported here are based on a network tested

on the standard A$\overline{\text{B}}$ task, with A and B trials with delays of three time steps, and with recurrent weights of 0.35. With these parameters, the network responds accurately on pre-trials and A trials in gaze/expectation and reach, but responds perseveratively on B trials in both output systems. Results from networks tested at different ages (with a range of recurrent weight values) are reported in the next section.
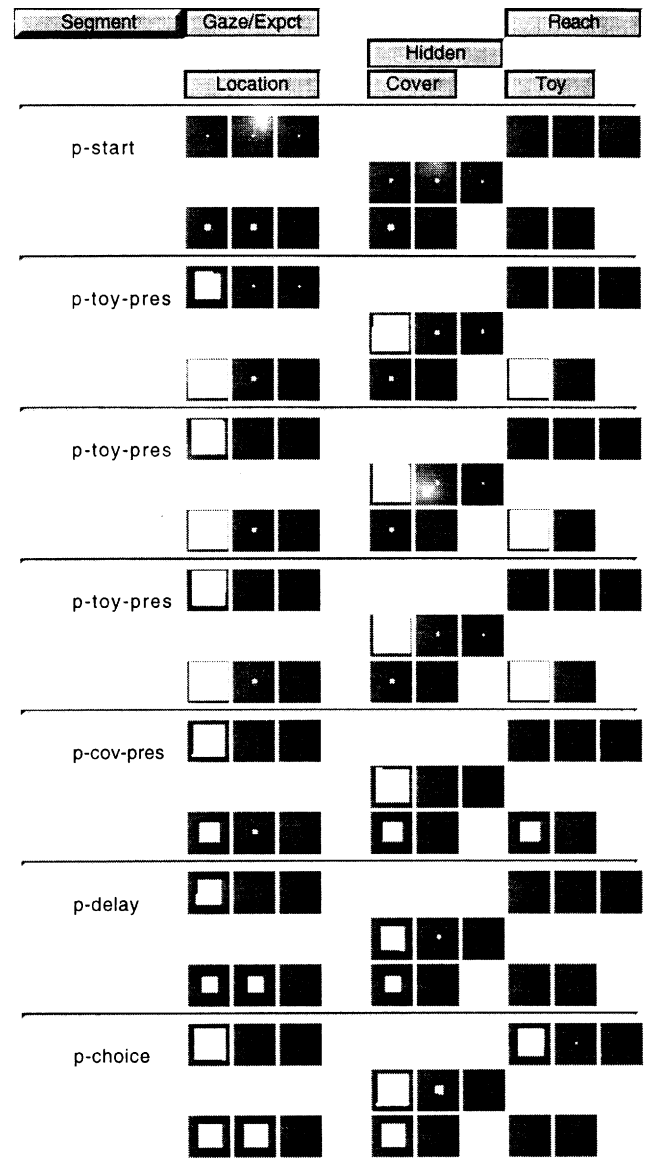

Figure 4 *Network activity during pre-trial: The activity of each unit in each layer (as indicated by the header) is represented by the size of the inner white square, and is shown for each segment of the first pre-trial. The input activity is clamped throughout the trial, and the Reach layer activity is clamped to 0 during all segments except* choice. *The network successfully represents and responds to the* A *location.*

Pre-trials

Figure 4 shows the activity of each unit in the network during each time step of the first pre-trial (subsequent pre-trials yield similar performance). The network represents and responds appropriately to *A*. During the *start* segment, the input units are only weakly activated, and in turn, the hidden and gaze/expectation units become only weakly active. During the presentation of the toy, the *A* location and the T1 units are strongly activated. Because the network enters the task with the *A* input unit strongly linked to the *A* hidden unit and the *A* hidden unit strongly linked to the *A* response units, the *A* hidden and gaze/expectation units also become strongly active. During the brief delay following *presentation*, the hidden and output activity at *A* decreases slightly, as the *A* and *B* location input units receive equal activation. However, the activity at *A* maintains enough of its strength to dominate over *B* at *choice*. At that segment of the trial, the gaze/expectation and reach systems respond to *A* 96% and 92% of the time, respectively. The slightly better performance of the gaze/expectation system is due to its ability to update throughout the trial. As a result, the *A* unit in the gaze/expectation system becomes active before *choice* is presented, giving it an advantage over the reach system which can become active only at *choice*.

A-trials

Network activity during *A* trials is similar to that during pre-trials, with slight differences due to the changes in connection weights accumulated over the pre-trials. In particular, the activity of the *A* hidden and response units is typically higher in the *A* trials than in the pre-trials. For example, at the first time step of *delay*, the hidden units respond to *A* 91% and 87% of the time in the second *A*-trial and first pre-trial, respectively. Such differences between pre-trials and *A*-trials are small; changes in connection weights have relatively little effect when there is cooperation between the active memory traces in the hidden units and the latent traces in connection weights. However, the influence of latent traces can become much more evident when latent traces compete with active traces.

B-trials

Figure 5 shows the activity of each unit in the network during each segment of the *B*-trial. The network responds perseveratively to *A* at *choice*, in both gaze/expectation and in reach.

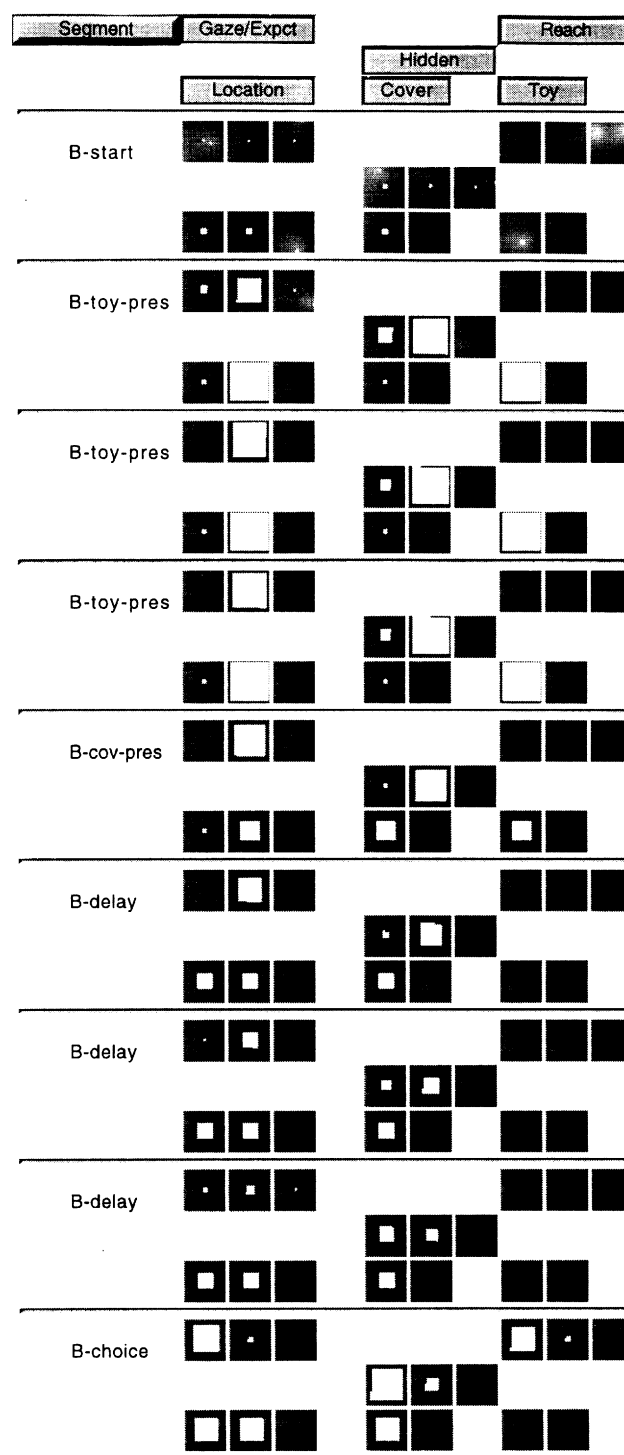Why does the network make the $A\overline{B}$ error? The

Figure 5   *Network activity during* B *trial: The network initially represents the* B *location, but this representation decays and the representation of the* A *location increases over the course of the delay. Just prior to choice, the hidden unit activity conveys little distinction between* A *and* B. *However, at choice, the network perseverates to* A *in both gaze/expectation and reaching.*

network represents and responds to *B* in gaze/expectation during *presentation* (Figure 5). However, during *delay*, this representation fades more rapidly than in prior trials (due to the influence of latent traces described below), so that the hidden unit activity shows little evidence of which location was recently-attended just prior to *choice*. If judged on the basis of active traces alone at that time, the network shows little memory of prior events. However, the network shows strong evidence of memory for the previous trials by making the A$\overline{\text{B}}$ error at *choice*, indicating the influence of latent traces. In particular, the network's connection weights favor activity at *A* over *B*.

