

J. Bruce Morton
Yuko Munakata

Department of Psychology
University of Denver, Denver
CO 80208

Active Versus Latent Representations: A Neural Network Model of Perseveration, Dissociation, and Decalage

Received 21 February 2001; Accepted 7 November 2001

ABSTRACT: Children of different ages often perseverate, repeating previous behaviors when they are no longer appropriate, despite appearing to know what they should be doing. Using neural network models, we explore an account of these phenomena based on a distinction between active memory (subserved by the prefrontal cortex) and latent memory (subserved by posterior cortex). The models demonstrate how (a) perseveration occurs when an active memory of currently relevant knowledge is insufficiently strong to overcome a latent bias established by previous experience, (b) apparent dissociations between children's knowledge and action may reflect differences in the amount of conflict between active and latent memories that children need to resolve in the tasks, and (c) differences in when children master formally similar tasks (decalage) may result from differences in the strength of children's initial biases. The models help to clarify how prefrontal development may lead to advances in flexible thinking. © 2002 Wiley Periodicals, Inc. *Dev Psychobiol* 40: 255–265, 2002. DOI 10.1002/dev.10033

Keywords: memory; perseveration; dissociation; decalage; prefrontal cortex development; neural network models

INTRODUCTION

The ability to act flexibly in light of changing circumstances, rather than by simple habits, is an important foundation of higher intelligence. For example, under most circumstances, we use a house key to unlock the front door of our house. However, we might use the

same key to cut a piece of string if we cannot find a knife or a pair of scissors. Although considerable research suggests that the prefrontal cortex plays an important role in this sort of flexible behavior, its precise role remains unclear. In this paper, we use neural network models to explore the role of prefrontal cortex in the development of flexible behavior in childhood.

Children often behave inflexibly in new circumstances by repeating habitual behaviors that are no longer appropriate. This kind of behavior, called perseveration, can be reliably observed in a number of simple tasks. For example, many 3-year-olds persist in sorting cards by color when asked to switch and sort the cards by shape (Zelazo, Frye, & Rapus, 1996).

Correspondence to: J. Bruce Morton or Y. Munakata
Contract grant sponsor: NICHD: 1R29HD37163-01
Contract grant number: NSF: IBN-9873492

Similarly, given conflicting cues to emotion in speech, many 6-year-olds persist in judging a speaker's feelings from what she says despite clear instructions to attend to her manner of expression or paralanguage (Morton & Munakata, in press; Morton, Trehub, & Zelazo, 2002).

In both tasks, children initially behave appropriately, sorting cards or judging the speaker's feelings according to the experimenter's instructions. However, when they are required to switch and view the same stimuli in a new way, children characteristically behave inflexibly; a majority continue to act according to their original responses rather than the experimenter's new instructions.

Perhaps most strikingly, even as children persist in using the old rules, they seem to show some knowledge of the new rules. Thus, their actions appear to be dissociated from their knowledge. For example, most 3-year-olds who persevere on initial color rules can point to where the cards should be placed according to the shape rules (Munakata & Yerys, 2001; Zelazo et al., 1996). Similarly, most 6-year-olds who inappropriately persist in judging a speaker's feelings from what she says accurately label instances of happy and sad paralanguage when content is obscured, and indicate that they intend to listen to the speaker's voice and not to what she says (Morton & Munakata, in press; Morton, Trehub & Zelazo, 2002). Such knowledge-action dissociations place important constraints on theories of perseveration, and can potentially shed light on the organization and development of the adult mind.

The prefrontal cortex is thought to play an important role in flexible behavior. For example, individuals with damage to their frontal lobes often behave perseveratively, continuing to sort cards according to an old rule despite negative feedback (Milner, 1963), or inappropriately performing familiar actions with everyday objects (e.g., attempting to use cutlery despite feeling sated and despite the absence of food; Lhermitte, 1986). Adult nonhuman primates with frontal lobe lesions also exhibit perseverative behavior by searching for rewards at previously correct rather than most-recent hiding locations (Diamond & Goldman-Rakic, 1989). Moreover, there is evidence that the *development* of the prefrontal cortex continues into adolescence and contributes to age-related advances in flexible behavior throughout this period of life (Bell & Fox, 1992; Casey, Durston, & Fossella, 2001; Diamond, 2000; Zald & Iacono, 1998). Thus, as the frontal lobes develop, children become better able to act appropriately in light of changing contextual demands and avoid perseverating on habitual responses.

Still, much remains unclear about this putative brain-behavior relation. Specifically, how does the development of the prefrontal cortex help children to avoid behaving perseveratively? What accounts for apparent dissociations between what children know and how they behave? Finally, why are children able to switch appropriately in a card-sorting task long before they are able to switch appropriately in a structurally similar speech-interpretation task? This phenomenon, termed *decalage* by Piaget (1941), represents a unique theoretical challenge, for any mechanism that is proposed to account for age-related advances in flexible behavior also needs to account for the fact that children continue to persevere on one task long after succeeding on a similar task.

