

## When Labels Hurt but Novelty Helps: Children's Perseveration and Flexibility in a Card-Sorting Task

Benjamin E. Yerys  
*University of Denver*

Yuko Munakata  
*University of Colorado, Boulder*

Children often persevere, repeating prior behaviors when inappropriate. This work tested the roles of verbal labels and stimulus novelty in such perseveration. Three-year-old children sorted cards by one rule and were then instructed to switch to a second rule. In a basic condition, cards had familiar shapes and colors and both rules were stated explicitly. In an uninformative-label condition, cards had familiar shapes and colors, but the first rule was not stated explicitly. In a novel-stimuli condition, both rules were stated explicitly but stimuli were novel on the first sorting dimension. More children switched to the second rule in the uninformative-label and novel-stimuli conditions than in the basic condition. Implications for theories of cognitive flexibility are discussed.

At one time or another, everyone engages in perseverative behavior—repeating an old behavior when it is not appropriate. Some people may continue to look in their pockets for their keys, even though they have already searched there. Others may repeatedly generate the same incorrect word for an item in a crossword puzzle.

Children are particularly robust perseverators. For example, infants reliably persevere in searching for toys in prior hiding locations, even when they have seen the toys hidden in new locations (Piaget, 1954). Three-year-olds persevere by continuing to sort pictures on cards by one rule (e.g., according to their shape), even when asked to switch and sort the cards by a second rule (e.g., according to their color; Zelazo, Frye, & Rapus, 1996). Young school-aged children persevere by continuing to judge a speaker's emotion by one feature (the content of what the speaker says), even when asked to switch and judge emotion by another feature (the tone of the speaker's voice; Morton, Trehub, & Zelazo, 2003).

To understand why such perseveration occurs, researchers have investigated the effects of a variety of experimental manipulations on perseveration. These studies span a range of populations, including typically developing children (e.g., Brooks, Hanauer, Padowska, & Rosman, 2003; Diamond, 1985; Jacques,

Zelazo, Kirkham, & Semcesen, 1999; Kirkham, Cruess, & Diamond, 2003; Marcovitch & Zelazo, 1999; Morton & Munakata, 2002b; Morton & Trehub, 2001; Munakata & Yerys, 2001; Perner & Lang, 2002; Towse, Redbond, & Cook, 2000; Wellman, Cross, & Bartsch, 1986; Zelazo, Mueller, Frye, & Marcovitch, 2003), neurologically intact adults (e.g., Diamond & Kirkham, 2005; Dunbar & Sussman, 1995), special populations of children and adults, with brain lesions or disorders (Casey, 2001; Diamond, 2000; Griffith, Pennington, Wehner, & Rogers, 1999; Milner, 1963), and members of other species (e.g., Diamond & Goldman-Rakic, 1989; Dias, Robbins, & Roberts, 1997; Hauser, 1999; Hood, Hauser, Anderson, & Santos, 1999).

The work reported here focuses on experimental manipulations of the novelty of stimuli and the use of informative labels, and their effects on children's perseveration in sorting cards. These studies were motivated by theories of perseveration that focus on competing memory systems, as described next. The findings should also inform and constrain alternative theories of perseveration, a point we return to in the Discussion.

Many researchers have proposed that a competition between two kinds of memory systems leads to perseveration, as well as ultimate success in behaving more flexibly (Butterworth, 1977; Cohen, Dunbar, & McClelland, 1990; Diamond, 1985; Goldman-Rakic, 1987; Munakata, 1998; Roberts & Pennington, 1996; Zelazo, Reznick, & Spinazzola, 1998). For example, a habit system may build up memories as a behavior is repeated (such as sorting cards by shape),

---

This research was supported by NICHD HD37163 and NSF IBN-9873492, and was based on the master's thesis of the first author. We thank members of the Cognitive Development Center for useful comments on drafts of this paper, and Amanda Bowles and Monika Offermann for lab assistance.