As specified by the Hebbian learning rule, the network learns which input features are associated with its hidden unit representation of *A* and its output responses to *A*, while learning little about the features associated with *B*. Input units *A*, C1, and T1 become more strongly connected to hidden unit *A* as those units are synchronously active during pre-trials and *A*-trials. In contrast, because hidden unit *B* shows little activity during those trials (Figure 4), its incoming weights undergo little change. Similarly, the *A* output unit receives stronger weights than the *B* output unit. The pattern is weaker for the weights to the Reach layer than for the weights to the gaze/expectation layer, because the Reach layer is active only during *choice*.

The influence of these weight changes on network activity is evident both when the task-relevant inputs are differentially active (during *presentation*) and when the task-relevant inputs are identically active (during *start*, *delay*, and *choice*). For instance, at the start of *presentation*, the hidden units respond to *A* in the *A*-trial 90% of the time and to *B* in the *B*-trial only 70% of the time. At the start of *delay*, the hidden units respond to *A* in the *A*-trial 91% of the time and to *B* in the *B*-trial only 78% of the time. And, although the hidden units favor *B* at the start of the *delay* in the *B* trial, and the input activity is identical for *A* and *B*, the hidden units switch to favor *A* at choice. All of these differences in unit activities result from the changes made to connection weights over the course of the A$\overline{\text{B}}$ task.

These changes might correspond to the following processes in infants: As infants repeatedly attend and respond to a location, they link their responses with features associated with that location (what it is, where it is, etc.). For example, they might link a toy, a cover, and their lefthand side with attending and reaching to the *A* location in the A$\overline{\text{B}}$ task. Such links would make infants increasingly likely to attend and respond to the *A* location, particularly with the toy and the cover there. Infants can still attend and respond elsewhere as they learn links such as these, but the links diminish infants'

attention and responsiveness to other locations. As a result, with repeated exposure to events at a given location, infants become less responsive to events at other locations.[6]

Effects of age on gaze/expectation and reaching performance

This section reports the responses of networks at different ages, that is, with different values of recurrent weights. The model simulates infants' performance in showing non-perseverative responding earlier in gaze/expectation than in reach, and leads to the novel prediction that under certain conditions infants may perseverate more in gaze/expectation than in reach.

Figure 6 shows the performance of the network on second *A* trials and *B* trials as recurrence varies from 0.3 to 0.5. The network performs well on *A* trials across this range of recurrence because latent weight changes favor *A* over *B*. The non-recurrent weights thus support enough activity at *A* that the network's ability to maintain activity at *A* has little effect on performance. In contrast, on *B* trials, the network progresses from responding perseveratively to *A* to responding correctly to *B* as the recurrent weights increase, as do infants with increasing age (Wellman *et al.*, 1986). The network's ability to maintain activity is critical to performance on *B* trials because the network must maintain a representation of *B* in the face of the latent bias to respond to *A*.

Figure 6 also shows that the network produces non-perseverative responding earlier in development in gaze/expectation than in reach, as do infants in the A$\overline{\text{B}}$ task (Hofstadter & Reznick, 1996; Ahmed & Ruffman, 1997). The dissociation in the network is based entirely on the lower frequency of reaching, which results in fewer opportunities for the reaching system to update based on a recently-presented location. Consider the network with recurrence of 0.38. Figure 7 shows the activity of each unit in the network during each segment of the *B*-trial. The gaze/expectation system responds to *B* 55% of the time, while the reach system responds to *B* only 37% of the time. With this level of recurrence, the network can sustain enough of a preference for the recently-presented *B* trial to benefit from the continual updating of the gaze/expectation system.

---

[6] An exception occurs in the inhibition-of-return (IOR) and spontaneous alternation (SA) paradigms, in which infants respond more to locations that they have not recently attended. However, the IOR and SA paradigms differ in an important way from the A$\overline{\text{B}}$ paradigm, in that their first '*A*' events are presented in isolation (without *B* as a choice), such that the '*B*' events to which infants subsequently attend/respond are more novel than those in the standard A$\overline{\text{B}}$ task.
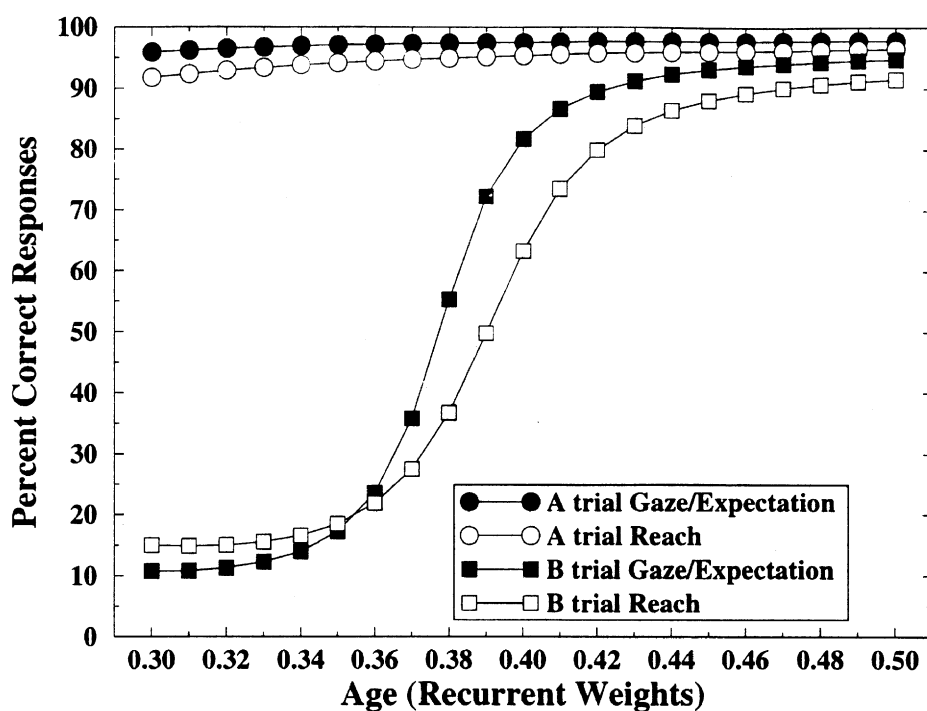
Figure 6 *Percent correct responses as a function of age: On A trials, the network is accurate across all levels of recurrence shown because any ability to maintain activity for A is supported by latent weight changes (built up across previous trials) that favor A over B. On B trials, the network responds non-perseveratively only as the recurrent weights get stronger. The gaze/expectation system shows non-perseverative responding earlier than the reach system.*

Similarly, infants may show earlier success in gaze/expectation variants of the A$\overline{B}$ task because they can constantly update their gazing and their expectations. As a result, they can counter perseverative tendencies on *B* trials by gazing at *B* and forming expectations about *B* during the *presentation*, *delay*, and *choice* trial periods. In contrast, infants can only reach at the *choice* point, by which time their representations have become more susceptible to perseverative biases.

Interestingly, Figure 6 shows that very 'young' networks show slightly more perseveration in gaze/expectation than in reach. Networks with relatively weak recurrent weights tend to default to the prepotent response, in which case the continual updating of the gaze/expectation system leads it to show more of this prepotent response than the reach system. In effect, when recurrence is high enough to keep the right location (*B*) in mind, then 'repeating' it, as the gaze/expectation system does, helps performance; in contrast, when recurrence is low so that the wrong, prepotent location (*A*) comes to mind, then repeating hurts performance. The difference between the gaze/expectation and reach systems decreases at the ends of the curves in Figure 6 because additional updating contributes less to performance as the system falls back to its default

activity (low recurrence) and as the system comes to favor *B* over *A* completely (high recurrence). This analysis of the difference between gaze/expectation and reach systems leads to the following prediction:

- There may be a period early in development during which infants perseverate more in gaze/expectation than in reaching. In contrast with the already-observed difference between gaze/expectation and reach measures in older infants, the simulations suggest that the predicted difference is likely to be of a smaller magnitude, and between two responses which are both perseverative overall (rather than dissociated).