A NEURAL NETWORK APPROACH

In this article, we use neural network models to explore how the development of the prefrontal cortex contributes to age-related advances in flexible behavior. Like traditional production system approaches (Anderson, 1983; Newell, 1990), neural network models allow for explicit computer simulations of task performance. Neural network models also incorporate important principles of neural computation (McClelland, Rumelhart & PDP Research Group, 1986; O'Reilly & Munakata, 2000; Rumelhart & McClelland, 1986). Typically, models consist of individual processing units grouped into separate layers and connected to form pathways. Like neurons, units become active as a nonlinear function of incoming excitatory and inhibitory signals from other units. A "presynaptic" unit affects the activity level of a "postsynaptic" unit based on the strength of the connection between the units and the activity level of the presynaptic unit. Learning occurs as the strength of connections between units is modified by experience. Networks represent information as patterns of activation across units, and process information by propagating this activation among units. As such, neural network models represent an important tool for understanding brain-behavior relationships and have been used to understand the mechanisms of typical cognition (McClelland, Rumelhart & PDP Research Group, 1986; O'Reilly & Munakata, 2000; Rumelhart & McClelland, 1986), the cognitive sequelae of different brain injuries (Farah, O'Reilly, & Vecera, 1993; Munakata, 2001), cognitive impairments associated with various disorders including schizophrenia (Cohen & Servan-Schreiber, 1992), Parkinson's disease (Amos, 2000; Monchi, Taylor, & Dagher, 2000), and Huntington's disease (Amos, 2000) as well as

cognitive changes that occur with normal brain development (Elman et al., 1996; Munakata & Stedron, 2001).

In this article, we explore the relation between the development of prefrontal cortex and perseveration, dissociation, and decalage by exploring a distinction between active and latent memory traces (Munakata, 1998). In neural network models, active representations take the form of sustained activity in particular processing units, and latent traces take the form of changes in the strength of the connections between processing units. According to the active-latent account:

(a) Perseveration occurs when an active trace for currently relevant information is insufficiently strong to compete against a latent trace for previously relevant information; (b) latent memory traces are formed in posterior cortex when processing a stimulus brings about a change in the subsequent processing of the stimulus. Because these traces are instantiated as changes in the strength of connections between units, they are not accessible to other brain regions, but influence processing in these regions through changes in the activation to subsequent stimuli; (c) active memory traces are formed in the prefrontal cortex when organisms actively maintain a representation of a stimulus. Active representations may be accessible to other brain regions, and can influence processing in these areas even if the stimulus is no longer perceptually available; and (d) flexible behavior can be understood in terms of the relative strengths of latent and active memory traces. With the development of the prefrontal cortex comes an improvement in the ability to actively maintain current rules, which allows for greater flexibility in tasks that involve card-sorting and speech interpretation.

The active-latent distinction that is central to our model is based on evidence from neurophysiological and behavioral studies which shows that neurons can remember information both in an active manner through sustained firing for a stimulus and in a latent manner through changes in firing thresholds or synaptic strength. For example, cells in the prefrontal cortex of monkeys show sustained firing to specific stimuli that must be remembered across delays (Fuster, 1973; Goldman-Rakic, 1987). This active memory is consistent with a number of imaging studies of the prefrontal cortex (e.g., Cohen et al., 1997). This type of active memory in the prefrontal cortex appears to be critical for maintaining information about a specific stimulus across intervening stimuli (Miller, Erickson, & Desimone, 1996). In contrast, simple recognition of familiar stimuli has been associated with decreased firing in cells in the inferotemporal cortex (Miller &

Desimone, 1994). One interpretation is that repeated exposure to a stimulus leads to the formation of a latent memory through changes in the strength of connections between neurons that represent the stimulus. These changes facilitate the processing of the stimulus on subsequent presentations, leading to reduced firing. A similar active-latent distinction for memories of spatial location has been observed in prefrontal and parietal cortex (Steinmetz, Connor, Constantinidis, & McLaughlin, 1994). It also is borne out in studies of humans with frontal lobe damage who show impaired performance on tasks that require the active maintenance and manipulation of information (e.g., working memory tasks; Kessels, Postma, Wijnalda, & de Haan, 2000), but normal performance on less strategic memory tasks such as priming (Shimamura, Gershberg, Jurica, Mangels, & Knight 1992).

The active-latent account shares important similarities with other models of perseveration and prefrontal cortex function. For example, several models view perseveration as arising out of a competition between different types of information that are similar to active and latent memory traces (Butterworth, 1977; Dehaene & Changeux, 1989; Diamond, 1985). As well, other models maintain that a basic function of the prefrontal cortex is the active maintenance of task-relevant information (Cohen & Servan-Schreiber, 1992; Goldman-Rakic, 1987; O'Reilly, Braver, & Cohen, 1999; Roberts & Pennington, 1996), with other functions, such as the inhibition of prepotent responses, dependent on this basic function. In this view, the role of the prefrontal cortex is to actively maintain task-relevant information (e.g., a rule, a recent hiding location, etc.) over delay periods or in the face of competing stimuli and to help sustain activity in the relevant regions of posterior cortex. Through local inhibitory connections, this sustained activity leads to the inhibition of inappropriate representations and responses.