Correspondence concerning this article should be addressed to Yuko Munakata, Department of Psychology, University of Colorado, Boulder, CO 80309-0345. Electronic mail may be sent to munakata@colorado.edu.

leading to a bias to repeat those behaviors. A working memory system may serve to actively maintain information (such as a new rule for sorting cards, by their color), allowing currently relevant information to guide behavior. When currently relevant information conflicts with previously relevant information, working memory and habit systems may compete with one another. Perseveration occurs when habits are stronger than working memories, whereas flexible switching is observed when working memories are stronger than habits (Diamond, 1985).

This kind of competition between memory systems has been explored in neural network models (e.g., Cohen et al., 1990; Cohen & Servan-Schreiber, 1992; Dehaene & Changeux, 1989; Morton & Munakata, 2002a; Munakata, 1998; Stedron, Sahni, & Munakata, 2005). Across all of these models, information can be represented and remembered in two distinct forms: active versus latent (Munakata, 1998). Active representations take the form of sustained activations of processing units, roughly corresponding to the maintained firing of neurons. This kind of maintained firing allows one to hold information actively in mind. Active memories consist of active representations in the face of little environmental support for this information (in contrast with active representations driven by stimuli in the environment). Active memories are viewed as one component of working memory. Neural recording, imaging, and patient studies indicate that active representations rely heavily on prefrontal cortical regions (e.g., Diamond, 2002; Miller & Cohen, 2001; Stuss & Benson, 1984). Latent representations take the form of the strength of connection weights between processing units, roughly corresponding to the efficacy of synapses. Changes in synaptic efficacy can facilitate processing of information, and may be one mechanism for laying down habits. Latent memories consist of latent representations formed through an individual's experience (in contrast with latent representations driven by evolutionary history). Latent representations are thought to tap posterior cortical and subcortical areas (e.g., Miller & Desimone, 1994; Steinmetz, Connor, Constantinidis, & McLaughlin, 1994).

Neural network models have demonstrated how a competition between active and latent representations can lead to perseveration (if the latent representations win) or flexible behavior (if the active representations win) (e.g., Cohen et al., 1990; Cohen & Servan-Schreiber, 1992; Dehaene & Changeux, 1989; Morton & Munakata, 2002a; Munakata, 1998; Stedron et al., 2005). For example, a neural network