It is important to stress that two aspects of the network's development challenge the tendency to explain infant behavior in terms of disembodied knowledge and deficits in ancillary systems. First, the network improved its performance as its ability to maintain active representations increased. Thus, the network progressed from responding perseveratively to responding correctly in the A$\overline{B}$ task based on the same process that allowed Munakata *et al.*'s (1997) network to progress from recognizing impossible events to reaching correctly in a single hiding location task. Second, the network showed
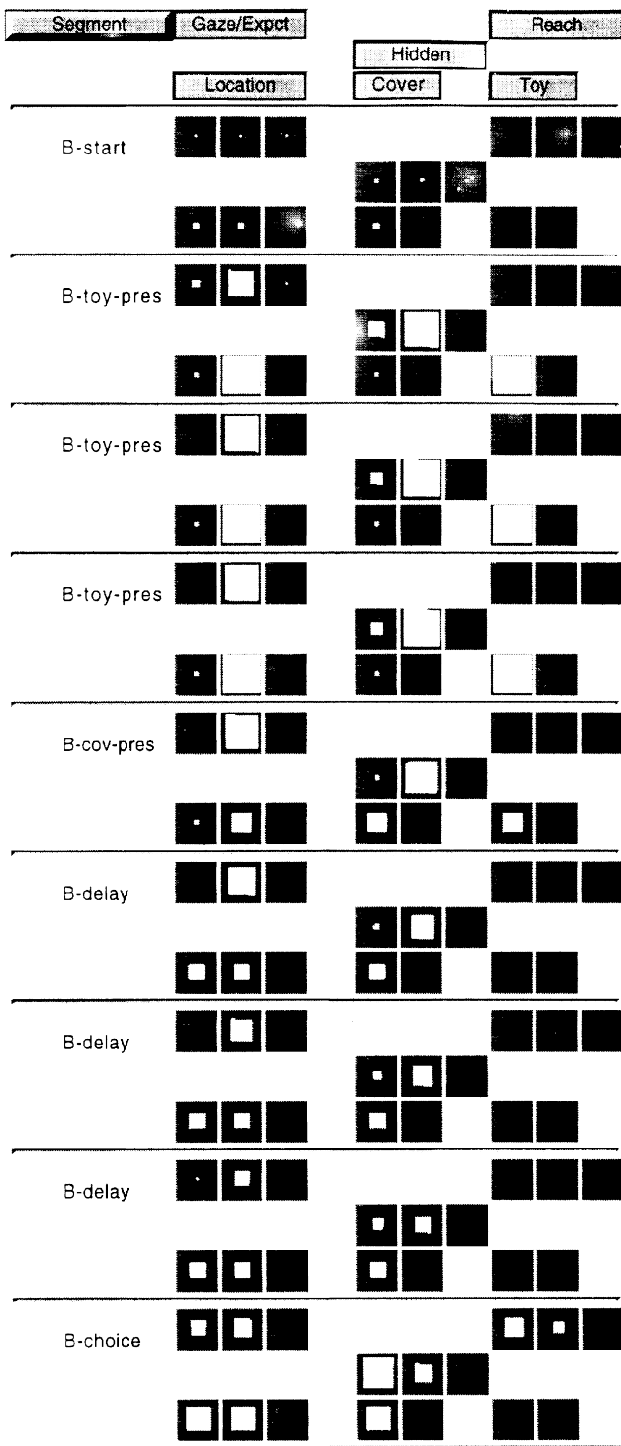
Figure 7 *Network activity during B trial with recurrence = 0.38: The gaze/expectation system shows better performance (responding to B 55% of the time) than the reach system (responding to B 37% of the time), due to the former's ability to update on each time step.*

a dissociation between gaze/expectation and reaching in the A$\overline{B}$ task based solely on the different frequencies of these behaviors. By updating more frequently, the gaze/expectation response was able to make different use of weak 'knowledge' representations on the network's hidden units. These two findings from the A$\overline{B}$ model run counter to standard accounts of infant behavior in this task, according to which gaze/expectation measures reveal that infants 'have the knowledge' that a toy was recently hidden at B, but they fail to reach accordingly due to reaching-specific deficits in inhibitory systems (Diamond, 1991; Hofstadter & Reznick, 1996; Ahmed & Ruffman, 1997) or problem-solving abilities (Baillargeon & Graber, 1988; Baillargeon *et al.*, 1989). Instead, the simulations suggest that a common process – the development of the ability to maintain active representations – may underlie infants' task-dependent progression for demonstrating sensitivity to hidden objects. Apparent improvements in infants' abilities to solve problems or to inhibit inappropriate responses may result from their increasing maintenance abilities. Thelen and Smith (1994) argued a similar point in their dynamic systems account of improvements in the A$\overline{B}$ task, in which 'there is not some internal ability called "inhibition" that causes performance; rather inhibition *is the product* of a system in a particular context with particular dynamic properties' (p. 308). Similar arguments have been made for a unitary mechanism underlying working memory and inhibition in adult normals (Roberts, Hager, & Heron, 1994; Cohen & O'Reilly, 1996; Cohen, Braver, O'Reilly, 1996), parietal-damaged patients (Cohen, Romero, Farah, & Servan-Schreiber, 1994), and schizophrenic patients (Cohen & Servan-Schreiber, 1992).

Effects of Observations at *A*

Because factors that influence infants' reaching in the A$\overline{B}$ task have been more extensively studied than factors influencing gaze/expectation, this and all subsequent simulation sections focus on reaching measures. This section shows that the model simulates infants' performance in the *A*-observation version of the task, reaching perseveratively after only observing events at *A*.

The input patterns for the *A*-observation version of the A$\overline{B}$ task are identical to those in the standard 2-location task (Figure 2). The network's lack of reaching to *A* is implemented by setting the learning rate to 0 in the weights to the reaching layer during the pre-trials' and *A* trials' *choice* segment (the only segment in which the network is normally permitted to reach). Two measures can then be recorded: 1) the network's

reaching on *B* trials after it has effectively had no experience reaching at *A*, because the learning rate of 0 during *choice* prevents any weight changes on the basis of reaching, and 2) the network's reaching on *A* trials, effectively where the network would have reached had it not been in the observation version of the task.

Figure 8 shows the percent correct reaching for the network on the second *A* trial (had it in fact been allowed to reach) and on the *B* trial (after it had effectively had no reaching experience to *A*). The network would have reached to *A* with high accuracy on the *A* trials, yet despite not actually doing so, reaches perseveratively to *A* on *B* trials. The network perseverates without actually reaching to *A* because nothing prevents it (or infants) from *attending* to *A*, that is, representing *A* on its hidden units. In this process, the weights from the input to hidden units change as they did when the network was allowed to reach, biasing the network to attend to *A*, and in turn, to respond to it. The *A*-observation network does differ from the reaching network in its reaching weights, which have not changed during pretrials and *A* trials to support perseverative reaching; however, the changes to the other weights in the network are sufficient to produce the A$\overline{\text{B}}$ error.

Effects of age and delay

This section reports simulation results across different ages and different delays. The model simulates infants' performance, improving as age increases and delay decreases, and leads to the novel prediction that under certain conditions infants' A$\overline{\text{B}}$ performance will worsen with age.

Figure 9 shows the performance of the network on *B* trials as recurrence varies from 0.3 to 0.5, and the number of time steps in *delay* varies from 0 to 5. The network's performance improves with age and shorter delays. And, like infants (Wellman *et al.*, 1986), the network progresses from perseverative to chance to correct performance with decreasing delays. As described previously, the network performs better as it ages because higher recurrent weights allow it to maintain a representation of *B* in the face of the latent bias toward *A*. The network performs better with shorter delays, because its active representation of *B* has less time to decay. Recall from Figure 5 that a network with recurrent weights of 0.35 entered the delay with a relatively strong representation of *B* that gradually decayed over subsequent time steps.
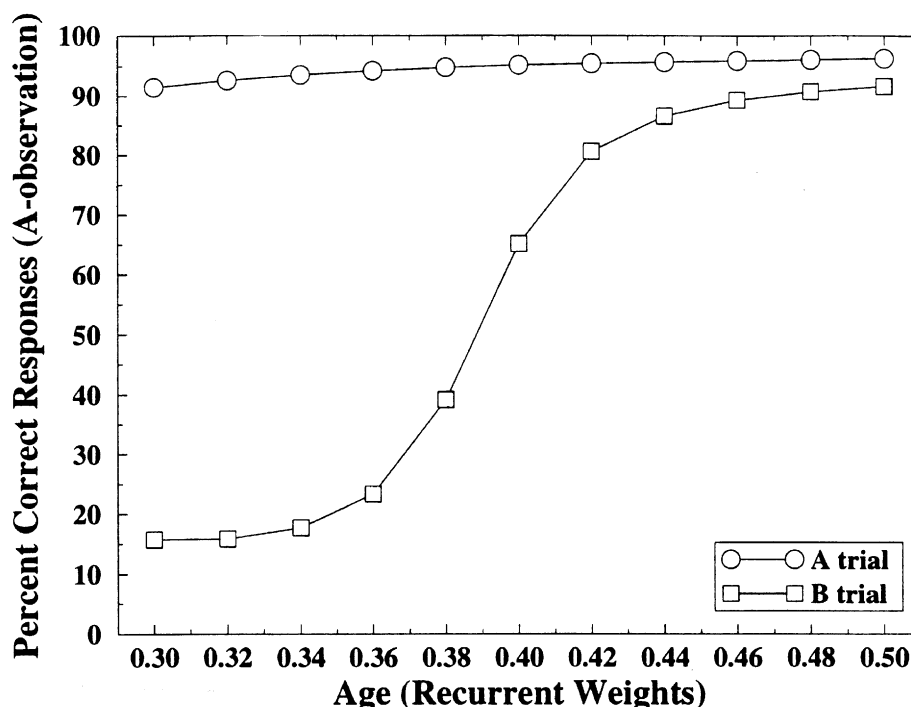

