The role of prefrontal cortex in perseveration:
Developmental and computational explorations

Yuko Munakata, J. Bruce Morton, and Jennifer Merva Stedron

Department of Psychology, University of Denver

Preparation of this chapter was supported by research grants from NICHD (1R29 HD37163-01) and NSF (IBN-9873492).

To appear in P. Quinlan (Ed.), *Connectionist Models of Development*, East Sussex: Psychology Press.

Introduction

One of the hallmarks of "higher" intelligence is the ability to act flexibly and adaptively, rather than being governed by simple habit. For example, we may drive the same route to and from work each day, but we can pull ourselves out of this habit when necessary (e.g., to stop at the store on the way home). Many findings point to the critical role of the prefrontal cortex in such flexible behavior. However, the exact nature of its role is uncertain. In this chapter, we use neural network models to explore the role of the prefrontal cortex in the development of flexible behavior in the first years of life.

Infants often demonstrate a lack of flexibility by *perseverating*, repeating prepotent or habitual behaviors when they no longer make sense. For example, as soon as infants will search for a toy that is presented and then hidden, they search perseveratively, continuing to reach back to old hiding locations after watching as the toy is hidden in a new location (Diamond, 1985; Piaget, 1954). Infants will even perseverate when objects are fully visible in front of them. For example, when faced with two towels to pull – one with a distant toy on it and one with a toy behind it – infants will choose the towel with the toy on it. However, if the towels are switched so that the towel that was to the infants' left (e.g., with the toy on it) is now to the infants' right, infants perseverate, continuing to pull the towel on the same side as before though it does not yield the toy (Aguiar & Baillargeon, 2000).

These perseverative behaviors are not limited to infancy; children also demonstrate them quite reliably. For example, most 3-year-olds can correctly sort cards according to experimenter instructions (e.g., with all of the blue cards going into one pile, and all of the red cards going into another pile). However, when instructed to switch to a different sorting rule (e.g., to sort the cards by their shape rather than their color, with all of the trucks going into one pile, and all of the flowers going into another pile), 3-year-olds perseverate, continuing to sort by the initial instructions (Zelazo, Frye, & Rapus, 1996). Six-year-olds show the same pattern when asked to judge a speaker's feelings from utterances with conflicting emotional cues (e.g., a sentence with happy content – "I won a prize" spoken in a sad tone of voice) (Morton & Trehub, 2001). When instructed to judge the speaker's feelings from the content of her utterances, most 6-year-olds succeed. However, when instructed to switch and judge the speaker's feelings from her tone of voice, many 6-year-olds perseverate, continuing to base their judgments on content (Morton & Munakata, in press-b; Morton, Trehub, & Zelazo, in preparation).

In all of these cases, infants and children appear quite sensible in their initial behaviors – searching in the correct location for the hidden toy, pulling the appropriate towel, and sorting cards and judging utterances according to experimenter instructions. However, they appear quite inflexible in their subsequent behaviors, perseverating with their previous responses when they no longer make sense – searching in the incorrect location for the hidden toy (making the "A-not-B" error), pulling the inappropriate towel, and sorting cards and judging utterances without apparent regard for the experimenter's current instructions.

Interestingly, even as infants and children perseverate with their previous responses, they sometimes seem to indicate through other measures that they have some awareness of the correct response. That is, they show *dissociations* in their behavior. For example, even as infants reach perseveratively to a previous hiding location for a toy, they occasionally gaze at the correct hiding location (Piaget, 1954; Diamond, 1985; Hofstadter & Reznick, 1996). Further, in violation-of-expectation variants of the A-not-B task, 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, 1998). Perhaps even more compelling, even as children sort perseveratively according to a previous rule, they can correctly answer questions about the new rule they should be using, such as where trucks should go in the shape game (Zelazo et al., 1996), or what aspect of a speaker's voice they should be listening to (Morton & Munakata, in press-b; Morton et al., in preparation). These dissociations in infants' and children's behavior provide an important constraint on theories of perseveration. Further, behavioral dissociations are a salient aspect of development more generally (e.g., Berthier, DeBlois, Poirier, Novak, & Clifton, 2000; Hood & Willatts, 1986; Piaget, 1952; Spelke, Breinlinger, Macomber, & Jacobson, 1992). Understanding such dissociations may thus be an important step in understanding the development and organization of our cognitive systems (Munakata, 2001a; Munakata & Stedron, in press).

Another important constraint on theories of perseveration is the remarkable *decalage* (Piaget, 1941, 1967), or temporal uncoupling of similar cognitive developments (Flavell, 1963), observed across various tasks. That is, infants succeed in the A-not-B task years before they succeed in the card-sorting task, and children succeed in the card-sorting task years before they succeed in the speech interpretation task. Children's apparent abilities to overcome perseveration and respond flexibly thus depend heavily on exactly what task they face.

The prefrontal cortex plays a critical role in reducing perseveration and in supporting flexible behavior more generally (Diamond, in press; Miller & Cohen, 2001; Miyake & Shah, 1999; O'Reilly, Braver, & Cohen, 1999; Roberts & Pennington, 1996; Stuss & Benson, 1984). For example, impaired prefrontal functioning often leads to perseverative behaviors. Some of the most dramatic examples come from human adults with prefrontal damage, who may exhibit Environmental Dependency Syndrome, inappropriately carrying out habitual responses supported by particular environmental stimuli (Lhermitte, 1986). For example, one such patient, upon entering her physician's home and seeing dirty dishes in the kitchen, began to wash them. Another patient, upon seeing paintings on the floor with a hammer and nails, began hanging the paintings. Such patients have also put on multiple pairs of glasses when presented with them individually, and even used a makeshift urinal when presented with it (Lhermitte, 1986). In all of these cases, patients simply carried out prepotent responses to particular stimuli, rather than responding flexibly based on the particular context, in which it would have been more appropriate to inhibit those behaviors. The prefrontal cortex has also been implicated in more systematic tasks like those described above. For example, patients with prefrontal damage perseverate in tasks like Zelazo et al.'s (1996) card sorting task (Milner, 1963). When the sorting rule is changed, patients respond based on habit (the first sorting rule) rather than responding flexibly to feedback indicating that the rule has changed. Similarly, adult monkeys with lesions to the prefrontal cortex perseverate in the A-not-B task (Diamond & Goldman-Rakic, 1989, 1986). And, human infants' eventual success in this task is correlated with event-related potential measures recorded over frontal cortex (Bell & Fox, 1992).

Thus, there is general agreement *that* the prefrontal cortex plays a role in reducing perseveration, *that* there is a decalage in when children succeed in overcoming perseveration, and *that* behavioral dissociations emerge in tasks that require flexibility, with one measure yielding perseveration and another measure suggesting awareness of the correct response. In this chapter, we use neural network models to explore three remaining questions about perseveration in infancy and childhood:

- *How* does prefrontal development reduce perseveration?
- *Why* are dissociations observed in perseveration?
- *Why* do children show the remarkable decalage in their flexibility, overcoming perseveration at such different ages across various tasks?

We consider each of these questions in turn, and close by comparing the answers that emerge from our neural network explorations to other accounts of perseveration.

Prefrontal Cortex and Perseveration

Neural network models have helped to understand how the development of the prefrontal cortex can reduce perseveration (Dehaene & Changeux, 1989, 1991; Morton & Munakata, in press-a; Munakata, 1998; Stedron, Munakata, & Sahni, 2002). In this chapter, we explore an account based on a distinction between “active” and “latent” memory traces (Munakata, 2001b). In the neural network framework, active traces take the form of sustained activations of network processing units (roughly corresponding to the firing rates of neurons), and latent traces take the form of changes to connection weights between units (roughly corresponding to the efficacy of synapses). According to the active-latent account:

- Perseveration occurs based on a competition between latent memory traces for previously relevant information and active memory traces for current information.
- Latent memory traces, subserved primarily by posterior cortex, result when organisms change their biases toward a stimulus after processing it, so that they may respond differently to the stimulus on subsequent presentations. These latent traces are not accessible to other brain areas, because synaptic changes in one part of the brain cannot be communicated to other areas. Rather, latent traces can only influence processing elsewhere in the system in terms of how they affect the processing of subsequent stimuli, and resulting patterns of activation.
- Active memory traces, subserved primarily by prefrontal cortex, result when organisms actively maintain representations of a stimulus. Unlike latent traces, such active representations may be accessible to other brain areas in the absence of subsequent presentations of the stimulus, because neuronal firing in one region can be communicated to other areas.
- Flexible behavior can be understood in terms of the relative strengths of latent and active memory traces. The increasing ability to maintain active traces of current information, dependent on developments in prefrontal cortex, leads to improvements in performance on tasks such as A-not-B.