Our model shares important features with and builds on an existing neural network model of the Stroop task (Cohen, Dunbar, & McClelland, 1990). In the task (for a review, see MacLeod, 1991), participants are presented with words printed in different ink colors (e.g., the word "RED" written in blue ink) and are required to name the ink color. Compared to a neutral condition in which the meaning of the word has very weak semantic associations with color (e.g., the nonsense word "SPUG"), participants are slower to name the color of the ink when the meaning of the word conflicts with the color to be named (e.g., the word "RED" written in blue ink), but are faster when the word and the ink color are consistent (e.g., the word "BLUE" written in blue

ink). Thus, strong, highly practiced responses (e.g., word-reading) interfere with weaker, less frequently practiced responses (e.g., color-naming), creating a need for a higher order control of processing. The neural network model explored how context representations (i.e., representations of task instructions) help to achieve this control (Cohen et al., 1990). In the model, separate pathways processed the ink color and the word meaning of various stimuli. Connections between processing units in the word-reading pathway were stronger than connections in the color-naming pathway, which reflected the fact that word-reading is more practiced than color-naming. Context units which represented task instructions (i.e., instructions either to color name or word read) were connected to and could modulate the strength of the color-naming and word-reading pathways. Because the word-reading pathway was considerably stronger than the color-naming pathway, the context units were instrumental in allowing the network to respond appropriately when instructed to color name. The model accounted for an impressive corpus of data (for a review, see MacLeod, 1991), including classic interference (Stroop, 1935) and practice effects (Dunbar & MacLeod, 1988). Moreover, it provided a basis for simulating cognitive impairments associated with frontal lobe dysfunction in schizophrenia. Individuals with schizophrenia are more prone to distraction than normally functioning individuals, possibly because frontal lobe dysfunction compromises their ability to hold contextual representations in mind. Consistent with this prediction, reducing the sensitivity of context units to input made them less able to bias activity in the lower order pathways, and resulted in increased interference effects—a result that precisely paralleled behavioral data (Cohen & Serven-Schreiber, 1992).

The active-latent model shares important similarities with the Stroop model, including the basic architectural design and the resultant competitive dynamics between active (or contextual) representations and latent (or automatic) biases. The active-latent model, however, extends these principles and shows how they can provide an account of perseveration, dissociation, and decalage. Moreover, the active-latent model is developmental in that it shows how age-related changes in the efficacy of active memory can account for changes in the incidence of perseveration and dissociation in early childhood.

METHOD

The following simulations explored 3- and 6-year-olds' perseveration in card-sorting (Zelazo et al.,

1996) and speech interpretation (Morton, Trehub, & Zelazo, 2002) tasks respectively. In both tasks, children use a pair of rules successfully in initial trials (referred to as preswitch trials), but then perseverate on these rules in later trials (referred to as postswitch trials) when instructed to switch to new rules. According to the model, children need to maintain a strong active representation of the new rules to overcome latent biases that are established or strengthened by successful use of the first rules. Failure to do so results in perseveration.

Architecture, Environments, and Learning Algorithm

Figures 1 and 2 show the architectures of the two models. Because these architectures were virtually identical and differed only in the strength of initial biases and the labeling of units, the underlying causes of perseveration and dissociation in both models were identical. Therefore, for simplicity, we focus on the card-sort model when discussing perseveration and dissociation and compare the models when discussing decalage.

The model consisted of three input layers (Visual Features, Verbal Features, and Rule), an internal representations layer, a prefrontal cortex (PFC) layer, and an output layer (see Figure 1). The Visual Features layer encoded the color (red and blue) and the shape (truck and flower) of the test cards. The Verbal Features layer encoded a variety of verbal statements that are an intrinsic part of the task, including descriptions of the rules (e.g., "Red ones go here, blue ones go there"), descriptions of the test cards (e.g., "Here's a red one") and knowledge questions (e.g., "Where do the red ones go?"). The Rule layer encoded the rule itself (e.g., "We're playing the color game").¹ The output units represented the trays into which children sort the test cards. In the task, sorting trays are marked by target cards that match each test card on only a single dimension. Thus, if the test cards feature red trucks and blue flowers, the target cards feature blue trucks and red flowers. To parallel this aspect of the task, the two output units were labeled red/flower and blue/truck, respectively.

Units interacted through feedforward, lateral, and recurrent connections. Feedforward connectivity included an initial bias to respond correctly to color and shape information and to accurately encode the current sorting rule. Thus, activation of the "flower" unit in either the Visual Features or the verbal

¹In the model, input units represent a fairly late stage in information processing.