Figure 8   *Percent correct responses on second* A *trials and* B *trials under* A-observation *conditions: The network effectively has had no experience reaching to* A, *but still makes the* A$\overline{B}$ *error. The* A *trial measures reflect percent correct responses had the network been able to reach.*
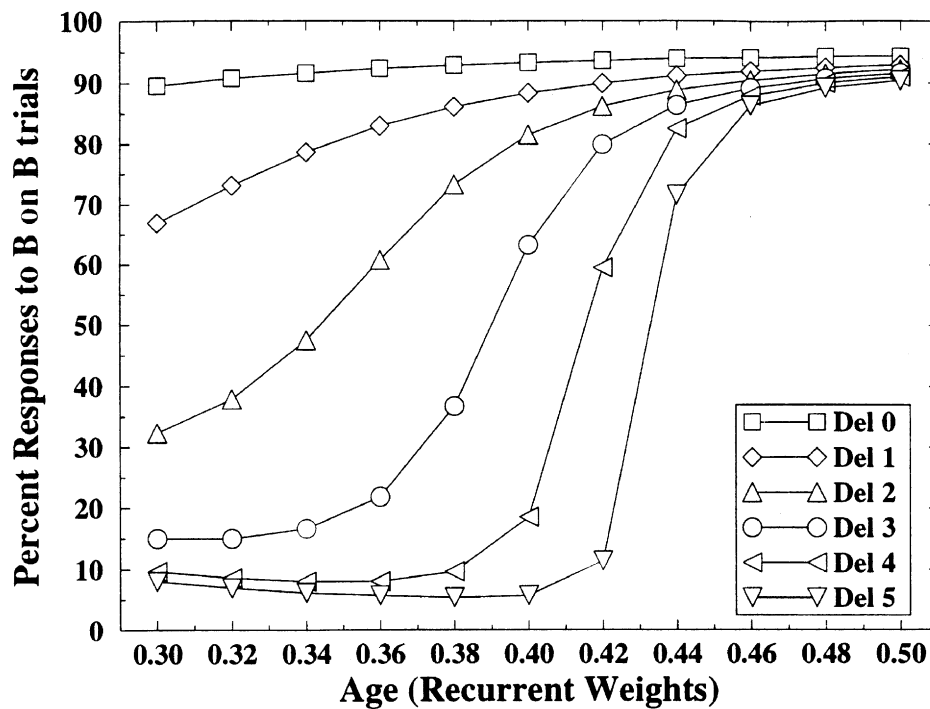
Figure 9   *Percent correct responses to B on B trials as a function of age and delay: The network performs better with higher recurrent weights and shorter delays.*

These effects of age and delay on network performance may seem straightforward; however Figure 9 reveals only a simple part of a more complicated picture of the network's behavior. Figure 10 includes the performance of the network with lower levels of recurrence. The developmental curves are actually U-shaped for longer delay periods. That is, the network shows its most perseverative responding in the middle of the developmental period shown.

The patterns of network activity during *A* and *B* trials reveal how increases in recurrence can hurt network performance. The representations of very young networks are so weak that they fade quickly over even *A* trial delays, leading to weak prepotent responses to *A*. As these representations become stronger, they fade less quickly over *A* trial delays, leading to stronger prepotent responses to *A*. In effect, the more the network keeps *A* in mind (as recurrence increases), the more biased the network becomes to respond to *A*. Becoming increasingly able to keep something in mind helps only if *B* can be 'kept in mind' long enough to sustain the delay; otherwise – if the system must perseverate – it is better off the less it keeps things in mind. Longer delay periods make the U-shape more prominent, because the recurrent weights influence the network's activity most during *delay*; the longer their period of influence, the more

evident their contribution. This analysis of the effects of recurrence leads to the following prediction:

● There may be a period quite early in development during which perseveration increases with age. In contrast with the already-observed decrease in perseveration with age, the predicted increase may occur before infants are able to reach; however, other measures (e.g., gaze) might follow the predicted pattern (as gaze/expectation in fact does in the simulation). The simulations suggest that $A\overline{B}$ tasks with longer delays are more likely to reveal the predicted increase in perseveration.

Effects of multiple locations and distinctive covers

The previous section showed that the model simulated the effects of two factors – age and delay – that emerged from Wellman *et al.*'s (1986) meta-analysis of the $A\overline{B}$ error. This section shows that the model simulates $A\overline{B}$ performance as a function of the two other factors – number and distinctiveness of hiding locations – that emerged from this meta-analysis, with performance improving with additional and distinctive
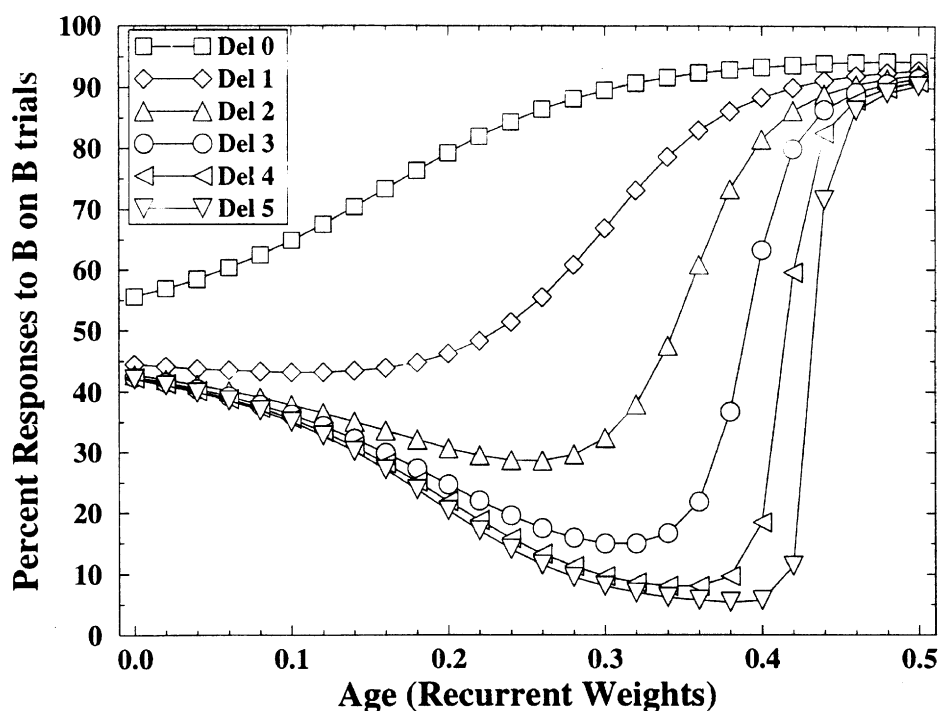
Figure 10 *Percent correct responses to B on B trials as a function of age and delay: Testing the network with lower levels of recurrence reveals that the curves for longer delay periods show a U-shape, with networks with middle levels of recurrence responding most perseveratively.*

hiding locations.[7] The model also leads to the novel predictions that infants may perform better with additional hiding locations or distinctive covers even if they are present only on *A* trials, and that the benefits of additional hiding locations will decrease across additional *B* trials while those of distinctive covers will increase.

Figure 11 shows the performance of the network as recurrence varies from 0.2 to 0.5, on *B* trials across three versions of the $A\overline{B}$ task: standard 2-location, 3-location, and distinctive covers (each with delays of three time steps). The network's performance improves with more locations and distinctive covers. In the network, both improvements rely on two aspects of

changes to the connection weights based on patterns of activity.