Behavioral and physiological data motivate the active-latent distinction central to the proposed theory of perseveration. For example, neurons can “remember” information in two distinct ways: through sustained firing for a stimulus (active), or through changes in firing thresholds or synapses that affect neurons’ subsequent firing to a stimulus (latent). 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 (Miller, Erickson, & Desimone, 1996). This active memory is consistent with a number of neural recording and imaging experiments in the prefrontal cortex (e.g., Cohen, Perlstein, Braver, Nystrom, Noll, Jonides, & Smith, 1997; Fuster, 1989; Goldman-Rakic, 1987a). In contrast, on an easier task that required memory for any familiar stimulus, monkeys solved the task using 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 (Miller & Desimone, 1994). 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. Neurons in prefrontal and posterior parietal cortex have shown the same active-latent distinction for memories of spatial locations (Steinmetz, Connor, Constantinidis, & McLaughlin, 1994). 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, whereas working memory requires information to be kept active for manipulation.[1]

The active-latent account shares several features with and builds upon existing accounts and computational models of perseveration. Many accounts and models similarly posit perseveration to arise based on a competition between two kinds of information, and describe something akin to the latent or active elements of the active-latent account (e.g. Butterworth, 1977; Dehaene & Changeux, 1991; Diamond, 1985; Harris, 1986; Wellman, Cross, & Bartsch, 1986; Thelen, Schoner, Scheier, & Smith, 2001). Further, many accounts and models have similarly emphasized working memory as a primary mechanism of prefrontal cortex, with other functions (e.g., inhibition of prepotent responses) emerging from this more basic mechanism (Cohen & Servan-Schreiber, 1992; Goldman-Rakic, 1987b; Kimberg & Farah, 1993; Miller & Cohen, 2001; O’Reilly et al., 1999; Roberts, Hager, & Heron, 1994). In this view, prefrontal cortex subserves the ability to maintain and manipulate information in working memory, in the absence of supporting stimuli (e.g., across delays) or in the face of interfering stimuli. Prefrontal cortex can thus serve to represent task-appropriate information, such as the most recent hiding location in the A-not-B task or the most recent rule in the card sorting task. These representations support flexible behaviors, and via competitive interactions throughout the cortex, the representations supporting inappropriate or habitual behaviors are inhibited.

We explore the active-latent account by testing neural network models in all of the developmental tasks described above: the classic A-not-B task (Piaget, 1954), the towel-pulling task (Aguiar & Baillargeon, 2000), the card sorting task (Zelazo et al., 1996), and the speech interpretation task (SIT, Morton & Munakata, in press-b; Morton et al., in preparation). The models provide a unified framework for understanding perseveration across a range of conditions (e.g., with hidden or visible objects, with or without explicit rules) and ages (from infancy through childhood). The models also lead to novel behavioral predictions. As outlined above, this unified approach to perseveration shares much in common with existing approaches; however, the active-latent account contrasts with other theories, as we elaborate in the Discussion section.

---

[1] We view active memory as only one component of (rather than equivalent to) working memory.

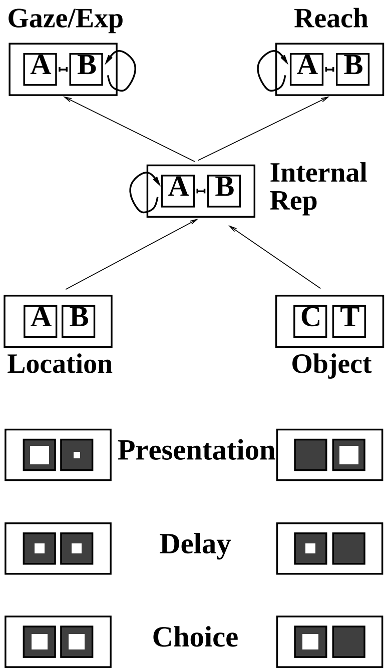

Figure 1: Simplified version of the A-not-B network and the elements of an A trial (adapted from Munakata, 1998a): The activation level of the input units for the three segments of the trial is shown by the size of the white boxes. The "Object" input indicated whether a cover ("C") or toy ("T") was visible.

A-not-B

This simulation (Munakata, 1998) explored infants' perseveration in searching for hidden objects. According to the active-latent account, after infants repeatedly attend to a hiding location and reach there, they lay down latent traces biasing them toward that location, making them more likely to reach there in the future. To overcome that tendency and reach to a new location, infants must maintain active memory traces for the most recent hiding location of an object. The full simulation evaluated many variants of the A-not-B task not covered here, so we simplify the presentation of the network architecture and environment to include only those elements covered in the simulations described in this section.

Architecture and environment

The network was comprised of two input layers that encoded information about the location and identity of objects, an internal representation layer, and two output layers for gaze/expectation and reach (Figure 1). The gaze/expectation layer responded (i.e., updated the activity of its units) to every input during the A-not-B task, while the reaching layer responded only to inputs corresponding to a stimulus within "reaching distance". This updating constraint was meant to capture the fact that infants' reaching is permitted at only one point during each A-not-B trial, when the apparatus is moved to within the infant's reach, whereas nothing prevents infants from forming expectations (which may underlie longer looking to impossible events) throughout each trial.[2]

---

[2] The model simplifies over nuances in infants' gazing and reaching. For example, infants' gaze is sometimes restricted during A-not-B experiments so that they cannot gaze continuously at a recent hiding location, whereas the model gazes continuously. And, infants may plan or imagine reaching movements prior to the point when they can reach to the A-not-B apparatus, so that they may activate brain areas relevant for reaching to some degree prior to the actual reach, whereas the model cannot activate its reaching

The network's feedforward connectivity included an initial bias to respond appropriately to location information, and also developed further biases based on the network's experience during the A-not-B task. The initial bias allowed the network, for example, to look to location A if something were presented there. Infants enter A-not-B experiments with such biases, so this manipulation may be viewed as a proxy for experience prior to the A-not-B study. The network's feedforward weights changed based on its experience during the study according to a Hebbian learning rule, such that connections between units that were simultaneously active tended to be relatively strong. The network's latent memory thus took the form of these feedforward weights; they reflected the network's prior experiences and influenced its subsequent processing.

Each unit in the hidden and output layers had a self-recurrent excitatory connection back to itself. These recurrent connections were largely responsible for the network's ability to maintain representations of a recent hiding location; units that are active tend to remain active when they receive their own activity as input through sufficiently large weights. The network's active memory thus took the form of maintained representations on the network's hidden and output layers, as supported by its recurrent connections. To simulate gradual improvements with age in the network's active memory, the strength of the network's recurrent connections was manipulated, with "older" networks having higher recurrence. This manipulation might be viewed as a proxy for experience-based weight changes that have been explored elsewhere (e.g., Munakata, McClelland, Johnson, & Siegler, 1997).

The simulated A-not-B task consisted 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. Each trial consisted primarily of three segments: the *presentation* of a toy at the A or B location, a *delay* period, and a *choice* period (Figure 1). During each segment, patterns of activity were presented to the input units corresponding to the visible aspects of the stimulus event. The levels of input activity represented the salience of aspects of the stimulus, with more salient aspects producing more activity. For example, the levels of input activity for the A and B locations were higher during *choice* than during *delay*, to reflect the increased salience of the stimulus when it was presented for a response.

Performance and internal representations

For all analyses of the network's performance, the network's percent correct response was computed as the activation of the appropriate output unit divided by the sum of activation over all possible output units. For example, the network's percent correct reaching on A trials was computed as the activity of $\frac{A}{A+B}$ for the reaching layer. The model simulated the A-not-B error (successful reaching on A trials with perseverative reaching on B trials), and improvements with age (Figure 2). The model also showed earlier sensitivity on B trials in its gaze/expectation than in its reach, a finding that we will return to in the Dissociations section.

The network performed well on A trials at all ages because latent changes to the feedforward weights, built up over previous trials in which the network represented and responded to A, favored A over B. These latent memories thus supported enough activity at A that the network's ability to maintain activity at A had little effect on performance. The internal representations of a relatively young network during an A trial (Figure 3) showed that even with relatively weak recurrent weights to support the active maintenance of the most recent hiding location, the network was able to strongly represent A during all three segments of the

---

units until the actual reach. Nevertheless, the model captures an essential difference between gaze/expectation and reach in the A-not-B task – infants have many more opportunities to gaze and to form expectations than to execute reaching responses.