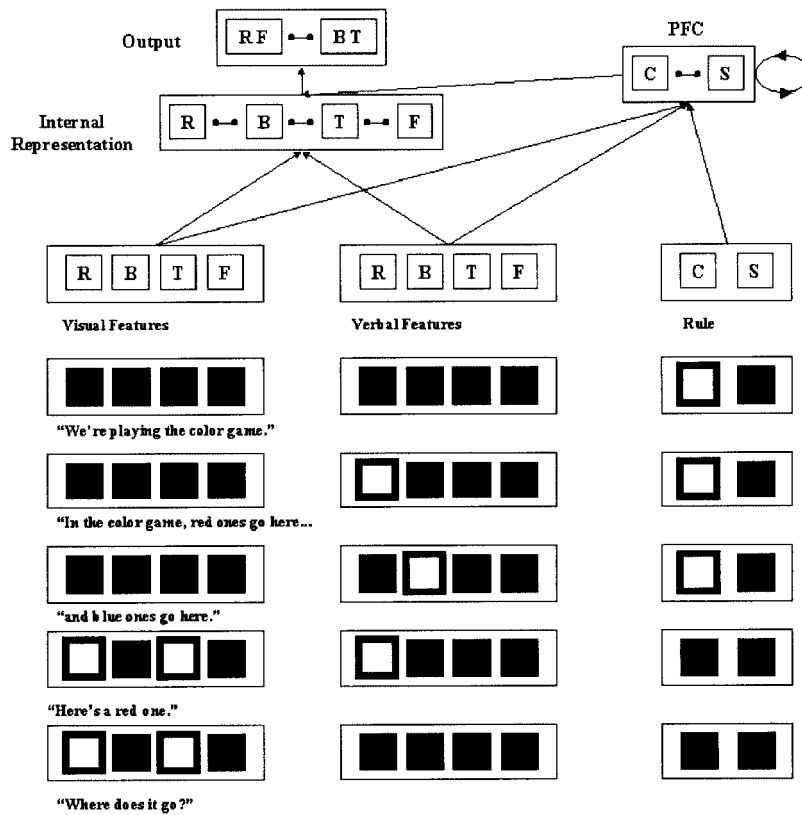


FIGURE 1 A simplified version of the card-sort network (R = red, B = blue, T = truck, F = flower), with the five inputs for a trial shown in the five rows below. Each trial included a presentation of the rule (first row), a description of the rule (second and third rows), a description of a test card (fourth row), and a presentation of a test card (fifth row). The description of the rules also included activation of the relevant output units (not shown).

descriptions layer led to activation exclusively in the “flower” hidden unit, which in turn, led to exclusive activation of the “red/flower” output unit. These biases reflected 3-year-olds’ ability to accurately label familiar shapes and basic colors. Similarly, activation of the “color” and “shape” units in the rule input layer led to activation of the “color” and “shape” PFC units, respectively. Input units that encoded shape (i.e., truck and flower) and color (i.e., red and blue) information also had feedforward connections to the color and shape units of the PFC layer, respectively, reflecting the fact that the presence of these features activates the associated dimension. Finally, the color and shape PFC units were connected to color (i.e., red and blue) and shape (i.e., truck and flower) hidden units. Thus, activation of the color PFC unit increased activation in the red and blue but not the truck and flower hidden units.

Feedforward connections between input and hidden units and between hidden and output units changed with experience according to a Hebbian learning rule such that connections between units that

were simultaneously active increased in strength. When the network sorted by color, for example, units coding for color showed simultaneous activity. Consequently, connections between these units became stronger. In this way, the network developed a latent memory for the color rule, which took the form of an experience-dependent bias in connection weights that favored the representation of color.

Feedforward connections to the PFC units also changed with experience, but at a much slower rate (approximately 1% of the learning rate for other feedforward connections), such that the PFC part of the model was less susceptible to bias than the rest of the system. This manipulation was meant to reflect the unique specialization of the PFC in helping other brain regions to overcome bias (Roberts & Pennington, 1996; Stuss & Benson, 1984).²

²The relatively slow learning rate for the PFC part of the model reflects one hypothesis regarding why PFC is less susceptible to bias. However, with such a low learning rate, the PFC might fail to learn from experience in ways that humans can. Understanding how PFC can both protect itself from bias and learn from experience remains an important challenge for modeling research.

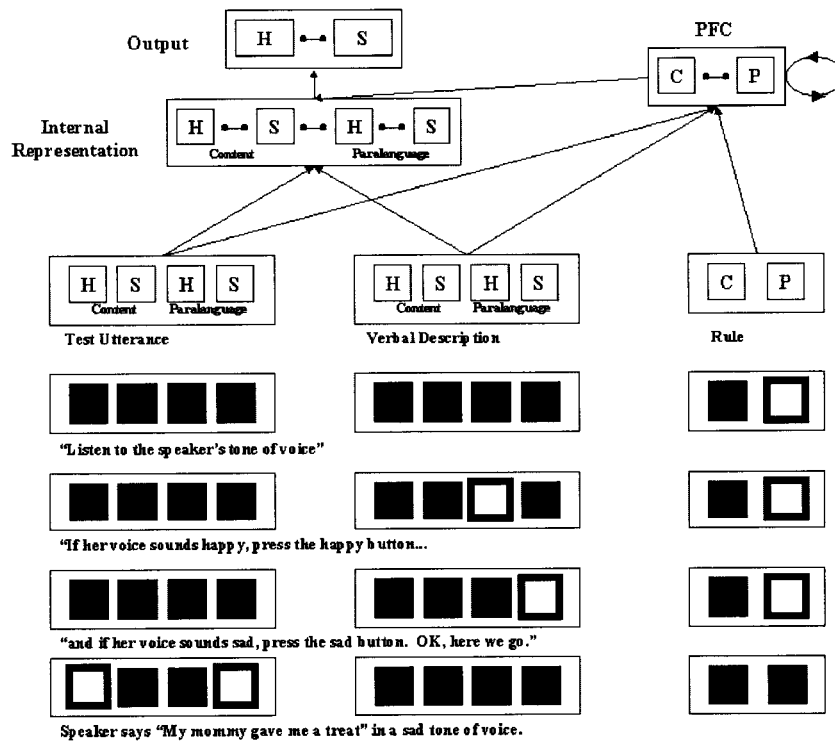


FIGURE 2 A simplified version of the speech-interpretation network (H=happy, S=sad, C=content, P=paralanguage), with inputs for part of the task shown the four rows below. This part of the task included a rule to respond to the speaker’s paralanguage (first row), a description of the rule (second and third rows), and a presentation of a test utterance (fourth row). During the description of the rule, relevant output units also were activated (not shown).