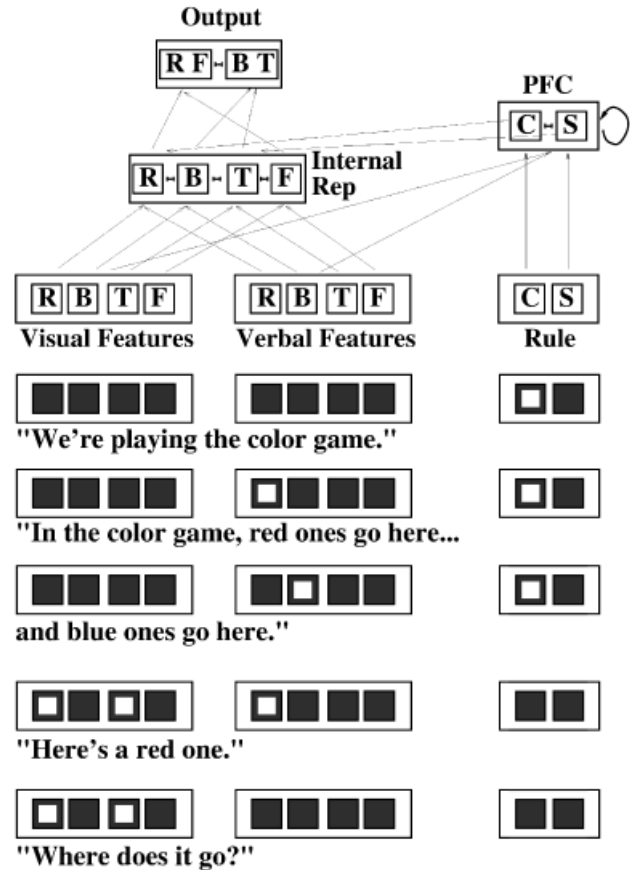


Figure 1. Card-sorting network and one trial (adapted from Morton & Munakata, 2002a): The network receives visual and verbal information about the stimuli (R = red, B = blue, T = truck, F = flower) and sorting rules (C = color, S = shape). Red trucks and blue flowers are presented to the network. Its response options for each card are to sort it with either the red flower target card or the blue truck target card. A trial consists of sequences of patterns of activity, shown below the network, with the corresponding verbal information presented to the child shown in quotations. The description of the rules also includes activation (not shown) of the relevant output units, corresponding to the experimenter drawing attention to the target cards to show where "red ones go" and where "blue ones go." Latent representations take the form of changes in feedforward connection weights; active representations take the form of sustained activation on the "PFC" layer. The model perseverates when latent representations of prior rules are stronger than active representations of the current rule, and switches when active representations are stronger than latent.

Note. From "Active versus latent representations: A neural network model of perseveration and dissociation in early childhood," by J. B. Morton and Y. Munakata, 2002a, *Developmental Psychology*, 40, p. 259. Copyright 2002 by John Wiley & Sons, Inc. Adapted with permission.

model simulated children's behavior on the card-sorting task described above (Morton & Munakata, 2002a). This model (Figure 1) demonstrated how la-

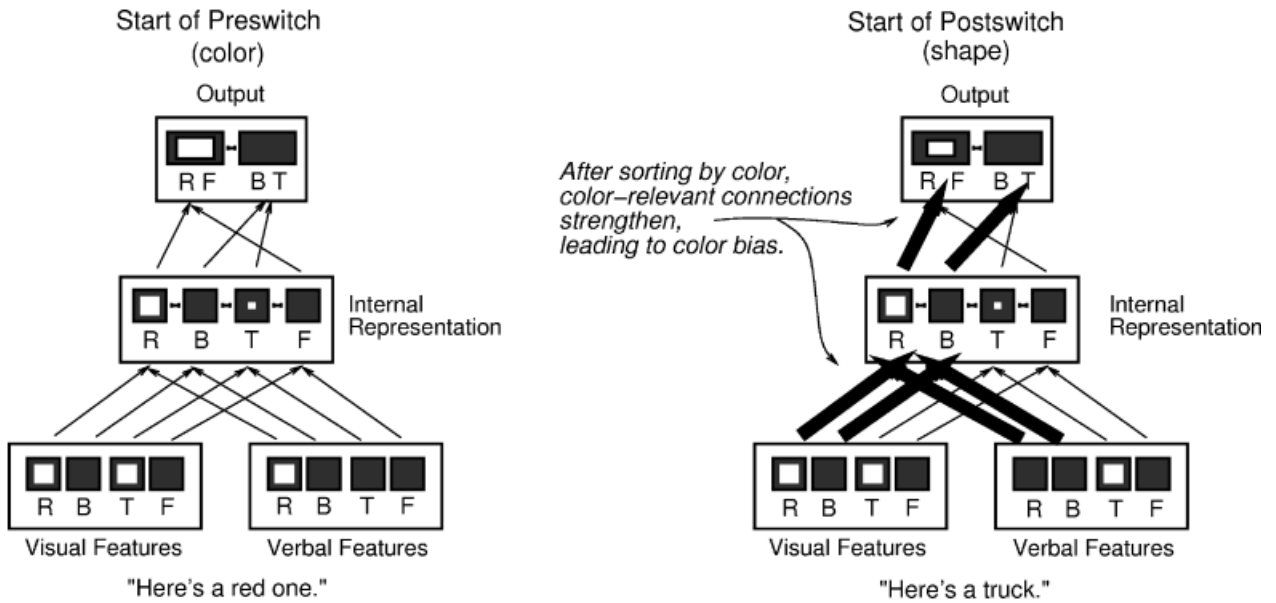


Figure 2. Schematic representation of preswitch and postswitch trials: Weights strengthen with experience with the preswitch rule, making it more difficult to use a postswitch rule. Active representations of the postswitch rule (not shown for simplicity) compete with these latent representations of the preswitch rule.