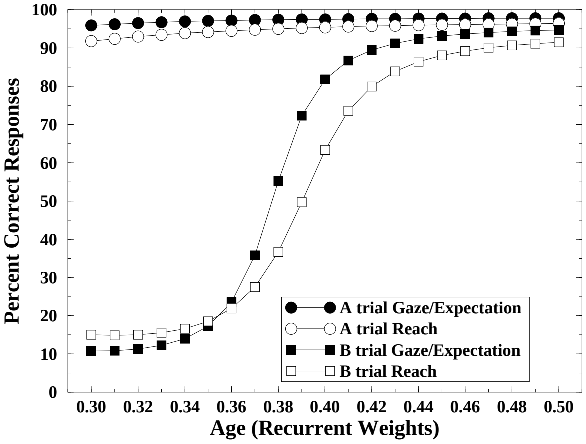

Figure 2: Percent correct responses as a function of age: On A trials, the network is accurate across all levels of recurrence shown. On B trials, the network responds non-perseveratively only as the recurrent weights get stronger.

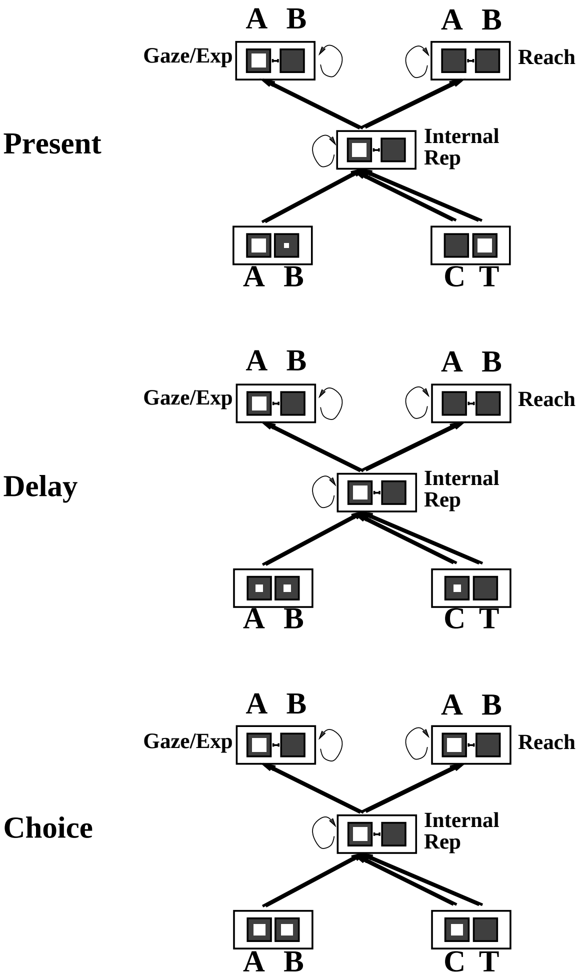

Figure 3: A young network's representations during an A trial (adapted from Munakata, 1998a): Only the strongest of the feedforward weights are shown; these reflect the network's latent bias toward A that developed during the practice trials. The network responds correctly to A in gaze/expectation and in reach.

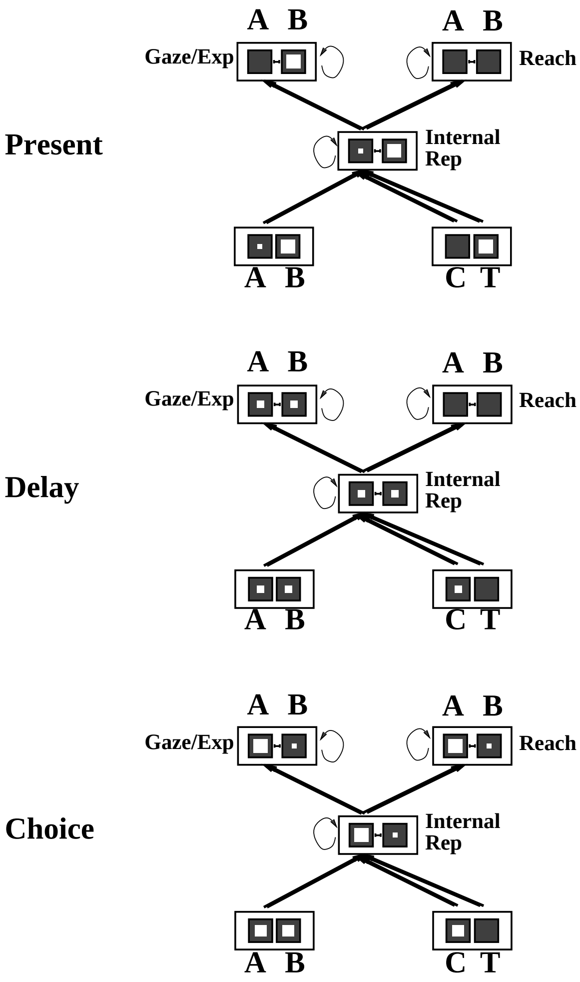

Figure 4: A young network's representations during a B trial (adapted from Munakata, 1998a): Again, only the strongest of the feedforward weights are shown. The weaker weights (not shown) that support a correct response to B allow the network to represent and gaze at B when a toy is presented there (as infants do). However, after the toy is hidden, the network's weak ability to maintain an active representation of B cannot compete against the network's latent bias toward A (reflected in the strong feedforward weights shown). The network thus responds perseveratively to A in gaze/expectation and in reach.

trial. Thus, the latent memories in the network's weights, biasing it toward A, allowed it to respond correctly toward A even with only a weak ability to actively hold the most recent hiding location in mind.

In contrast, the network's ability to maintain activity for the most recent hiding location was critical to its performance on B trials, because the network had to maintain a representation of B in the face of the latent bias to respond to A. The activity of the units of the young network during a B trial (Figure 4) indicated that the network appropriately represented and responded to B in gaze/expectation during the presentation of the toy at B, when the B input unit was strongly activated. Infants in the A-not-B task similarly look to B when an object is hidden there. However, the memory for B faded during the *delay*, when the A and B input units were equally activated, so that the internal representation activity showed little evidence of which location was recently-attended. If judged at that time on the basis of active traces alone, the network showed little memory of prior events. However, the network showed strong evidence of memory for the previous trials by making the A-not-B error at *choice*, indicating the influence of latent traces. In particular, the network's connection weights had learned to favor activity at A over B, based on the preceding pre-trials and A trials. Thus, by repeatedly attending and responding to one location, the network became increasingly likely to attend and respond there. Stronger recurrent weights allowed an older network to maintain an active memory of B during the delay. That is, the older network was better able to hold information about a recent hiding location in mind, rather than simply falling back to its biases for previous locations.

Predictions

The A-not-B model led to the novel prediction that infants' may show a U-shaped pattern of performance in their perseveration, at first showing worse performance with increasing age and then better performance. That is, networks at the youngest ages (weakest recurrent weights) in Figure 2 showed *more* perseveration than slightly younger networks. This prediction of U-shaped performance has since been supported (Clearfield & Thelen, 2000).

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.

Towel-pulling

This simulation (Stedron et al., 2002) explored infants' perseveration when retrieving visible objects. As described earlier, infants presented with two towels, one with a distant toy on it and one with a toy behind it, will pull the correct towel to retrieve the toy, but perseverate and pull the incorrect towel (with the toy behind it) when the toy's location is switched (Aguiar & Baillargeon, 2000). Infants also perseverate in other tasks with fully visible objects. For instance, when presented with a toy inside a transparent box, infants will perseveratively attempt to reach the toy through the closed top of the box, rather than through an open side (Diamond, 1981). Perhaps counterintuitively, a competition between active and latent memory traces can also account for such perseverative behaviors with visible objects. Latent memory traces result from repeated behaviors (e.g., pulling the initially correct towel) or prepotent tendencies (e.g., to reach directly for visible objects). Active memory traces represent currently relevant information (e.g., which towel should be pulled, how the toy in the transparent box can be retrieved). When latent memory traces are stronger than the active memory traces, infants perseverate.