There were lateral inhibitory connections among units within the PFC, hidden, and output layers. As a result, units within these layers had to compete with each other to remain active. The color and shape units in the PFC layer inhibited each other so that activity in the color unit led to a suppression of activity in the shape unit, and vice versa. Similarly, in the output layer, activation of the red/truck unit led to a suppression of activity in the blue/flower output unit (and vice versa). Finally, in the hidden layer, category exemplars inhibited each other (e.g., red inhibited blue, and blue inhibited red), as did the two categories (e.g., each color unit inhibited both shape units, and each shape unit inhibited both color units). The competitive dynamic among units within layers led to a competition for representation, which is considered an important principle of neural computation (McClelland, 1993).

Finally, each PFC unit had a recurrent excitatory connection to itself, which allowed the units to remain active in the absence of external input. Thus, the network’s active memory of a sorting rule took the form of self-sustained activity in the PFC units. Age-related improvements in active memory were simulated as increases in the strength of the recurrent connection.

The simulated task consisted of a demonstration in which two test cards were sorted by the preswitch rule (in this case, color), followed by six preswitch and six postswitch trials. Each pre- and postswitch trial consisted of a statement of the color and shape rules, respectively, followed by a presentation and description of the test card.

RESULTS

Performance and Internal Representations

We measured the network’s performance as a percent correct response, calculated as the activation of the correct output unit divided by the sum of activation across both output units. For example, when a red truck was presented on a color-game trial, percent correct response was calculated as the output activity of RF/RF+BT. As shown in Figure 3, the network simulated good preswitch performance across all ages, perseveration in postswitch trials, and age-related improvements.

Networks of all ages sorted correctly in preswitch trials. Together, the demonstration trials and the labeling of the color of each test card led the network to

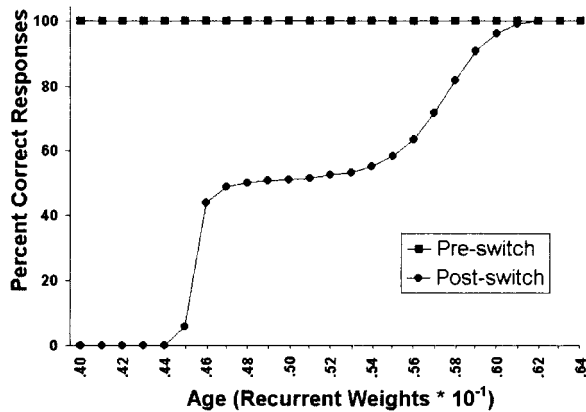


FIGURE 3 Percent correct responses as a function of age. The Network performed perfectly on preswitch trials, but perseverated on postswitch trials when recurrence was low.

form color-biased internal representations of the cards (e.g., presenting a blue flower led to activation of the blue but not the flower hidden unit). These biased internal representations, in turn, caused the network to respond to the color rather than the shape of the test cards, even in the absence of an active representation of the color rule. As the network gained experience sorting test cards by color, feedforward connections between units that processed color became stronger, and a strong latent memory for the preswitch rule was established.

Postswitch performance was directly related to the network's age (i.e., the strength of the recurrent connection). Higher recurrence led to greater sustained activation in the PFC unit representing the shape rule, and this had important consequences for performance. Young networks with low recurrence were unable to maintain a strong active representation of the new shape rule, and were therefore unable to overcome the strong latent bias to represent the color rather than the shape of the test cards. Consequently, they perseverated by sorting the test cards according to their color rather than their shape. In contrast, older networks with stronger recurrence were able to actively maintain a representation of the new shape rule that was sufficiently strong to overcome the latent bias to color. Consequently, they switched appropriately in the postswitch trials. In sum, the model illustrates how perseveration can be understood in terms of a competition between active and latent memory traces.