First, the more alternatives the network has on *A* trials, the less *A* stands out. Additional locations or distinctive covers increase the average activity of the input layer, reducing the impact of input activity associated with *A*. As a result, the connections to the *A* hidden unit from its usual inputs do not strengthen as much. The same pattern holds for changes to the connections to the *B* hidden unit on *B* trials. However, alternatives have greater influence on changes to the *A* hidden unit connections because these undergo greater change during pre-trials and *A*-trials, leading to the net effect of a weaker prepotent response to *A*.

Second, the presence of alternative locations or cover types favors non-prepotent responses. Across pre-trials and *A*-trials, the network learns that alternatives are not associated with *A* while learning little about *B*, so the presence of alternatives detracts more from representing and responding to *A* than to *B*. In effect, 'an enemy's enemy is a friend'; alternative locations or cover types serve as B's friends because they are enemies of *B*'s enemy, *A*. For example, *C* input favors *B* over *A* because the network learns that *C* input is not associated with hidden unit *A* activity over the pre-trials and

[7] It should be noted that Diamond, Cruttenden, and Neiderman (1994) argued that improvements in performance with additional hiding locations are an artifact of procedural differences between two-well and multiple-well variants of the $A\overline{B}$ task. In particular, they argued that infants can maintain attention on the correct well more easily in multiple-well tasks because only the correct well is uncovered and re-covered, whereas in two-well studies, both wells are covered simultaneously. This procedural difference may contribute to infants' improvements with multiple hiding locations; however, the multiple-well advantage remains after the procedural difference is controlled for (Bjork & Cummings, 1984), so other factors must also contribute.
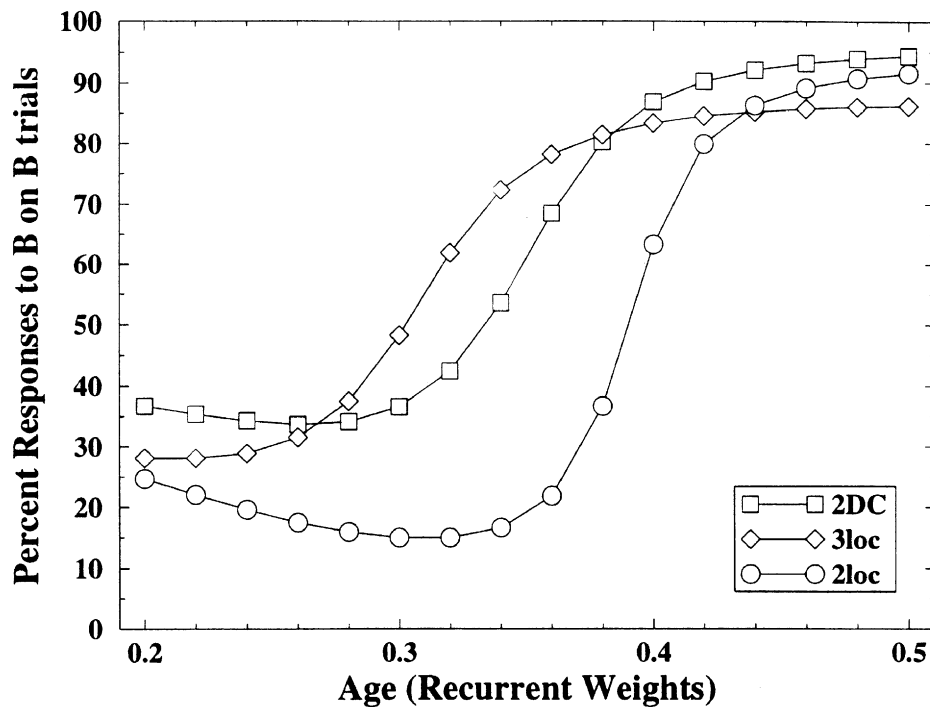
Figure 11 *Percent correct responses to B on B trials as a function of age and task version: The network shows non-perseverative responding earlier with distinctive covers ('2dc') and multiple locations ('3loc') than in the standard 2-location ('2loc') task.*

A-trials, so the connection from input unit *C* to hidden unit *A* weakens. Although *C* input is similarly unassociated with hidden unit *B* activity, the connection between them does not weaken as much because hidden unit *B* shows little activity across pre-trials and *A* trials. In effect, the network does not attend to *B* and so learns little about what is and is not associated with it. The net effect is that *C* favors activity at *B* over *A*. Similarly, C2 input favors *B* over *A* because C2 is less active than C1 when *A* is active during the cover-presentation in *A* trials, so the connection from C2 to hidden unit *A* weakens. In addition, C2 input favors *B* over *A* because C2 is synchronously active with *B* during presentation of the *B* trial. However, this effect is small because C2 and C1 are equally active during all other segments of *A* and *B* trials, so the network has difficulty binding each to a unique location.[8] It is important to note that additional *B* trials have different effects on hidden unit *B*'s incoming weights from *C* and C2; the *C* weight weakens as the network learns that *C* is not associated with *B*, whereas the C2 weight strengthens as the network learns that C2 is associated with *B*.

[8] A network – or infant – with more capability to solve the binding problem would presumably benefit more from the presence of a distinctive cover at *B*.

These factors – weakening of the prepotent response to *A*, presence of alternatives that favor *B* over *A* – can be mapped onto infants' processing in the A$\overline{\text{B}}$ task in a relatively straightforward way. Infants may respond to *A* on *A* trials in all three versions of the task; however, the presence of an additional location or a distinctive cover reduces infants' attention to aspects of the *A* trial, leading to weaker associations between these aspects and the response to *A*. On *B* trials, the presence of an additional hiding location or distinctive cover reduces A$\overline{\text{B}}$ errors by providing a stimulus that is not associated with infants' responses to *A* (and neutral with respect to *B*, because infants have yet to learn much about *B*). The model thus leads to the following novel predictions:

- Infants may show non-perseverative responding earlier with distinctive covers and multiple locations than in the standard 2-location task, even if the distinctive cover or additional location is present only on *A* trials.
- With additional *B* trials, infants may show less benefit from the presence of additional hiding locations, and more benefit from the presence of distinctive covers.

The first prediction is based on the presence of alternatives weakening the prepotent response to *A*; the second

is based on the time course of alternatives favoring *B* over *A*.

Effects of hidden toys vs. covers alone

This section reports the network's responses on four $A\overline{B}$ variants resulting from crossing *A* and *B* trials with hidden toy and cover alone conditions. The model simulates infants' performance, showing similar performance on *A* trials regardless of toy presence, and similar $A\overline{B}$ errors across all variants except the *A*-cover/*B*-toy (cover-toy) variant. The model also shows how a hidden toy can matter and yet sometimes fail to show its influence in behavior.

Figure 12 shows the network's performance on the two types of *A* trials (*A*-cover and *A*-toy) and on the *B* trials of the four variants of the $A\overline{B}$ task. The network is accurate on both *A* trial types. The network is biased to respond to *A* based on the latent weight changes built up across pre-trials and *A* trials, and so level of input activity (higher for Toy than for Cover) and ability to maintain activity, within the range shown, have little effect on performance. On *B* trials, the network per-

forms similarly on toy-toy and cover-cover variants because strong activations at *B* compete against strong weight changes toward *A* in the former, and weak activations at *B* compete against weak weight changes toward *A* in the latter. The network shows its best performance in the cover-toy version because strong activations at *B* compete against weak weight changes toward *A*. The competition between weights and activations can thus account for the network's performance on three of these four versions of the $A\overline{B}$ task.