tent representations can build up as cards are sorted repeatedly according to one rule (the “preswitch” rule). As red trucks and blue flowers were repeatedly sorted according to color, corresponding units in the network were repeatedly activated. Connections between units changed according to a Hebbian learning rule, whereby “units that fire together, wire together.” As a result, strong connections were created between units representing color and units representing corresponding locations for sorting the cards. These latent representations led to a bias to respond to stimuli according to the first rule used (Figure 2). When the rule changed (e.g., to sort according to shape), the network’s ability to respond flexibly to the new task depended on the strength of active representations of the new rule (the “postswitch” rule). This strength was increased through increases to recurrent excitatory connections, corresponding to developmental changes in prefrontal cortical regions. These recurrent excitatory connections allowed units to activate themselves to maintain their activity. With weak active representations of the new rule to guide behavior, the network’s internal representations were dominated by the latent representations, leading to perseveration. Stronger active representations of currently relevant information led to flexible behavior.

In this simulation of a competing systems account, representations (both active and latent) are graded in nature (Munakata, 2001). That is, representations vary in strength rather than being all-or-nothing,

present or absent (e.g., Farah, Monheit, & Wallace, 1991; Farah, O’Reilly, & Vecera, 1993; Munakata, McClelland, Johnson, & Siegler, 1997; Rumelhart & McClelland, 1986). For active representations, strength might be instantiated in terms of a number of factors, including the number of neurons contributing to a representation, and the firing rates and coherence of those neurons. For latent representations, strength might be instantiated in terms of factors such as the number and efficacy of postsynaptic receptors, and the amount of neurotransmitters released. The strength of representations may vary based on one’s state of development and prior history, and the current environmental context. As a result of such gradedness, behavior may be highly task dependent (see also Thelen, Schoner, Scheier, & Smith, 2001; Thelen & Smith, 1994).

For example, in the card-sorting task, a child with a weak ability to actively represent new rules might use these representations to succeed on some tasks but not on others. In the card-sorting model described above, weak active representations allowed the model to successfully answer questions about new rules (such as where trucks go in the shape game), but these representations were too weak to allow the model to successfully sort cards with red trucks on them. The model instead sorted such cards perseveratively, according to their color. Stronger active representations were needed for sorting such cards, because the conflicting information from the other dimension (e.g., the redness of the truck) tap-

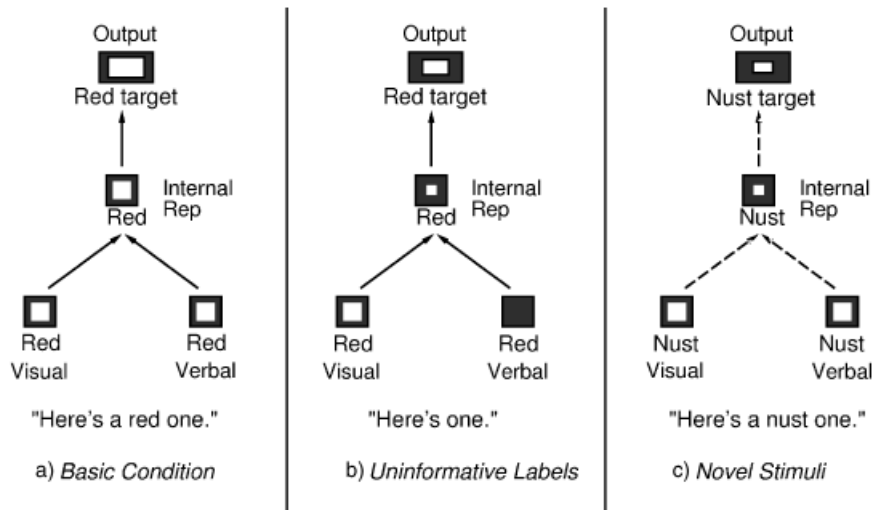


Figure 3. Predicted effects of manipulating the informativeness of labels and the novelty of stimuli, shown for just one feature for simplicity: (a) The basic condition includes visual and verbal inputs and connections established for familiar stimuli, while (b) the uninformative-labels condition is missing verbal input and so has weaker activation (denoted by smaller white squares), which will in turn lead to weaker changes to connection weights, and (c) the novel-stimuli condition has weaker connections (denoted by dashed lines), leading to weaker activations. As described in the Method section, “nust” refers to a relatively unfamiliar green-brown color.