Why is active memory helpful with fully visible objects? Active memory can bolster representations of fully visible information, allowing one to attend to the critical features of the environment (e.g., De Fockert, Rees, Frith, & Lavie, 2001). For example, an adult may not normally attend to the color of a companion's shirt, but active memory may help the adult attend to that feature if he or she is trying to follow the companion through a large crowd. Similarly, in the towel-pulling task, active memory can help infants' attend to which towel currently holds the toy. The stronger the representation of this critical information, the more

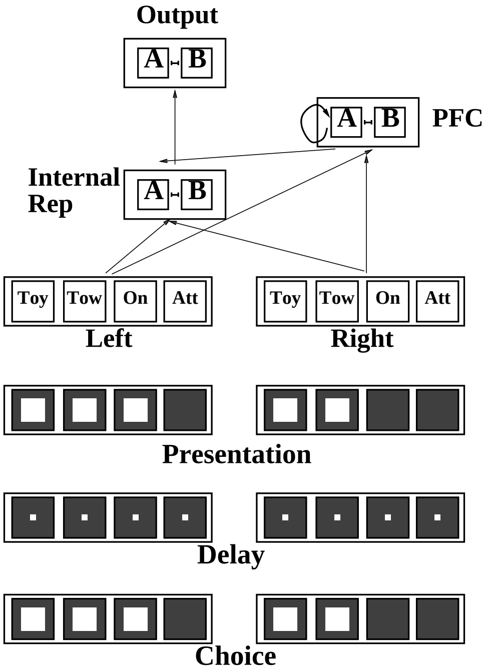

Figure 5: The towel-pulling network and the elements of an A trial: The input units encoded information about the identity of objects (toy and towel) and their placement (on and attached). A toy sitting behind a towel would activate the toy and towel units only (as on the right side of the trial shown). A toy sitting on a towel would activate the toy, towel, and on units (as on the left side of the trial shown). A toy that attached to its supporting towel would activate the toy, towel, on, and attached units; this condition will be discussed in the Decalage section.

likely active memory will prevail over latent memory, enabling the infant to reach to the correct towel. We explore these ideas in a simulation of infant behavior in the towel-pulling task.

Architecture and environment

The network was comprised of two input layers that encoded information about the location, identity and placement (e.g., toy *on* towel) of the objects, and three layers that represented the two locations of the objects: an internal representation layer (hidden layer), a prefrontal cortex (PFC) layer, and an output layer for reaching (Figure 5).

Feedforward connections linked the network's input layers to the hidden and PFC layers, and the hidden layer to the output layer. This feedforward connectivity included an initial bias to accurately encode the location of various aspects of the display. For example, the weight from the left "On" input unit was strongly connected to the left hidden unit and the left PFC unit, and the weight from the left hidden unit was strongly connected to the left output unit. This initial bias allowed the network to represent and respond to the location of a toy on a towel preferentially over a toy behind a towel. Infants appear to bring such a bias into the towel-pulling task. Like the A-not-B model, the network developed a bias (latent traces in the form of stronger feedforward weights) for the towel location that originally supported the toy based on its experience with the task, according to a Hebbian learning rule.

Each prefrontal unit had a self-recurrent excitatory connection back to itself, and an excitatory connection to the corresponding hidden unit. These recurrent connections were largely responsible for the network's ability to maintain an active representation of the current (and visible) location of the towel supporting the toy. This model and subsequent ones thus elaborated the A-not-B model architecture by incorporating a separate prefrontal layer for this active memory function, rather than simply using self-recurrent connections on the hidden layer as a proxy. Again, the network's aging was simulated by strengthening the recurrent connections.

The towel-pulling task consisted of four A trials, in which the same towel (at location A) supported the toy, and one B trial, in which the other towel (at location B) supported the toy. Each trial consisted of three segments: the *presentation* of the toys placed on (or behind) the towels at the A and B locations, a short *delay* in which reach was prevented (simulating a brief period in the behavioral studies when a screen was placed between the infant and the towels to prevent immediate reach), and a *choice* period when the towels and toys were again visible (Figure 5). During each segment, patterns of activity were presented to the input units corresponding to the visible aspects of the stimulus event. Activity was low and equally distributed across all input units during the delay period, to reflect the lack of any particular visual input due to the occluding screen.

Performance and internal representations

As in the A-not-B model, the network's correct response was computed as the activation of the appropriate output unit divided by the sum of activation over both output units (e.g., $\frac{A}{A+B}$ on A trials). The model simulated infants' correct towel-pulling on A trials, their perseverative towel-pulling on B trials, and improvements with age.

The network performed well on A trials at all ages because of the network's initial bias to reach to the towel supporting the toy. Because the network responded to the same towel supporting the toy throughout the four A trials, an initial bias for the towel supporting the toy at the A location was strengthened as a result of latent changes to the feedforward weights.

During the B trials, this latent bias to respond to the old towel competed with the network's ability to strongly represent information about the new towel supporting the toy. As with infants, the information about the two towel choices was fully visible to the network. The total input activation on the correct side was greater than the activation on the previous side, because the "on" unit was active only on the correct side. The recurrent connections were necessary for amplifying this greater activation for the correct side, thus more strongly influencing the activation of the hidden units. In younger networks with weaker recurrent weights, the latent bias for the old location was stronger than the activation for the correct location provided by the PFC layer, leading to perseveration. In older networks with stronger recurrent weights, the stronger PFC representation of the correct location provided stronger input to the hidden layer, and stronger competition against the latent bias for the old towel, leading the network to reach to the correct towel.

Predictions

The towel-pulling model led to the novel prediction that infants at different points in development may show an interesting pattern in how quickly they pull the towel on B trials. In the model, reaction time is measured in terms of processing cycles required before the model settles on a stable response. The network produced a developmental inverted U-shaped reaction time curve in its performance, responding most quickly when it was very young and perseverating and when it was very old and succeeding. In contrast, the network responded slowly at a transitional age, just prior to its first success and just after its first success. These differences in reaction time resulted from differences in the degree of competition between the active representation for the current location of the towel supporting the toy and the latent bias for the old location. Specifically, the more imbalanced this competition, the faster the competition was resolved, and the faster the reach. These results suggest that infants should be fastest when they are either really perseverating or really succeeding (i.e., when they are far from the transition period from perseveration to success), and they should be slowest in the transition period. Thus, the model makes clear developmental predictions about reaction times that we are now testing.

Card sorting and Speech interpretation task

These simulations (Morton & Munakata, in press-a) explored children's perseveration in using prior rules rather than current rules for sorting cards (Zelazo et al., 1996) and judging the emotion of a speaker (Morton & Munakata, in press-b; Morton et al., in preparation). According to the active-latent account, children need to maintain a strong active representation of the current (post-switch) rule to overcome latent biases that are established or strengthened by use of the prior (pre-switch) rule. Failure to do so results in perseveration.

The networks for the two tasks had virtually equivalent architectures, differing only in the strength of initial biases and the labeling of units. Consequently, the underlying causes of perseveration and dissociation were identical for both models. To simplify the presentation, we focus on the results from the card sort model, and save a discussion of the differences between the models for the Decalage section.

Architecture and environment

The networks consisted of three input layers, an internal representation layer, a PFC rule layer, and an output layer (Figures 6 and 7). In the card sorting network, the input layers encoded the shape and color of the test cards, verbal descriptions of these features, and the sorting rule. The two output units represented the sorting trays in which children place the test cards, and were labeled red/flower and blue/truck respectively.

The network's feedforward connectivity included an initial bias to respond appropriately to color and shape information, and also developed further biases based on the network's experience during the card sorting task. The "Red" hidden unit, for example, became active when either the "Red" visual features or the "Red" verbal features units were active, and the "Red/Flower" output became active when either the "Red" or the "Flower" hidden units were active. Children appear to bring such biases with them into the card sorting task. As in the A-not-B and towel-pulling models, these connections changed with experience according to a Hebbian learning rule, such that the network's latent memory took the form of stronger connections between units. And as in those models, the network's aging was simulated by strengthening the recurrent connections supporting the network's active representations.

The simulated card sorting task consisted of 2 demonstrations of the pre-switch rule, 6 trials with the pre-switch rule, and 6 trials with the post-switch rule. Each simulated trial of the card sort task included a statement of the rules followed by a presentation of a test card (Figure 6).

Performance and internal representations

We measured the network's percent correct response as the activation of the appropriate output unit divided by the sum of activation over both output units. For example, the network's percent correct response on a trial with a red flower in the color game was computed as the output activity of $\frac{RF}{RF+BT}$. The network simulated good pre-switch performance at all ages, and perseveration in the post-switch trials with age-related improvements.