However, one might expect the network to perform its worst in the toy-cover variant, in which weak activation at *B* competes against strong weight changes toward *A*. Instead, as with infants, performance on the toy-cover variant is comparable to that on the toy-toy and cover-cover versions. This is due to the similarity between *A* and *B* trials in the toy-toy and cover-cover versions. In effect, the more *B* trials resemble *A* trials, the more they 'remind' the network of *A* trials and so bias the network to respond prepotently. This perseverative 'reminding' effect is strong enough to balance the greater attention to *B* in the toy-toy version relative to the toy-cover version, and strong enough to balance the weaker prepotent response to *A* in the cover-cover version relative to the
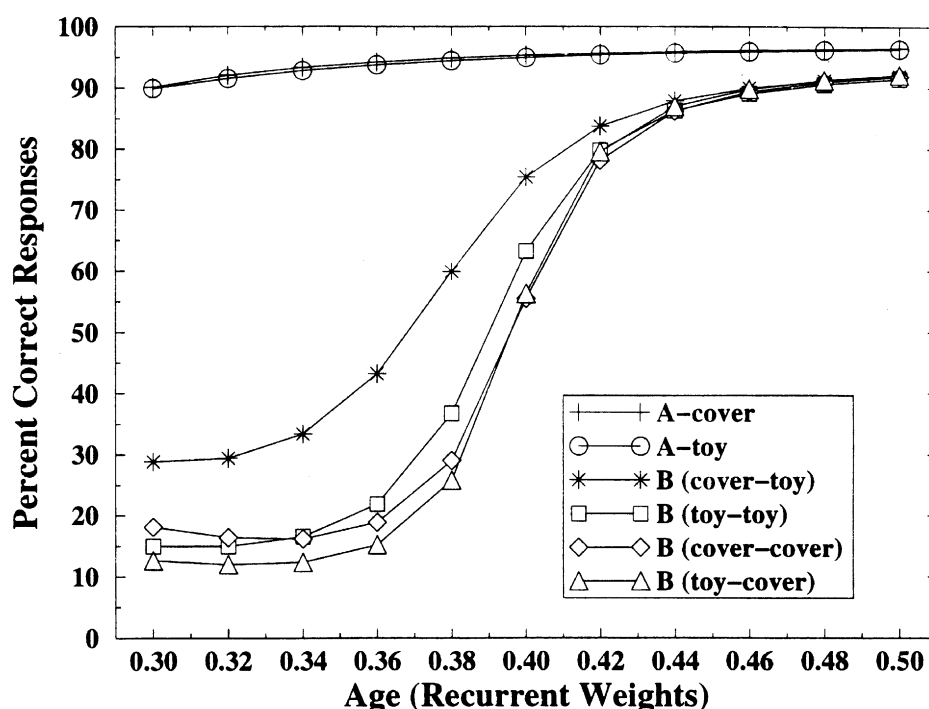

Figure 12 *Percent correct responses as a function of age and toy presence: On* A *trials, the network is accurate with both cover and toy trials across all levels of recurrence shown, because any activity for* A *is supported by latent weight changes that favor* A *over* B. *On* B *trials, the network perseverates on all task variants at some point in development, but shows non-perseverative responding earliest on the* A-cover/B-toy *(cover-toy) variant.*

toy-cover version. That is, on *B* trials, the toy in the toy-toy version does provide stronger activation at *B* than the cover in the toy-cover version. However, because the toy has been strongly associated with *A* over pre-trials and *A* trials, it also provides more activity to *A* than does the cover in the toy-cover version. In effect, the toy directs the network's attention to *B* while also reminding it of *A*. The net effect is an A$\overline{\text{B}}$ error similar to that in the toy-cover version. Similarly, on *A* trials, the toy in the toy-cover version does result in a stronger weight advantage for hidden unit *A* from the Location units relative to the cover-cover version. However, because the cover is presented only briefly during toy trials whereas it is the focus during cover trials, the cover-cover version results in a stronger weight advantage for hidden unit *A* from the C1 unit relative to the toy-cover version. As a result, the reminding role of the cover in the cover-cover version balances the stronger weight changes from the Location units toward *A* in the toy-cover version, so that the network makes comparable A$\overline{\text{B}}$ errors in the two versions.

In contrast with Smith *et al.*'s (1995) account in which the toy does not matter, the model thus captures Munakata's (in press) empirical findings showing that the toy does matter both in *recency* (in *B* trials, reducing the A$\overline{\text{B}}$ error following *A* cover trials) and in *history* (in *A* trials, reducing the influence of a toy on *B* trials). It is clear that the toy matters to the network, that is, it is treated differently from a cover alone – resulting in different activation levels and patterns of weight changes; however, these differences are not always evident in the network's behavior. The simulations thus demonstrate how a system that distinguishes toys and covers could nonetheless behave at times as if it did not, based on the underlying dynamic between active and latent traces.

The corresponding processes in infants might be summarized as follows: Infants distinguish covers alone from the more interesting case of toys underneath covers, but their behavior does not always reveal this distinction. First, infants may not show their ability to distinguish toys and covers until faced with a competing prepotent response. Infants tend to reach correctly to *A* on both types of pre-trials because they are sufficiently interested in each to span the short pre-trial delay. And, the experience of representing and responding to *A* sets up a prepotent response, so that infants subsequently reach correctly to *A* on both types of *A* trials. Second, infants may distinguish toys and covers on *B* trials in a way that balances the distinction on *A* trials, such that the distinctions are not evident in their behavior. That is, infants perform similarly on the toy-toy and cover-cover variants, not because the toy does not influence their

behavior, but because the toy's influence on *A* trials is balanced by its influence on *B* trials. Third, the less *B* trials resemble *A* trials, the less they remind infants of *A* and the less they bias them to respond prepotently. Thus, infants do not perform substantially worse on the toy-cover variant, despite the toy presented at *A* being more interesting than the cover presented at *B*, because the *B* trial does not strongly remind infants of the *A* trial. Infants perform best in the cover-toy variant because they distinguish the more interesting toy at *B* from the less interesting cover presented at *A*, and the toy at *B* does not remind them of *A*.

General discussion

This section discusses the contributions of the A$\overline{\text{B}}$ model proposed in this paper, compares it to an existing model of A$\overline{\text{B}}$ performance, and raises remaining issues for exploration in future models.

Contributions of the model

The model proposed in this paper provides an explicit set of processing mechanisms to account for a large and diverse set of data on infant performance in the A$\overline{\text{B}}$ task (Table 1). The model demonstrates how basic processes – the formation of latent memory traces and their interaction with developing active memory traces – can provide a unifying framework for understanding why and when infants perseverate, and what factors influence the extent of their perseveration. In addition, this unifying framework may span improvements in object permanence tasks throughout the first year of life, from early success in visual habituation studies, to retrieval of occluded toys, to eventual success in the A$\overline{\text{B}}$ task.

In contrast, often a unique explanatory factor is invoked for infant performance on each object permanence task (and even each variant of each object permanence task). For example, infants 'know' about object permanence from very early in life, but: fail to notice certain violations in visual habituation paradigms due to reasoning deficits (Baillargeon, 1993); fail to reach for hidden objects due to means-ends deficits (Baillargeon, Graber, DeVos, & Black, 1990; Diamond, 1991; Willatts, 1990); perseverate in the A$\overline{\text{B}}$ task due to inhibitory deficits (Diamond, 1991; Hofstadter & Reznick, 1996; Ahmed & Ruffman, 1997); and perseverate without hidden toys because hidden toys don't actually matter (Smith *et al.*, 1995). And so on. At best, these 'one finding, one explanation' accounts provide a list of unrelated factors that influence infant behavior.

At worst, they form a set of ad hoc proposals that prove insufficient in the final analysis.

Comparison to Dehaene and Changeux's (1989) model

As mentioned earlier, Dehaene and Changeux (1989) proposed an A$\overline{\text{B}}$ model that also taps the distinction between active and latent traces. Their model is comprised of two levels, the first with direct connections from input to output, the second passing through a layer of self-excitatory mutually inhibitory memory units. The first level can retain only latent traces in the form of changes to the connection weights, whereas the second can retain active traces in the form of maintained activations. The model makes a strong perseverative response when only the first level is functional, and responds correctly when the second level is added.

Dehaene and Changeux's (1989) model and the model proposed in this paper both tap the active-latent distinction (the 'memory'-'synaptic efficacy' distinction for Dehaene and Changeux) that may be critical to understanding infant performance in the A$\overline{\text{B}}$ task. However, the current model includes several important advances over Dehaene and Changeux's model.

First, the current model's A$\overline{\text{B}}$ performance improves through gradual and continuous increases in the network's ability to maintain active representations. In contrast, Dehaene and Changeux's (1989) model improves primarily from a less plausible, discontinuous 'switching on' of its memory units, thought to reside in prefrontal cortex.

Second, the current model provides a closer match to infants' perseveration. The current model, like infants, shows a natural progression in its performance with increasing age and decreasing delay (Figure 9). In contrast, it appears that Dehaene and Changeux's (1989) 'young' model (with only level 1) would make the A$\overline{\text{B}}$ error regardless of delay, because it is incapable of maintaining an active representation and so must rely on its latent representations. When the second level of memory units is added, the network then shows sensitivity to delay but fails to show true perseveration. With a short delay, the network responds correctly on *A* and *B* trials around 90% of the time; with a longer delay, the network performs similarly on *A* trials and responds correctly on *B* trials around 80% of the time. Although the network's performance worsens from *A* trials to *B* trials with the longer delay, 80% correct performance on *B* trials is far from perseveration.[9] Infants in Diamond's (1985) data set (which Dehaene and Changeux claim to model) show approximately 35% correct performance on *B* trials.