ped the latent representations biasing the system toward this dimension. Graded representations in this model thus led to the same “knowledge–action” dissociations observed in children in this task (Munakata & Yerys, 2001; Zelazo et al., 1996), and reported anecdotally for patients with damage to prefrontal cortex in a similar task (Milner, 1963). In addition, the model predicted that such knowledge–action dissociations would disappear when tasks were equated for conflict (e.g., by asking where *red* trucks go in the shape game, and requiring a red truck to be sorted). This prediction has been confirmed in children (Munakata & Yerys, 2001; see also Morton & Munakata, 2002b).

Similarly, an infant with a weak ability to actively represent a toy’s new hiding location might succeed on some tasks but not on others. For example, weak active representations of a hidden toy allowed a neural network model to show looking–reaching dissociations like those observed in infants (Diamond, 1985; Hofstadter & Reznick, 1996; Piaget, 1954). That is, the model could use weak active representations to gaze at a toy’s new hiding location, but those representations were too weak to support successful reaching to that location (Munakata, 1998). The model instead reached perseveratively to the previous hiding location. Stronger active representations were needed for reaching than for gazing, because these systems differed in how often they could update, in a manner designed to parallel infants’ experience in studies with hidden

toys. Specifically, the gazing system could continually update its behaviors and so could make use of relatively weak representations, whereas the reaching system could respond only at particular times.

If perseveration and flexibility reflect a competition between two kinds of graded memory systems, performance should be influenced by experimental manipulations designed to affect the strength of the different memory systems. This study investigates two such manipulations in children’s card sorting. One manipulation focuses on the strength of active representations formed during the task. Such active representations might be relatively weak if relevant linguistic information is removed from the task. Although we have focused on top-down contributions to active memory from the prefrontal cortex, bottom-up contributions from the environment may also affect the strength of active representations. For example, children might form weaker active representations for a stimulus if an informative verbal label for it (such as “red”) is no longer provided (Figures 3a vs. b). The second manipulation focuses on the strength of latent representations brought into a task from prior experience. Such latent representations should be relatively weak for unfamiliar stimuli (e.g., nonsense shapes or atypical colors), because they have not been processed repeatedly and therefore connections have not been shaped extensively by them. For example, children should have weaker latent representations for novel stimuli than for familiar stimuli in terms of the connections