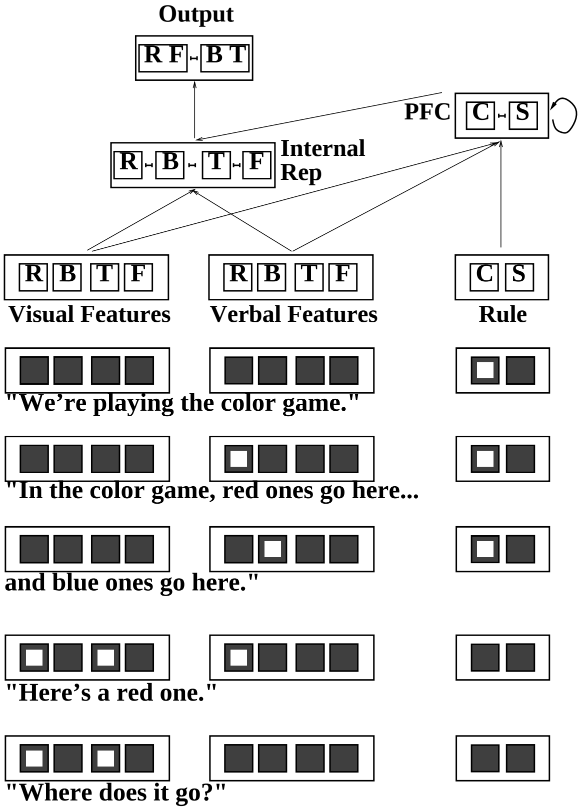

Figure 6: Simplified version of the card sort network and the elements of a trial (R=red, B=blue, T=truck, F=flower, C=color, and S=shape). In the inputs with "go here," the corresponding output unit was activated for the network to indicate where the card should go.

In pre-switch trials, networks at all ages sorted correctly because the demonstration trials had slightly biased the feedforward weights in favor of the preswitch rule. For example, the internal representation of a young network presented with a blue flower showed that even in the absence of a strong representation of the color rule, the network strongly represented the blue aspect but not the flower aspect of the test card in its hidden layer. This, in turn, caused the network to sort in terms of the color rather than the shape of the test card. Continued experience correctly sorting the test cards further strengthened the feedforward weights in favor of the preswitch rule.

The network's age (strength of recurrent connections) played a larger role in performance on the post-switch trials. Young networks were unable to maintain a strong active representation of the shape rule for the full duration of a trial, and therefore were unable to overcome the latent bias to internally represent the color rather than the shape of the test cards. Consequently, young networks responded to color and not shape. In contrast, older networks had stronger recurrent connections that allowed them to maintain a strong active representation of the shape rule. This active representation strengthened the representation of shape in the hidden layer, allowing older networks to overcome the latent bias to represent color.

Predictions

According to the active-latent account, children's ability to rapidly switch rules in response to verbal instructions depends on variations in the strength of active memory. This leads to the prediction that in-

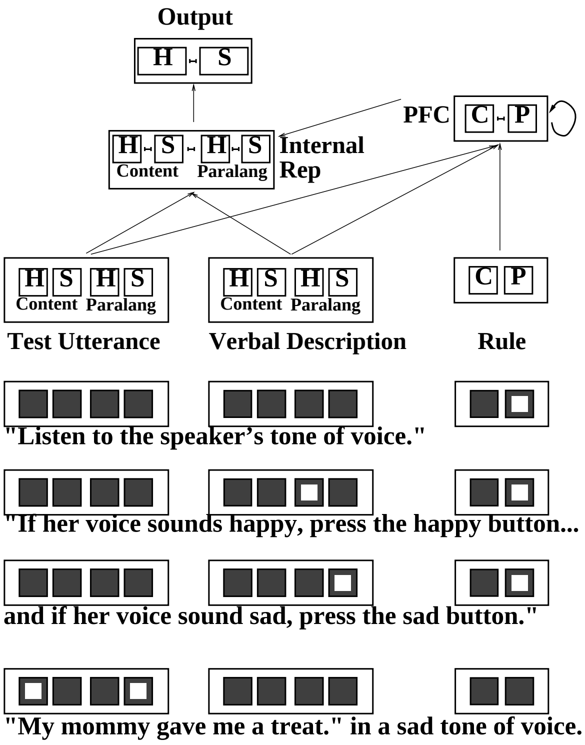

Figure 7: Simplified version of the speech interpretation network and the elements of a trial (H=happy, S=sad, C=content, and P=paralanguage). In the inputs with "press the ... button," the corresponding output unit was activated for the network to indicate which button should be pushed.

creasing the delay between the delivery of a new rule and test should compromise children's ability to switch. A longer delay would give more opportunity for the active representation of the new rule to decay, making it more difficult to overcome a latent bias for using the old rule. Conversely, repeating the new rule more frequently might facilitate switching, due to a strengthening of the active representation of the new rule.

In addition, the model suggests a more effective method than direct instruction for helping perseverating children switch to a new rule. The model predicts that children may be more likely to switch to a new rule (e.g., sorting cards by color after sorting them by shape), if they could gain some experience with the new rule that would strengthen latent memories for it. For example, after sorting cards by shape, children could easily sort cards by color if the cards were completely blue or red, without any shapes on them to conflict with the color cues. According to the model, this experience would strengthen children's latent representations for the new color rule, making them more likely to switch to color when presented with the original cards (e.g., red trucks and blue flowers). In contrast, direct instructions to switch to a new rule would be less effective, if children could not maintain active representations of the new rule. We are currently testing these predictions.

Dissociations

The preceding simulations demonstrate how improvements in active memory can reduce perseveration across a range of conditions (e.g., with hidden objects, fully visible objects, and explicit rules). However, as discussed earlier, infants and children sometimes show dissociations in their behavior, perseverating despite seeming to demonstrate that they know what they should do. How can the problem then be one of active memory? For example, if children can answer questions about the correct card sorting rule, and infants can look at the correct hiding location in the A-not-B task, how could their incorrect perseverative responses reflect limitations in remembering the rule or the hiding location?

This type of challenge builds on the assumption that memory is an all-or-nothing construct – either present or absent. From this standpoint, if individuals succeed on one memory test (e.g., by answering a question about a rule), their memory must be fine, so perseveration must be attributed to other factors.

In contrast, if one allows that various capacities may be graded in nature rather than all or nothing, memory limitations can in fact explain perseveration and dissociations in behavior (reviewed in Munakata, 2001a). That is, memories, perceptions, rules, and so on may vary in their strength rather than simply being present or absent (e.g., Farah, O'Reilly, & Vecera, 1993; McClelland, Rumelhart, & PDP Research Group, 1986; Mathis & Mozer, 1996). Strength might be instantiated by the number of neurons contributing to a particular representation and the firing rates and coherence of those neurons. Weak representations might suffice for some tasks but not others, leading to dissociations in behavior. For example, some degree of memory for a card sorting rule might support the ability to answer questions about the rule, but not to correctly sort a card when faced with the conflicting features present in it. When children perseverate with an old rule (e.g., sorting a red truck by its shape rather than by its color) despite appearing to know the new rule (by correctly answering the question, "Where do red things go in the color game?"), the sorting measure requires resolving a conflict between rules (i.e., deciding where to place an object that is both red and a truck) whereas the knowledge question does not. A weak memory for the color rule might allow children to correctly answer non-conflict questions, but not to respond correctly when presented with the conflict inherent in the sorting task.

Thus, if one allows that memory may be graded in nature, with weaker representations sufficing for some tasks but not others, behavioral dissociations can be understood in terms of memory limitations. We use neural network simulations to explore the role of graded representations in dissociations observed across perseverative tasks.

A-not-B

As shown in Figure 2, the A-not-B model, like infants, showed earlier sensitivity on B trials in its gaze/expectation than in its reaching (Munakata, 1998). This dissociation can be understood by considering a network slightly older than the one shown in Figures 3 and 4. (The younger networks perseverated in both gaze/expectation and reaching.) With stronger recurrent weights, the slightly older network was better able to hold in mind information about a recent location (Figure 8). The gaze/expectation system was able to take advantage of this information with its constant updating, showing correct responding during *presentation* and *delay*, which carried over to *choice*. In contrast, the reach system was only able to respond at *choice*. Because the network's active memory for the most recent location faded with time, by the *choice* point the network's internal representation reflected more of the network's latent memory of A. The gaze/expectation system was thus able to make better use of relatively weak active representations of the recent hiding location.