Third, the current model provides a closer match to

the learning process in the infant. Dehaene and Changeux's (1989) network learns on the basis of reinforcement – small positive reinforcement for correct orienting during the presentation of events, medium positive reinforcement for correct responses at test, and strong punishment for incorrect responses at test. In contrast, the current model learns without resorting to reinforcement, on the basis of associations between its perceptions, representations, and responses. The lack of reinforcement in the learning algorithm is critical to modelling infant behavior, because infants do not perform in the A$\overline{\text{B}}$ task on the basis of reinforcement. Infants receive no reinforcement for attending to events during their presentation. And, infants perform similarly regardless of whether they are reinforced for making A$\overline{\text{B}}$ errors (Diamond, 1983).

Remaining issues

Although the model simulates a wide range of infant behaviors on the A$\overline{\text{B}}$ task and its variants, certain aspects of cognitive processing are not captured in the model and may be required for the network to simulate additional data.

First, there are no top-down influences in the network. As a result, the responses in the reaching system have no influence on the responses in the gaze/expectation system, and vice versa. The responses of the two systems are of course related because they are based upon the same hidden layer patterns of activity, but each is independent of the other's response to this pattern of activity. The network would thus show similar responses in gaze in a gaze-only variant of the A$\overline{\text{B}}$ task and in the standard task in which reaching accompanies gazing. Although this comparison has not been tested empirically,[10] I would predict that contrary to the network prediction, reaching would influence gazing such that infants would show more perseveration in their gaze responses in the standard A$\overline{\text{B}}$ task than in a gaze-only variant. The network might simulate this pattern of data with the introduction of some type of top-down processing, for example, from the reaching and gaze/expectation layers back to the hidden layer.

Second, the network fails to habituate. Unlike infants,

---

[9] Dehaene and Changeux (1989) argue in a footnote that their model shows a larger perseverative response when they introduce a large noise level and a change to the weight update rule. However, this manipulation and its results are not presented in enough detail to allow sufficient evaluation of the resulting match to the data or the underlying assumptions about development.

[10] Hofstadter and Reznick's (1996) data set contains the relevant data, but this analysis has not been conducted (Reznick, personal communication, June 1996).

the network fails to get bored with the A$\overline{\text{B}}$ task or with the same toy presented repeatedly. As a result, the network is unable to simulate three aspects of infant behavior: 1) deteriorated performance (making errors with equal frequency on *A* and *B* trials) with long delays (Diamond, 1985); the network only becomes increasingly perseverative with longer delays, 2) improved performance when a new toy is hidden on *B* trials (Schuberth, Werner, & Lipsitt, 1978), and 3) occasional errors even when an object is visible at *B* (Butterworth, 1977; Harris, 1974; Bremner & Knowles, 1984). The network shows only slight improvement when a new toy (activating the T2 input unit) is hidden on *B* trials. The network improves for the same reasons it improves when a distinctive cover is used on *B* trials – because novel/distinct input is less associated with *A* than the standard input – but the effect is smaller with the novel toy because the novel toy is more transient than the distinct cover. If the network included some mechanisms of habituation, it might show a more substantial effect of a novel toy on *B* trials (based on its habituation to the familiar toy over pre-trials and *A* trials), and deteriorated performance with long delays and occasional errors with visible objects (based on its habituation to the task).

Third, the network learns little about the A$\overline{\text{B}}$ apparatus itself with repeated exposures. That is, the network 'sees' the apparatus in the same way from the very start of the experiment, with different locations, covers, and toys clearly demarcated in their activation patterns across only a few units. This constancy in the simplified input, together with the network's built-in bias to represent and respond to presented locations, makes the task of distinguishing different locations rather simple for the network. As a result, the network benefits little in this task from additional trials. Instead, additional trials primarily serve to influence the strength of the network's prepotent response, such that additional *A* trials lead to a stronger A$\overline{\text{B}}$ error. However, it is unclear whether infants make more A$\overline{\text{B}}$ errors following more *A* trials. The number of *A* trials did not emerge as a factor in Wellman *et al.*'s (1986) meta-analysis. Subsequently, Smith *et al.* (1995) showed that the number of reaches to *A* does influence A$\overline{\text{B}}$ errors in two ways: 1) when infants reach to a neutral location using a different apparatus on pre-trials, they show fewer A$\overline{\text{B}}$ errors, and 2) when infants reach spontaneously to one location of an A$\overline{\text{B}}$ apparatus (without the experimenter drawing their attention to it), the number of these 'A' trials does predict how readily infants switch to reach to the other, 'B' location when the experimenter draws attention to it. This pattern of results might be due to a combination of factors associated with additional 'A' experience,

including a stronger prepotent response (leading to more A$\overline{\text{B}}$ errors), increased ability to distinguish *A* and *B* (leading to fewer A$\overline{\text{B}}$ errors), and habituation to the task (leading to deteriorated performance) which might be reduced when the infant is allowed to reach spontaneously or when the apparatus is changed after pre-trials. In its current form, the model incorporates only the first of these factors.

Fourth, the network represents space in a very limited way. It represents only the three potential hiding locations in the A$\overline{\text{B}}$ task, and nothing about other possible locations in the display, such as the table surface. Because of this limitation, the network would capture Bremner's (1978a) finding that different colors to the table surface do not change infants' A$\overline{\text{B}}$ performance; however, the network could not capture Smith *et al.*'s (1995) finding that infants' performance does change when the experimenter taps a marker on the table surface, away from the A$\overline{\text{B}}$ apparatus. In addition, the network does not represent proximity relations, for example, the fact that location *B* is closer to *A* than *C* is, and so could not capture aspects of spatial gradients in infants' errors (Diamond *et al.*, 1994; Bjork & Cummings, 1984). With its fairly limited representation of space, the network captures many of the fundamental aspects of infants' perseverative behavior, but a richer representation of space will be needed to account for additional findings.

Future simulations will incorporate these aspects of cognitive processing and test their effects on A$\overline{\text{B}}$ performance. It is important to note however, that even with such further enhancements, this model – like all models – will by necessity ignore certain aspects of the interface between brain and behavior. For instance, the model will not simulate the effects of neurotransmitters (though other models do) nor anything approximating the complexity of reaching (though other models do). However, these limitations do not detract from the model's main contribution: an explicit set of processing mechanisms that provide a unifying framework for understanding infants' perseveration. Challenges to the current theory, and to the model that instantiates it, must take the form of alternative theories that can account for the data without appealing to the active-latent distinction and gradedness of representations that are fundamental to the proposed theory.

References

Ahmed, A., & Ruffman, T. (1997). *Why do infants make* A *not* B *errors in a search task, yet show memory for the location of hidden objects in a non-search task?*. Manuscript submitted for publication.

Baillargeon, R. (1993). The object concept revisited: New directions in the investigation of infants' physical knowledge. In C. Granrud (Ed.), *Visual perception and cognition in infancy. Carnegie Mellon Symposia on Cognition.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Baillargeon, R., DeVos, J., & Graber, M. (1989). Location memory in 8-month-old infants in a non-search A$\overline{\text{B}}$ task: Further evidence. *Cognitive Development*, **4**, 345–367

Baillargeon, R., & Graber, M. (1988). Evidence of location memory in 8-month-old infants in a non-search A$\overline{\text{B}}$ task. *Developmental Psychology*, **24**, 502–511.

Baillargeon, R., Graber, M., DeVos, J., & Black, J. (1990). Why do young infants fail to search for hidden objects? *Cognition*, **36**, 255–284.

Bjork, E. L., & Cummings, E. M. (1984). Infant search errors: Stage of concept development or stage of memory development. *Memory and Cognition*, **12**(1), 1–19.

Brainerd, C., & Kingma, J. (1985). On the independence of short-term memory and working memory in cognitive development. *Cognitive Psychology*, **17**, 210–247.

Brainerd, C., & Reyna, V. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, **10**, 3–47.

Brainerd, C., & Reyna, V. (1993). Memory independence and memory interference in cognitive development. *Psychological Review*, **100**(1), 42–67.

Bremner, J. G. (1978a). Egocentric versus allocentric spatial coding in nine-month-old infants: Factors influencing the choice of code. *Developmental Psychology*, **14**, 346–355.

Bremner, J. G. (1978b). Spatial errors made by infants: Inadequate spatial cues or evidence of egocentrism? *British Journal of Psychology*, **69**, 77–84.