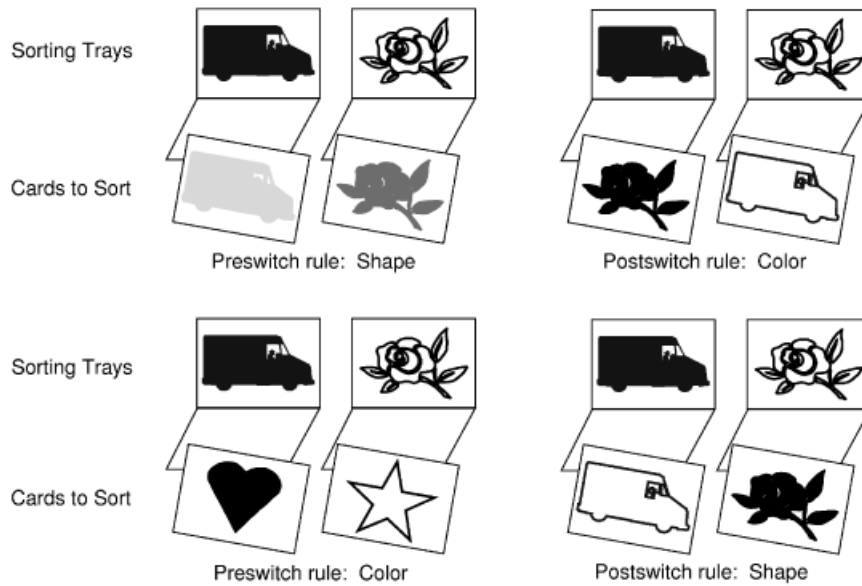


Figure 4. Sample stimuli for basic-PPC and uninformative-labels conditions. The top row shows how cards should be sorted when the preswitch rule is shape and the postswitch rule is color, and the bottom row shows how cards should be sorted when the preswitch rule is color and the postswitch rule is shape. Each preswitch card matches one target according to the relevant dimension only; the feature in the other dimension does not match either target. Each postswitch card matches one target according to one dimension, and the other target according to the other dimension. This version of the task is similar to Zelazo et al.'s (2003) "partial-partial change" version, and was chosen for reasons described in the text. Black and white shading correspond to red and blue respectively, and light gray and dark gray shading correspond to yellow and green respectively.

from perceptual inputs to internal representations, as well as the connections from those internal representations to the corresponding responses (Figures 3a vs. c).

Note that active and latent representations are conceptually distinct, but highly interactive. For example, latent representations should be relatively weak for novel stimuli. This would lead to weaker active representations when those stimuli are presented, which would in turn lead to smaller changes to the connections with experience because connection changes are a function of activity levels. Similarly, stimuli with uninformative labels should lead to weaker active representations of the stimuli, which in turn would lead to weaker habits being laid down while processing those stimuli. Thus, although we focus on weaker latent representations for novel stimuli and weaker active representations for uninformative labels, this is only the starting point, with each of these manipulations ultimately affecting the other type of representation.

By varying whether the card-sorting stimuli are basic, novel, or presented with uninformative labels, there are nine possible preswitch-postswitch combinations involving this experimental manipulation. The effects of this manipulation could be tested most directly by holding preswitch constant and manipulating postswitch, or by holding postswitch

constant and manipulating preswitch.<sup>1</sup> Either approach could address the theoretical motivations for the study. We chose to hold postswitch constant and manipulate preswitch, because the experimental manipulations might have their greatest effect built up across preswitch trials. Moreover, the predictions in this direction seemed more counterintuitive and thus potentially more interesting, as described next.

First, if active representations are relatively weak when linguistic information is removed from the task, children should show more flexibility in switching after sorting stimuli with uninformative labels. With relatively weak active representations during preswitch, the latent representations created would in turn be weaker. Children's active representations of the new sorting rule should be better able to compete against these relatively weak latent representations of the preswitch rule. This should occur even if children sort the stimuli with uninformative labels just as often and just as successfully as in the standard condition during preswitch. This prediction may be somewhat counterintuitive, because informative labels often aid children's performance on cognitive tasks (Deak & Bauer, 1996; Gentner & Loewenstein, 2001; Jacques & Zelazo, 2003; Kirkham et al., 2003; Towse et al., 2000). In many of these cases however, these labels are relevant to the task at hand, rather than to the prior task