Similarly, infants may show earlier success in gaze/expectation variants of the A-not-B task because they can constantly update their gazing and their expectations. As a result, they can counter perseverative

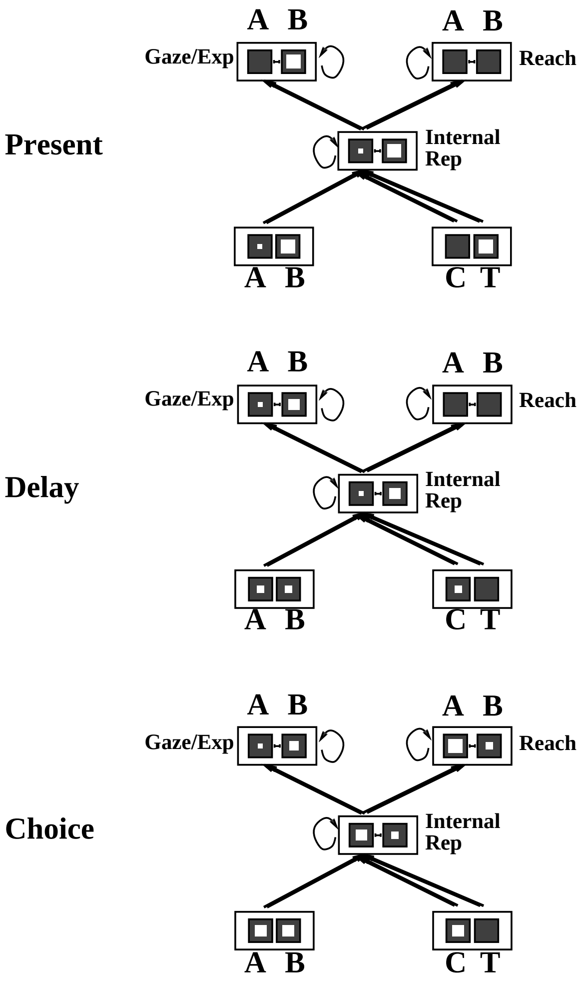

Figure 8: A slightly older network's representations during a B trial (adapted from Munakata, 1998a): The network responds correctly to B in gaze/expectation, but reaches perseveratively to A.

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 memories have become more susceptible to perseverative biases.

Unexpectedly, very young networks showed slightly *more* perseveration in gaze/expectation than in reach (Figure 2). The basis for this difference was again the different rates of updating in the two output systems. Networks with relatively weak recurrent weights tended to default to the prepotent response, in which case the continual updating of the gaze/expectation system led it to show more of this prepotent response than the reach system. In effect, when recurrence was high enough to keep the right location (B) in mind, then "repeating" it, as the gaze/expectation system did, helped performance; in contrast, when recurrence was low so that the wrong, prepotent location (A) came to mind, then repeating hurt performance. In other words, if infants can hold B in mind to some degree, they can show more sensitivity to this information in their continually updating gaze/expectation systems than in their reaching. However, if infants cannot hold this new location in mind (and in fact switch to representing A based on their latent biases), they will show less sensitivity to the correct location in their continually updating gaze/expectation systems, which are now updating more frequently than the reach system on *incorrect* information. The simulation thus yielded the novel empirical prediction that infants may perseverate more in gaze/expectation than in reaching early in development, a prediction that we are now testing.

Card sorting and Speech interpretation task

Like 3-year-olds, young models (both card sort and SIT models) correctly answer simple questions about the post-switch rules, but fail to sort cards according to these rules in the post-switch trials. These behaviors can be understood by considering the degree of conflict that must be resolved in the two tasks. Sorting involves a high degree of conflict, because both previously and currently relevant sorting dimensions are presented (e.g., an object to be sorted is both a truck and blue). Under these circumstances, a weak representation of the post-switch dimension does not lead to a correct response because this representation cannot compete against a strong representation of the pre-switch dimension. In contrast, answering simple knowledge questions (e.g., “Where do the trucks go in the shape game?”) involves no conflict, because only information about the post-switch dimension is presented. Under these circumstances, a weak representation of the post-switch dimension can support correct performance because there are no other competing representations.

This account led to the prediction that children would no longer show an advantage on knowledge questions relative to sorting if the knowledge questions contained the same amount of conflict as test cards. This prediction was confirmed. When knowledge questions contained information about both the pre- and the post-switch dimensions (e.g., “Where do the blue trucks go in the shape game?”), 3-year-olds’ knowledge of the post-switch rules no longer outstripped their sorting behavior (Munakata & Yerys, 2001). This pattern of findings has also been observed with 6-year-olds in the speech interpretation task (Morton & Munakata, in press-b).

Finally, the model suggests that active representations are required in complex tasks involving conflict, whereas latent representations can suffice in simple tasks that contain little or no conflict. For example, networks needed to maintain a strong active representation of the shape rule to switch from sorting by color to sorting by shape, and to correctly answer conflict knowledge questions. However, weak latent representations sufficed for answering simple questions that contained no conflict. Indeed, networks continued to answer simple knowledge questions correctly even if active memory was eliminated altogether by setting the weight of the recurrent connections to 0. This may imply that with development, representations become stronger both quantitatively and qualitatively. Early in development, performance in certain tasks may be supported almost exclusively by latent representations. These latent representations may become stronger with development, and may additionally benefit from an increasing contribution from active memory. Thus, developmental changes in the strength of representations may comprise both quantitative and qualitative changes.

Decalage

The preceding simulations demonstrate how graded representations might lead infants and children to show dissociations in their behavior, succeeding on one task (e.g., looking to a hidden object, answering a question about a rule) while failing another task meant to measure the same knowledge (e.g., reaching for the hidden object, sorting cards based on the rule). This section explores a related phenomenon: Why do infants and children master formally similar tasks at different ages? This phenomenon of decalage (Piaget, 1941,

1967) poses a challenge to many theories of development. Most theories attribute infants' or children's success in a task to the development of a certain ability or collection of abilities. However, if a putative ability is operative in one task at an early age, why does the same ability not appear to be operative in a formally similar task until later in development?[3]

In this section, we consider two instances of decalage in perseverative tasks: one within variants of the towel-pulling task, the other across the card sort and speech interpretation tasks. Infants show a decalage in variations of the towel-pulling task from 7 to 11 months. Infants at all of these ages pull the correct towel on A trials (i.e., the towel that will yield the toy), but succeed or fail on B trials depending on the task variant. Nine-month-old infants succeed on B trials if they are shown that the toy is *attached* to the towel during presentation, but they perseverate on B trials if the toy is simply placed (unattached) on the towel. Eleven-month-old infants succeed in both versions of the towel pulling task while 7-month-old infants perseverate in both versions. The attached and unattached versions of the task are formally equivalent (choose a towel on the B trials based on the same features used to make the decision on the A trials, rather than perseverating to location), but infants show a decalage in when they succeed on these tasks.

The towel-pulling model simulated this decalage naturally, based on the strength of active representations and their ability to compete with latent representations of previous behaviors. Specifically, the network had stronger representations for a toy *attached* to a towel than for a toy *on* a towel, allowing younger networks to successfully overcome prepotent responses. The network's input units represented the stimuli by encoding the presence of the towel, the toy, whether the toy was on the towel, and whether the toy was attached to the towel (Figure 5). Thus, the *toy*, *towel*, and *on* units were activated on the correct side for both variants of the task, but the *attached* unit provided additional activity on the correct side for the attached condition only. With mid-range recurrent weights (reflecting 9-month-olds) this additional input was sufficient to override the prepotent response supported by the latent representations. However, when the toy merely sat on the towel, the resulting lower level of input activity was not sufficient for the network's active representations to override the prepotent response. At younger ages (lower recurrent levels), the additional input provided by the *attached* unit could not be maintained enough to overcome the latent bias. At older ages (higher recurrent levels), the network could maintain activation for the correct side to overcome the latent bias, regardless of the additional input provided by the attached unit.

According to this account, infants with some active memory abilities may overcome perseverative tendencies, if their developing active memories are bolstered by strong input from the environment supporting correct responding. That is, infants can better maintain a representation of the correct choice if that choice is made more salient. When a toy is attached to a towel, this provides such environmental support for the correct choice. Nine-month-old infants can use this additional information to strengthen their active representation of the correct choice, to succeed with attached toys while perseverating with toys that are simply on towels. Younger infants' active memory abilities are too weak to benefit from the greater input for the correct choice, and older infants' active memory abilities are too strong to see benefits from this greater input.

The card sorting and speech interpretation tasks provide another interesting example of decalage. These tasks are formally equivalent (switch from one dimension to a conflicting dimension for classifying stimuli),

---

[3] Dissociations might be viewed as a broad class of discrepancies in behavior (any case where one measure shows success, and another measure meant to tap the same knowledge shows failure), with instances of decalage as a particular type of dissociation (i.e., behavioral discrepancies across tasks that are formally similar).

but most 4-year-olds pass the card sorting task while most 6-year-olds fail the speech interpretation task. The card sorting and speech interpretation models simulated this decalage naturally, based on the strength of latent representations underlying prepotent responses. Specifically, the network had stronger latent representations to override for the speech interpretation task than for the card sorting task. At the outset of the simulations, feedforward connections were stronger for content than paralanguage in the speech interpretation task, whereas they were equal in strength for shape and color in the card sort model. This manipulation reflected the fact that children come into the speech interpretation task with a strong pre-existing bias to respond to content, a bias that is stronger than any bias that children bring to the card sorting task. As a result of these stronger latent representations, the model required stronger recurrent connections to overcome the prepotent response in the speech interpretation task than in the card sorting task (see also Cohen & Servan-Schreiber, 1992). That is, the model showed a decalage, by succeeding in these formally similar tasks at different ages.