Bremner, J. G., & Knowles, L. S. (1984). Piagetian stage IV errors with an object that is directly accessible both visually and manually. *Perception*, **13**, 307–314.

Butterworth, G. (1974). *The development of the object concept in human infants.* Unpublished doctoral dissertation, University of Oxford.

Butterworth, G. (1977). Object disappearance and error in Piaget's stage IV task. *Journal of Experimental Child Psychology*, **23**, 391–401.

Carey, S., & Gelman, R. (Eds.). (1991). *The epigenesis of mind.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cohen, J. D., Braver, T. S., & O'Reilly, R. C. (1996). A computational approach to prefrontal cortex, cognitive control, and schizophrenia: Recent developments and current challenges. *Philosophical Transactions of the Royal Society, Series* B, **351**, 1515–1527.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing model of the stroop effect. *Psychological Review*, **97**(3), 332–361.

Cohen, J. D., & O'Reilly, R. C. (1996). A preliminary theory of the interactions between prefrontal cortex and hippocampus that contribute to planning and prospective memory. In M. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.), *Prospective memory: Theory and applications.* Hove, England: Lawrence Earlbaum Associates.

Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, **386**, 604–608.

Cohen, J. D., Romero, R. D., Farah, M. J., & Servan-Schreiber, D. (1994). Mechanisms of spatial attention: The relation of macrostructure to microstructure in parietal neglect. *Journal of Cognitive Neuroscience*, **6**(4), 377–387.

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review*, **99**(1), 45–77.

Cummings, E. M., & Bjork, E. L. (1983a). Perseveration and search on a five-choice visible displacement hiding task. *Journal of Genetic Psychology*, **142**, 283–291.

Cummings, E. M., & Bjork, E. L. (1983b). Search behavior on multi-choice hiding tasks: Evidence for an objective conception of space in infancy. *International Journal of Behavioral Development*, **6**, 71–87.

Dehaene, S., & Changeux, J.-P. (1989). A simple model of prefrontal cortex function in delayed-response tasks. *Journal of Cognitive Neuroscience*, **1**(3), 244–261.

Diamond, A. (1983). Behavior changes between 6–12 months of age: What can they tell us about how the mind of the infant is changing? *Dissertation Abstracts International*, **44**(01B), 337. (University Microfilms No. AAD8311882).

Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by infants' performance on A$\overline{\text{B}}$. *Child Development*, **56**, 868–883.

Diamond, A. (1991). Neuropsychological insights into the meaning of object concept development. In (Carey & Gelman, 1991), Chap. 3, pp. 67–110.

Diamond, A., Cruttenden, L., & Neiderman, D. (1994). A$\overline{\text{B}}$ with multiple wells:: 1. Why are multiple wells sometimes easier than two wells? 2. Memory or memory + inhibition? *Developmental Psychology*, **30**(2), 192–205.

Diamond, A., & Goldman-Rakic, P. S. (1986). Comparative development in human infants and infant rhesus monkeys of cognitive functions that depend on prefrontal cortex. *Neuroscience Abstracts*, **12**, 742.

Diamond, A., & Goldman-Rakic, P. S. (1989). Comparison of human infants and rhesus monkeys on Piaget's A$\overline{\text{B}}$ task: Evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research*, **74**, 24–40.

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

Evans, W. F. (1973). The stage-IV error in Piaget's theory of object concept development: An investigation of the role of activity. *Dissertation Abstracts International*, **34**(11B), 5651. (University Microfilms No. AAD74–11845).

Fischer, K. W., & Bidell, T. (1991). Constraining nativist inferences about cognitive capacities. In (Carey & Gelman, 1991), Chap. 7, pp. 199–236.

Flavell, J. H. (1984). Discussion. In R. J. Sternberg (Ed.), *Mechanisms of cognitive development*. New York: Freeman.

Fuster, J. (1989). *The prefrontal cortex*. New York: Raven Press, 2nd edition.

Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In *Handbook of physiology: The nervous system V* (Chap. 9, pp. 373–417).

Gratch, G., Appel, K., Evans, W., LeCompte, G., & Wright, N. (1974). Piaget's stage IV object concept error: Evidence of forgetting or object conception? *Child Development*, **45**, 71–77.

Harris, P. L. (1974). Perseverative search at a visibly empty place by young infants. *Journal of Experimental Child Psychology*, **18**, 535–42.

Harris, P. L. (1986). Bringing order to the A-not-B error: Commentary on Wellman, Cross, and Bartsch (1986). *Monographs of the Society for Research in Child Development*, **51**(3, Serial No. 214).

Harris, P. L. (1989). Object permanence in infancy. In A. Slater, & G. Bremner (Eds.), *Infant development* (Chap. 5, pp. 103–121). Hove, England: Lawrence Erlbaum Associates, Inc.

Hofstadter, M. C., & Reznick, J. S. (1996). Response modality affects human infant delayed-response performance. *Child Development*, **67**, 646–658.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA,* **79**, 2554–2558.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. USA,* **81**, 3088–3092.

Horobin, K., & Acredolo, L. (1986). The role of attentiveness, mobility history, and separation of hiding sites on stage IV search behavior. *Journal of Experimental Child Psychology*, **41**, 114–127.

Johnson, M. H., & Morton, J. (1991). *Biology and cognitive development: The case of face recognition*. Oxford: Blackwell.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge, MA: MIT Press.

Lecuyer, R., Abgueguen, I., & Lemarie, C. (1992). 9- and 5-month-olds do not make the AB̄ error if not required to manipulate objects. In (Rovee-Collier, 1992).

Matthews, A. (1992). Infants' performance on two versions of AB̄: Is recall memory a critical factor? In (Rovee-Collier, 1992).

McClelland, J. L. (1994). The interaction of nature and nurture in development: A parallel distributed processing perspective. In P. Bertelson, P. Eelen, & G. D'Ydewalle (Eds.), *Current advances in psychological science: Ongoing research*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

Munakata, Y. (1997). Perseverative reaching in infancy: The roles of hidden toys and motor history in the AB̄ task. *Infant Behavior and Development*, **20**(3), 405–416.

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review*, **104**(4), 686–713.

O'Reilly, R. C., & McClelland, J. L. (1992). *The self-organization of spatially invariant representations* (Parallel Distributed Processing and Cognitive Neuroscience PDP. CNS. 92. 5). Carnegie Mellon University, Department of Psychology.

Petrides, M. (1989). Frontal lobes and memory. In F. Boller, & J. Grafman (Eds.), *Handbook of neuropsychology*, Vol. 2 (Chap. 3, pp. 75–90). New York: Elsevier.

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

Roberts, R., Hager, L., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General*, **123**(4), 374–393.

Rovee-Collier, C. (Ed.). (1992). *Abstracts of papers presented at the eighth international conference on infant studies*, Norwood, New Jersey. Ablex Publishing Corporation.

Schuberth, R. E., Werner, J. S., & Lipsitt, L. P. (1978). The stage IV error in Piaget's theory of object concept development: A reconsideration of the spatial localization hypothesis. *Child Development*, **49**, 744–748.

Siegler, R. (1996). *Emerging minds: The process of change in children's thinking*. New York: Oxford University Press.

Smith, L. B., McLin, D., Titzer, B., & Thelen, E. (1995). The task dynamics of the A-not-B error. In L. B. Smith (Chair), *Tests of a dynamic systems theory: The object concept*. Symposium conducted at the 1995 Meeting of the Society for Research in Child Development, Indianapolis, IN.

Smith, L. B., & Thelen, E. (Eds.). (1993). *A dynamic systems approach to development: Applications*. Cambridge, MA: MIT Press.

Sophian, C., & Yengo, L. (1985). Infants' search for visible objects: Implications for the interpretation of early search errors. *Journal of Experimental Child Psychology*, **40**, 260–278.

Steinmetz, M., Connor, C., Constantinidis, C., & McLaughlin, J. (1994). Covert attention suppresses neuronal responses in area 7a of the posterior parietal cortex. *Journal of Neurophysiology*, **72**(2), 1020–1023.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Wellman, H. M., Cross, D., & Bartsch, K. (1986). Infant search and object permanence: A meta-analysis of the A-Not-B error. *Monographs of the Society for Research in Child Development*, **51**(3, Serial No. 214).

Willatts, P. (1990). Development of problem-solving strategies in infancy. In D. F. Bjorklund (Ed.), *Children's strategies: Contemporary views of cognitive development*

(Chap. 2, pp. 23–66). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Author note

Yuko Munakata, Department of Psychology.