Predictions

According to our active-latent model, perseveration occurs when active memory is unable to maintain a representation of the current sorting rule. Thus, the model predicts that children who perseverate with old

rules might switch to the new rules if the use of the new rules was scaffolded in a more implicit way rather than verbally instructed in an explicit way. The model makes this prediction because, even in young networks, scaffolding the use of the new rules would lead to changes in connection weights that favor the new rules whereas verbal instructions require the ability to actively maintain the new rules, which young networks cannot do very well. We explored this prediction by providing a young network (i.e., a network with no recurrence) with a series of trials that scaffolded the use of the new rules (i.e., rules based on the previously ignored features of the test cards). First, we trained the network to sort conflict cards (i.e., red trucks and blue flowers) by color. Then, we had the network sort eight "scaffold" cards that contained only shape information (i.e., truck and flower silhouettes). Finally, we tested the effect of the scaffold trials by presenting the network with two additional conflict cards (i.e., one red truck and one blue flower). In the absence of competing color information, the network correctly sorted the eight scaffold cards by shape. This experience caused a shift in how the network sorted the conflict cards. Prior to the scaffold trials, the network sorted conflict cards by color; however, after the scaffold trials, the network sorted the same cards by shape. This shift in sorting behavior was not attributable to active memory because recurrence was set to zero, and the network was not given any explicit instructions that it needed to maintain. Instead, the performance shift was due to a "bottom-up" strengthening of connections between shape-processing units. We are currently testing these predictions in a behavioral study involving 3-year-olds.

Dissociations

Our simulations illustrated the role of active memory in flexible behavior by showing how an improvement in the ability to hold new rules in mind reduces the tendency to perseverate on old rules. However, as discussed earlier, children typically perseverate while also showing an apparent knowledge of the new rules. That is, they sort cards according to the preswitch rules while correctly answering simple questions about the postswitch rules (Zelazo et al., 1996). For example, 3-year-olds who perseverate on color rules can correctly point to where trucks and flowers go in the shape game. Similarly, 6-year-olds who perseverate on content accurately label instances of happy and sad paralinguage, and indicate they intend to listen to the speaker's voice and not to what she is saying. How then is perseveration a memory problem?

Challenges of this sort assume that memory is all or nothing—either completely present or completely absent. Thus, correct responses to simple questions about the postswitch rules are considered evidence that children remember the new rules and persevere for other reasons. However, mental capacities such as memory may not be all or nothing. Instead, memory may vary continuously in strength, due to variation in a number of physiological factors including the number of neurons involved in a particular representation, and the frequency and coherence of their firing rates. By adopting a graded view of memory, one can account for knowledge–action dissociations in terms of memory limitations (for a recent review, see Munakata, 2001). In some instances, dissociations may arise when different tasks (or measures) are used to assess the same underlying representation. When this representation is weak, it can lead to success in the less demanding task but failure in the more demanding task. In the card-sorting task, for example, children’s ability to act on postswitch rules is measured under conditions of conflict because both the currently and previously relevant dimensions are present on each test card. Resolving conflicts of this nature requires a strong active representation of the postswitch rule to overcome the latent bias to represent the previously relevant feature of the test card. In contrast, standard knowledge questions contain no conflict because they only contain information about the currently relevant dimension (e.g., “Where do the trucks go in the shape game?”). In the absence of conflict, no active memory of the postswitch rule is necessary, and latent representations suffice. Thus, dissociations can arise when knowledge and action measures contain different amounts of conflict.

This account led to the prediction that children would no longer show an advantage on knowledge questions if the questions contained the same degree of conflict as the test cards. We tested this idea by presenting two different knowledge questions to networks of different ages. The first was a standard knowledge question with no conflicting cues, simulated by activating the appropriate unit in the verbal descriptions layer. For example, the question, “In the shape game, where do the trucks go?” was simulated by activating the shape unit in the rule input layer and the truck unit in the verbal descriptions layer. The second was a conflict knowledge question that, like the test cards, contained information about both the currently and the previously relevant dimensions (e.g., “In the shape game, where do the blue trucks go?”). We simulated this question by activating the shape unit in the rule input layer as well as the truck and the blue units in the verbal descriptions layer.

Networks of all ages passed the standard knowledge question, but only older networks with strong recurrence passed the conflict knowledge question. Younger networks with weaker recurrence answered the conflict question in terms of the preswitch rules (i.e., according to color). These predictions have been confirmed in behavioral studies of 3- and 6-year-olds (Morton & Munakata, *in press*; Munakata & Yerys, 2001). In one study, 3-year-olds were administered the standard card-sorting task, and their knowledge was tested using both a standard and a conflict knowledge question. Children who perseverated typically answered the standard knowledge question correctly, but the conflict knowledge question incorrectly (Munakata & Yerys, 2001); 6-year-olds who perseverated in the speech-interpretation task showed the same pattern of performance, answering standard knowledge questions correctly, but conflict knowledge questions incorrectly (Morton & Munakata, *in press*). Thus, knowledge and action only appear dissociated when their respective measures contain different amounts of conflict. Taken together, the simulation and behavioral findings suggest that knowledge–action dissociations are more apparent than real.

Decalage

In the preceding sections, we have shown how two phenomena—perseveration and dissociation—may be related to memory problems. Perseveration occurs when active memory is insufficiently strong to overcome latent biases, and dissociation occurs when different tasks (i.e., tests of knowledge and action) involve different amounts of conflict, and therefore depend differentially on active and latent representations. A somewhat-related phenomenon is decalage (Piaget, 1941, 1967), which refers to instances in which formally similar tasks are mastered at different ages. Differences in performance across similar conservation tasks provide a classic illustration of this phenomenon. Children are shown two identical balls of clay and then watch as one gets rolled into a sausage-like strip. In a conservation of mass task, children are asked which object contains more clay; in a conservation of weight task, they are asked which object weighs more. Despite their similarities, conservation of mass tasks are solved approximately a year earlier than conservation of weight tasks. Piaget referred to this phenomenon as horizontal decalage, defined as a temporal uncoupling of similar cognitive developments (Flavell, 1963).