In this way, our active-latent account provides a natural framework for understanding why infants and children show decalages in their mastery of formally similar tasks. Tasks may have the same formal demands, but differ in how much support they provide for active representations of the correct response (as in the *on* and *attached* versions of the towel-pulling task), or in how strong the latent representations are that must be overcome (as in the card sorting and speech interpretation tasks). Infants, children, and networks may thus succeed or fail on formally similar tasks depending on the competitive dynamic between latent representations underlying perseverative responses and active representations supporting currently relevant information.

Discussion

Our active-latent account of perseveration, dissociation, and decalage shares much with existing approaches (e.g., Cohen & Servan-Schreiber, 1992; Dehaene & Changeux, 1991; Goldman-Rakic, 1987b; Thelen et al., 2001). We discuss the relation between our active-latent account and these similar theories elsewhere (Morton & Munakata, in press-a; Munakata, 1998; Munakata, Sahni, & Yerys, 2001; Stedron et al., 2002). Our active-latent account contrasts with several other theories. We discuss three such alternatives here: working memory and inhibition, miscategorization, and reflective consciousness.

Working memory and inhibition

Our account of perseveration and prefrontal cortex has focused on the primary construct of active memory, a component of working memory. In this view, prefrontal cortex does not serve specifically to inhibit prepotent responses; instead inhibition falls out of the more basic mechanism of working memory (Cohen & Servan-Schreiber, 1992; Goldman-Rakic, 1987b; Kimberg & Farah, 1993; Miller & Cohen, 2001; O’Reilly et al., 1999; Roberts et al., 1994). In contrast, other accounts have focused on working memory and inhibition as separate mechanisms subserved by prefrontal cortex (Diamond, 1991, 1998; Fuster, 1989).

The critical difference between the working memory and inhibition accounts is whether a separate mechanism of inhibition needs to be attributed to prefrontal cortex. Our simulations serve as an existence proof that working memory alone may be sufficient, because perseveration is reduced solely by changes to the working memory system. Two additional types of evidence further suggest that inhibition may fall out of the more basic mechanism of working memory, rather than being a separate construct. First, impairing working memory impairs inhibitory abilities. For example, adults have more difficulty inhibiting inappropriate eye movements in the anti-saccade task when simultaneously engaging in a working memory task (Roberts et al., 1994). Such data suggest that the same mechanisms may contribute to working memory and inhibition. Second, the nature of cortical connectivity suggests that inhibition is not a function localized to one region (i.e., prefrontal cortex) that inhibits other regions. Instead, inhibitory interneurons show very diffuse patterns of connectivity within circumscribed regions of cortex, and long-range intracortical connections are excitatory (Shepherd, 1992; White, 1989). This evidence supports the idea that the inhibition of prepotent responses arises from the use of working memory, dependent on prefrontal cortex, to support appropriate behaviors; the inhibition of inappropriate behaviors falls out of competitive interactions dependent on inhibitory interneurons throughout the cortex.

Miscategorization

The miscategorization account proposes that infants (and adults) perseverate because they fail to encode relevant information when the task changes (Aguiar & Baillargeon, 2000; Baillargeon & Wang, 2002). According to this account, they would succeed if they only attended to the relevant information, but their miscategorization of the task as old leads them to ignore the task changes and rely on prior solutions. For example, after repeatedly searching in the correct hiding well or pulling the correct towel, infants miscategorize a new trial (with a new location for the hidden toy or for the correct towel for retrieving a toy) as old; they perseverate to the same location because they fail to notice that the toy has moved. Similarly, after repeatedly sorting cards by color or making judgments about emotion based on content, children might perseverate with these rules because they fail to notice that the experimenter has specified a new set of rules. Thus, the miscategorization account focuses on what infants and children encode, whereas the active memory account focuses on what infants and children can maintain in memory.

Miscategorization may contribute to perseveration in some tasks but does not appear to explain all perseverative behavior. For example, children do much better at switching to a new rule in the card sorting task if negative feedback is provided when they perseverate (Yerys & Munakata, 2001). The negative feedback may change children’s assumption that the task is the same as the previously mastered task (with the old sorting rule), thus enabling them to encode and act on the new rule. However, even with negative feedback, some children still sort perseveratively, suggesting that miscategorization cannot be their only difficulty. In addition, infants who perseverate in the A-not-B task receive similar negative feedback (the well they search in is empty and they do not get to play with the toy), yet they often continue to perseverate in subsequent trials (Butterworth, 1977).

The miscategorization and active memory accounts may lead to different predictions about infants’ reaction times, which would help to address the role of these factors in perseveration (Stedron et al., 2002). As described earlier, the active memory model predicts an inverted U-shaped reaction time curve across development: infants will have the slowest reaction times just before and after they succeed at the task, whereas much younger (perseverating) infants and much older (non-perseverating) infants will be much faster. In contrast, the miscategorization model of this task may not predict a significant change in reaction time for infants who are first succeeding and older infants, because both can encode and remember changes to the task, and so should perform similarly.

Reflective consciousness

Cognitive Complexity and Control (CCC) Theory (Zelazo & Frye, 1997) and the related Levels of Consciousness framework (Zelazo, 2000), emphasize the role of reflection and higher-order representations in the development of executive control. For example, according to the Levels of Consciousness framework, reflective consciousness allows infants to maintain representations of hidden objects in working memory. In the absence of reflection, awareness of an object is confined to the present, and unrecoverable if the object is removed from view. As a result, infants fall prey to prepotent responses when searching for hidden toys. Similarly, according to CCC theory, 3-year-olds in the card sorting task fail to use post-switch rules they evidently know because they are unable to reflect on these rules and subordinate them to a higher-order rule. In sum, age-related advances in reflective consciousness allow increasingly complex representations to govern action. This account has been explored through a neural network model (Marcovitch & Zelazo, 2000) using the cascade correlation learning algorithm (Fahlman & Lebiere, 1990), according to which networks recruit additional units as they are needed to solve tasks.

Accounts that emphasize the role of reflective consciousness contrast with our active-latent account in several ways. For example, the Levels of Consciousness framework argues that infants must be able to reflect on representations to maintain them in working memory. In contrast, our active-latent account focuses on the more basic mechanism of recurrence for understanding infants' abilities to maintain active memories. Further, CCC theory argues that the difficulty of a task is related to its formal complexity: tasks that require the use of embedded rules are more difficult than tasks that involve the use of simple rules. However, as described earlier, children show a decalage in performance across tasks with equivalent formal complexity (card sorting and speech interpretation). It appears that this decalage may not be easily interpretable within the basic CCC framework. In contrast, our active-latent account naturally accounts for this decalage in terms of the strength of latent biases that need to be overcome.

In addition, whereas CCC theory argues that knowledge-action dissociations occur because children are unable to reflect on their knowledge, our active-latent account maintains knowledge-action dissociations are more apparent than real. Knowledge appears to outstrip action only when they are measured under different conditions. When knowledge and action are measured under more equivalent conditions, systematic dissociations disappear (Morton & Munakata, in press-b; Munakata & Yerys, 2001).

Conclusion

Our neural network explorations suggest the following answers to the questions that guided this chapter:

- How does prefrontal development reduce perseveration?

  The development of prefrontal cortex can support the strengthening of active representations – of hidden toys, towels to pull, or rules for sorting cards or judging utterances. Stronger active representations can compete better against latent representations that build over repeated experience and support perseverative behaviors.

- Why are dissociations observed in perseveration?

  The strengthening of active representations is not an all-or-nothing process; these representations are graded in nature. Weak representations may allow infants and children to succeed in some tasks (e.g., those that require less effort or involve little conflict) but not others (e.g., those that require more effort or involve greater conflict).

- *Why* do children show the remarkable decalage in their flexibility, overcoming perseveration at such different ages across various tasks?

  The strength of active and latent representations influences when infants and children can pass a task. One version of a task may support stronger active representations (e.g., for a toy that is attached to a towel in addition to being on it) than those of a formally similar task (e.g., for a toy that is simply on a towel); younger infants may succeed only in the task that supports stronger active representations. In addition, stronger latent representations (e.g., representing a bias to speech content after years of processing speech) require stronger active representations to overcome them than do relatively weak latent representations (e.g., those established during a few trials in a card sorting experiment).

In exploring each of these questions, we have found neural network models to be particularly useful tools, for specifying our theories in working models, testing the abilities of the models to simulate the developmental timecourse of behaviors observed in infants and children, and generating testable empirical predictions that may help to distinguish our theories from alternative theories. For all of these reasons, neural network models provide a useful tool for understanding cognitive development more generally (Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996; Munakata & Stedron, 2001).