Horizontal decalage poses a challenge for stage theories of development. Piaget (1967) argued, for

example, that to master problems of conservation, children need to understand the principles of compensation, identity, and reversibility. Thus, to master a conservation of mass problem, children need to understand that increases in the length of the strip compensate decreases in its width (the principle of compensation), no amount of clay has been added to or taken away from the ball as it was transformed into the strip (principle of identity), and the strip could be rolled back up to recreate the original ball (principle of reversibility). What is unclear from this analysis, however, is why young children who understand the principles of compensation, identity, and reversibility in the context of a problem concerning the conservation of mass fail to apply these same principles to a very similar problem concerning the conservation of weight. If a putative ability is operative in one task at an early age, why does same ability not appear to be operative in a formally similar task until later in development?

The card-sorting and speech interpretation tasks provide another example of decalage. Both tasks require a switch from one pair of rules to another. However, while most 4-year-olds pass the card-sorting task, most 6-year-olds fail the speech-interpretation task. How might neural network models inform thinking about this instance of decalage? One approach is to consider differences in preexisting biases that are operative in different tasks, and characterize these differences in the models as differences in initial-connection strengths (Cohen & Servan-Schreiber, 1992). Tasks that involve strong preexisting biases may place greater demands on active memory mechanisms than tasks that involve weak preexisting biases. In the card-sorting task, for example, children typically do not show an initial bias for color or shape. Instead, bias (for color or shape) is established during the preswitch trials. In contrast, children come into the speech-interpretation task with a strong prepotent bias for content (Friend, 2000; Morton & Trehub, 2001), which is then further strengthened in preswitch trials. Because of these differences in initial-bias strength, the speech-interpretation task may place more demands on active memory mechanisms (i.e., require higher recurrence in the model) than the card-sorting task.

We explored this hypothesis in a series of simulations involving four models: the card-sorting model, which had no initial bias, and three speech interpretation models with weak, moderate, and strong preexisting biases to content. Of interest was whether demands on active memory would vary under these different conditions. This proved to be true. Models with stronger preexisting biases required stronger

recurrent connections to achieve perfect postswitch performance. Specifically, the card-sort model achieved perfect postswitch performance with a recurrent connection strength of 6.15×10^{-2} whereas speech interpretation models with weak, moderate, and strong preexisting biases to content achieved perfect postswitch performance with recurrent connection strengths of 6.64×10^{-2} , 6.82×10^{-2} , and 8.0×10^{-2} , respectively. Thus, stronger preexisting biases placed greater demands on active memory mechanisms.

CONCLUSIONS

We have presented a model of perseveration, disassociation, and decalage that is based on a distinction between active and latent memory and which finds broad support in behavioral and neural network studies. Our model contrasts in important ways with other accounts of these phenomena. For example, in one view, children perseverate on old color rules despite appearing to know new shape rules because they are unable to reflect on their knowledge (Zelazo & Frye, 1997). Reflection allows children to subordinate lower order perspectives (e.g., views based on shape and color) to a higher order representation and leads to greater cognitive coordination and control. What is unclear from this account is why reflection should emerge earlier in card-sorting than in speech-interpretation tasks. In another view, perseveration reflects the failure of working memory and inhibitory control (Gerstadt, Hong, & Diamond, 1994). We suggest a more parsimonious account in which advances in inhibitory control are inextricably linked to advances in active memory.

Finally, our account departs from another neural network model (Dehaene & Changeux, 1991) that explored perseveration by frontal lobe patients in the Wisconsin Card Sorting Task (WCST; Grant & Berg, 1948). In the task, participants sort cards that vary along several dimensions (e.g., shape, number, color), and use feedback to determine what sorting rule to apply. When feedback contingencies change, participants need to switch to a new sorting rule. Frontal lobe patients often perseverate under these circumstances by continuing to sort by the old rule (Milner, 1963). According to the model, perseveration occurs when feedback is ignored, and units that maintain the sorting rules receive no signal to change to a new rule. This is not a plausible model of children's perseveration in the card-sorting and speech-interpretation tasks, however, because in both tasks, children are explicitly instructed to switch rules and are not given performance feedback. It seems

unlikely, therefore, that age-related declines in the incidence of perseveration in these tasks is attributable to changes in children's sensitivity to feedback.

Still, challenges for our active-latent model remain. With these basic tasks, our models (and children) simply do what they are told, as long as they can actively maintain this information. However, in more complex situations, people must remember what to do and when, and be able to update these plans without experimenter instructions. To simulate such performance, our models would need to incorporate mechanisms that go beyond environmentally specified behaviors. Characterizing these mechanisms without invoking something like a homunculus remains an important challenge for the field (for a review, see Monsell & Driver, 2000). One recent proposal is a conflict-detection mechanism (Botvinick, Braver, Barch, Carter, & Cohen, 2001) that becomes active in the context of competing responses and modulates the contribution of controlled processes to ongoing cognitive activity. Other proposals include gating mechanisms that regulate the updating of active memory in anticipation of future reward (Braver & Cohen, 2000). Whether and how these mechanisms might contribute to the development of cognitive control will need to be informed by further empirical and computational research.

NOTES

We thank Christina Von Stroh for helping to develop one of these models as part of her undergraduate honors thesis, and members of the Cognitive Development Center for constructive suggestions on a draft of this article.

REFERENCES

- Amos, A. (2000). A computational model of information processing in the frontal cortex and basal ganglia. *Journal of Cognitive Neuroscience*, 12, 505–519.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Bell, M., & Fox, N. A. (1992). The relations between frontal brain electrical activity and cognitive development during infancy. *Child Development*, 63, 1142–1163.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108, 624–652.
- Braver, T. S., & Cohen, J. D. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. In S. Monsell & J. Driver (Eds.), *Attention and performance XVII*. Cambridge, MA: MIT, 2000.
- Butterworth, G. (1977). Object disappearance and error in Piaget's stage IV task. *Journal of Experimental Child Psychology*, 23, 391–401.
- Casey, B. J., Durston, S., & Fossella, J. A. (2001). Evidence for a mechanistic model of cognitive control. *Clinical Neuroscience Research*, 1, 267–282.
- Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, 97, 332–361.
- 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.
- Dehaene, S., & Changeux, J.P. (1989). A simple model of prefrontal cortex function in delayed-response tasks. *Journal of Cognitive Neuroscience*, 1, 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. (1985). Development of the ability to use recall to guide action, as indicated by infants' performance on AB. *Child Development*, 56, 868–883.
- Diamond, A. (2000). Close interrelation of motor development and cognitive development and of the cerebellum and prefrontal cortex. *Child Development*, 71, 44–56.
- Diamond, A., & Goldman-Rakic, P. S. (1989). Comparison of human infants and rhesus monkeys on Piaget's AB task: Evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research*, 74, 24–40.
- Dunbar, K., & MacLeod, C. M. (1988). Training and Strooplike interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 14, 126–135.
- 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.
- 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: Van Nostrand.
- Friend, M. (2000). Developmental changes in sensitivity to vocal paralinguistic. *Developmental Science*, 3, 148–162.
- Fuster, J. M. (1973). Unit activity in prefrontal cortex during delayed-response performance: Neuronal correlates of transient memory. *Journal of Neurophysiology*, 36, 61–78.
- Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3½–7 years old on a Strooplike day–night test. *Cognition*, 53, 129–153.

- Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. *Handbook of Physiology: The Nervous System*, 5, 373–417.
- Grant, D. A., & Berg, E. A. (1948). The performance of topectomized patients on the University of Wisconsin Card Sorting Test. *American Psychologist*, 3, 360.
- Kessels, R. P. C., Postma, A., Wijnalda, E. M., & de Haan, E. H. F. (2000). Frontal lobe involvement in spatial memory: Evidence from PET, fMRI, and lesion studies. *Neuropsychology Review*, 10, 101–113.
- 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.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109, 163–203.
- McClelland, J. L. (1993). Toward a theory of information processing in graded, random, interactive networks. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience—A Silver Jubilee Volume* (pp. 655–688). Cambridge, MA: MIT Press.
- 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., & 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 prefrontal 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.
- Monchi, O., Taylor, J. G., & Dagher, A. (2000). A neural model of working memory processes in normal subjects, Parkinson’s disease, and schizophrenia for fMRI design and predictions. *Neural Networks*, 13, 953–973.
- Monsell, S., & Driver, J. (2000). *Attention and performance XVII*. Cambridge, MA: MIT Press.
- Morton, J. B., & Munakata, Y. (in press). 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, 834–843.
- Morton, J. B., Trehub, S. E., & Zelazo, P. D. (2002). 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 AB task. *Developmental Science*, 1, 161–184.
- Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences*, 5, 309–315.
- 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., & Yerys, B. E. (2001). All together now: When dissociations between knowledge and action disappear. *Psychological Science*, 12, 335–337.
- Newell, A. (1990). *Unified theories of cognition. The William James lectures, 1987* (pp. 549). Cambridge, MA: Harvard University Press.
- 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.
- O’Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain* (pp. 504). Cambridge, MA: MIT Press.
- Piaget, J. (1941). Le mecanisme du developpement mental et les lois du groupement des operation. *Arch. Psych., Genev e*, 28, 215–285.
- Piaget, J. (1967). *Six psychological studies*. New York: Random House.
- Roberts, R. J., & Pennington, B. F. (1996). An interactive framework for examining prefrontal cognitive processes. *Developmental Neuropsychology*, 12, 105–126.
- Rumelhart, D. E., McClelland, J. L. & PDF Research Group (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations*. Cambridge, MA: MIT Press.
- Shimamura, A. P., Gershberg, F. B., Jurica, P. J., Mangels, J. A., & Knight, R. (1992). Intact implicit memory in patients with frontal lobe lesions. *Neuropsychologia*, 30, 931–937.
- 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, 1020–1023.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.
- Stuss, D., & Benson, D. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin*, 95, 3–28.
- Zald, D. H., & Iacono, W. G. (1998). The development of spatial working memory abilities. *Developmental Neuropsychology*, 14, 563–578.
- 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: Benjamins.
- Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, 11, 37–63.