References

Aguiar, A., & Baillargeon, R. (2000). Perseveration and problem solving in infancy. In H. Reese (Ed.), *Advances in child development and behavior*, Vol. 27 (pp. 135–180). New York: Academic Press.

Ahmed, A., & Ruffman, T. (1998). 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? *Developmental Psychology*, *34*, 441–453.

Baillargeon, R., & Wang, S.-h. (2002). Event categorization in infancy. *Trends in Cognitive Sciences*, *6*, 85–93.

Bell, M., & Fox, N. A. (1992). The relations between frontal brain electrical activity and cognitive development during infancy. *Child Development*, *63*, 1142–63.

Berthier, N., DeBlois, S., Poirier, C., Novak, M., & Clifton, R. (2000). Where's the ball? Two- and three-year-olds reason about unseen events. *Developmental Psychology*, *36*, 394–401.

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

Clearfield, M. W., & Thelen, E. (2000). Reaching really matters: The development of infants' perseverative reaching. Talk presented at the 2000 meeting of the International Conference on Infant Studies, Brighton, England.

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

De Fockert, J., Rees, G., Frith, C., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science*, *291*, 1803–1806.

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.

Dehaene, S., & Changeux, J. P. (1991). The Wisconsin Card Sorting Test: Theoretical analysis and modeling in a neuronal network. *Cerebral Cortex*, *1*, 62–79.

Diamond, A. (1981). Retrieval of an object from an open box: The development of visual-tactile control of reaching in the first year of life. *Society for Research in Child Development Abstracts*, *3*, 78.

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

Diamond, A. (1991). Neuropsychological insights into the meaning of object concept development. In S. Carey, & R. Gelman (Eds.), *The epigenesis of mind* (Chap. 3, pp. 67–110). Hillsdale, NJ: Lawrence Erlbaum Associates.

Diamond, A. (1998). Understanding the A-not-B error: Working memory vs. reinforced response, or active trace vs. latent trace. *Developmental Science*, *1*, 185–189.

Diamond, A. (in press). A model system for studying the role of dopamine in prefrontal cortex during early development in humans: Early and continuously treated phenylketonuria (PKU). In M. Johnson, Y. Munakata, & R. Gilmore (Eds.), *Brain development and cognition: A reader*. Oxford: Blackwell.

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$\overline{B}$ task: Evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research*, *74*, 24–40.

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

Fahlman, S. E., & Lebiere, C. (1990). The Cascade-Correlation learning architecture. In D. S. Touretzky (Ed.), *Advances in Neural Information Processing Systems 2* (pp. 524–534). San Mateo, CA: Morgan Kaufmann.

Farah, M. J., O'Reilly, R. C., & Vecera, S. P. (1993). Dissociated overt and covert recognition as an emergent property of a lesioned neural network. *Psychological Review*, *100*, 571–588.

Flavell, J. H. (1963). *The developmental psychology of Jean Piaget*. New York: D. Van Norstrand Company Inc.

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

Goldman-Rakic, P. S. (1987a). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In F. Plum, & V. Mountcastle (Eds.), *Handbook of physiology: The nervous system V* (Chap. 9, pp. 373–417). Bethesda, MD: American Physiological Society.

Goldman-Rakic, P. S. (1987b). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. *Handbook of Physiology — The Nervous System*, *5*, 373–417.

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).

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

Hood, B., & Willatts, P. (1986). Reaching in the dark to an object's remembered position: Evidence for object permanence in 5-month-old infants. *British Journal of Developmental Psychology*, *4*, 57–65.

Kimberg, D. Y., & Farah, M. J. (1993). A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex, organized behavior. *Journal of Experimental Psychology: General*, *122*, 411–428.

Lhermitte, F. (1986). Human autonomy and the frontal lobes: Part II. Patient behavior in complex and social situations: The "environmental dependency syndrome". *Annals of Neurology*, *19*, 335–343.

Marcovitch, S., & Zelazo, P. D. (2000). A generative connectionist model of the development of rule use in children. *Proceedings of the Twenty-second Annual Conference of the Cognitive Science Society* (pp. 334–339). Lawrence Erlbaum Associates: Mahwah, NJ.

Mathis, D. A., & Mozer, M. C. (1996). Conscious and unconscious perception: A computational theory. In G. Cottrell (Ed.), *Proceedings of the 18th Annual Conference of the Cognitive Science Society* (pp. 324–328). Mahwah, NJ: Lawrence Erlbaum.

McClelland, J. L., Rumelhart, D. E., & PDP Research Group (Eds.). (1986). *Parallel distributed processing. Volume 2: Psychological and biological models*. Cambridge, MA: MIT Press.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *24*, 167–202.

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

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefontal cortex of the macaque. *Journal of Neuroscience*, *16*, 5154–5167.

Milner, B. (1963). Effects of different brain lesions on card sorting. *Archives of Neurology*, *9*, 90–100.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Press.

Morton, J. B., & Munakata, Y. (in press-a). Active versus latent representations: A neural network model of perseveration and dissociation in early childhood. *Developmental Psychobiology*.

Morton, J. B., & Munakata, Y. (in press-b). Are you listening? Exploring a knowledge action dissociation in a speech interpretation task. *Developmental Science*.

Morton, J. B., & Trehub, S. E. (2001). Children's understanding of emotion in speech. *Child Development*, *72*(3), 834–843.

Morton, J. B., Trehub, S. E., & Zelazo, P. D. (in preparation). Representational inflexibility in 6-year-olds' understanding of emotion in speech.

Munakata, Y. (1998). Infant perseveration and implications for object permanence theories: A PDP model of the $A\overline{B}$ task. *Developmental Science*, *1*, 161–184.

Munakata, Y. (2001a). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences*, *5*(7), 309–315.

Munakata, Y. (2001b). Task-dependency in infant behavior: Toward an understanding of the processes underlying cognitive development. In F. Lacerda, C. v. Hofsten, & M. Heimann (Eds.), *Emerging cognitive abilities in early infancy* (pp. 29–52). Mahwah, NJ: Lawrence Erlbaum Associates.

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*, 686–713.

Munakata, Y., Sahni, S. D., & Yerys, B. E. (2001). An embodied theory in search of a body: Challenges for a dynamic systems model of infant perseveration. *Behavioral and Brain Sciences*, *24*, 56–57.

Munakata, Y., & Stedron, J. M. (2001). Neural network models of cognitive development. In C. Nelson, & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (pp. 159–171). Cambridge, MA: MIT Press.

Munakata, Y., & Stedron, J. M. (in press). Memory for hidden objects in early infancy. In J. Fagen, & H. Hayne (Eds.), *Advances in infancy research, volume 14*. Norwood, NJ: Ablex Publishing Corporation.

Munakata, Y., & Yerys, B. E. (2001). All together now: When dissociations between knowledge and action disappear. *Psychological Science*, *12*(4), 335–337.

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically based computational model of working memory. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control.* (pp. 375–411). New York: Cambridge University Press.

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. (1941). Le méchanisme du développement mental et les lois du groupement des opération. *Arch. Psych., Genève*, *28*, 215–285.

Piaget, J. (1952). *The origins of intelligence in childhood.* New York: International Universities Press.

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

Piaget, J. (1967). *Six psychological studies*. New York: Random House.

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.

Roberts, R. J., & Pennington, B. F. (1996). An interactive framework for examining prefrontal cognitive processes. *Developmental Neuropsychology*, *12*(1), 105–126.

Shepherd, G. M. (1992). *Foundations of the neuron doctrine.* New York: Oxford University Press.

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

Stedron, J., Munakata, Y., & Sahni, S. D. (2002). In plain sight: Simulating the role of memory in perseveration with visible solutions. Manuscript submitted for publication.

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.

Stuss, D., & Benson, D. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin*, *95*, 3–28.

Thelen, E., Schoner, G., Scheier, C., & Smith, L. B. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, *24*, 1–86.

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).

White, H. (1989). Learning in artificial neural networks: A statistical perspective. *Neural Computation*, *1*, 425–464.

Yerys, B. E., & Munakata, Y. (2001). Feedback improves children’s flexibility: Rethinking perseveration in a card sorting task. Manuscript in preparation.

Zelazo, P. D. (2000). Self-reflection and the development of consciously controlled processing. In P. M. . K. Riggs (Ed.), *Children’s reasoning and the mind* (pp. 169–189). London: Psychology Press.

Zelazo, P. D., & Frye, D. (1997). Cognitive complexity and control: A theory of the development of deliberate reasoning and intentional action. In M. Stamenov (Ed.), *Language structure, discourse, and the access to consciousness* (pp. 113–153). Amsterdam: John Benjamins.

Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37–63.