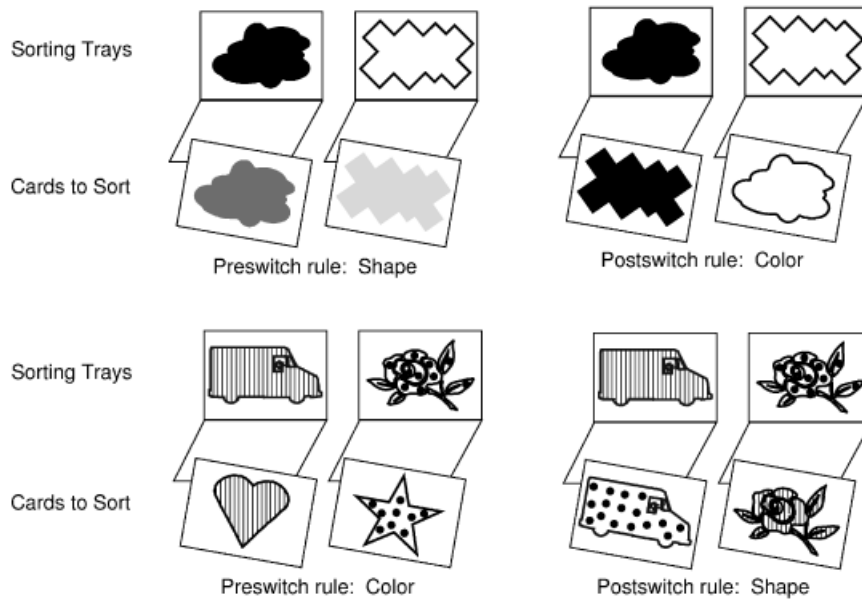


Figure 5. Sample stimuli for Novel stimuli condition. The top row shows how cards should be sorted when the preswitch rule is shape and the postswitch rule is color, and the bottom row shows how cards should be sorted when the preswitch rule is color and the postswitch rule is shape. As in the other conditions, each preswitch card matches one target according to the relevant dimension only; the feature in the other dimension does not match either target. Each postswitch card matches one target according to one dimension, and the other target according to the other dimension. For the novel shapes, the cloud-like shape was a “dax” and the jagged shape was a “gub.” Again, black and white shading correspond to red and blue respectively, and light gray and dark gray shading correspond to yellow and green respectively. For the novel colors, stripes represent a purple-gray color (“flirp”), and polka dots represent a green-brown color (“nust”).

as in this study. In one case however (Jacques & Zelazo, 2001a), informative labels relevant to a prior task improve performance, a point we return to in the Discussion. In other circumstances, labels can impair performance (Chambers & Reisberg, 1985; Liittschwager & Markman, 1994; Schooler & Engstler-Schooler, 1990).

Second, if latent representations are relatively weak for novel stimuli, children should show more flexibility in switching after sorting stimuli according to unfamiliar features. Again, children’s active representations of the new sorting rule should be better able to compete against these relatively weak latent representations of the previous sorting rule. And, this should occur even if children sort the unfamiliar stimuli during preswitch just as often and just as successfully as in the standard condition. This prediction may also be somewhat counterintuitive, because novelty often serves to draw children’s attention, rather than helping them to attend to other information (e.g., Kersten & Smith, 2002; Ross, 1974). However, in some circumstances, it is easier to switch from a less familiar task to a more familiar one (MacLeod & Dunbar, 1988; Morton et al., 2003).

To test these ideas, we assessed 3-year-old children’s abilities to switch from a preswitch rule to a

postswitch rule in three conditions of the card-sorting task. For reasons described below, these conditions built on a variation of the standard card-sorting task, known as the “partial–partial change” version (Zelazo et al., 2003), which differs from the standard version in the stimuli used during the preswitch phase. Specifically, each preswitch sorting card matched one target card according to the relevant dimension only; the feature in the other dimension did not match either target (Figures 4 and 5). For example, if the target cards were red trucks and blue flowers and the preswitch rules were shape, the sorting cards might be yellow trucks and green flowers. As in the standard card-sorting task, each postswitch sorting card matched one target card according to one dimension, and the other target card according to the other dimension. For example, if the target cards were red trucks and blue flowers, the sorting cards would be red flowers and blue trucks.

This version of the task was chosen for multiple reasons. First, our piloting work building on the standard version of the task revealed potential floor effects in children’s postswitch performance, so that the effects of our experimental manipulations could not be assessed. Children show a trend to switch more readily in the partial–partial change version

than in the standard version (Zelazo et al., 2003); hence we adopted that variant of the task to allow comparisons across our experimental manipulations with this age group. The partial–partial change version may be easier because the features change between preswitch and postswitch in the to-be-attended dimension, perhaps drawing children’s attention. We consider another possible reason in the Discussion. Second, we speculated that the presence of conflicting features during preswitch, as in the standard task, might increase the likelihood of children spontaneously labeling the relevant features in an attempt to deal with this conflict. For example, with target cards of red trucks and blue flowers, a child faced with a blue truck to sort by shape might be more likely to label it as a “truck,” in an attempt to deal with the conflicting color information, than a child faced with a yellow truck to sort. Any such spontaneous labeling would reduce differences between our conditions with and without informative labels. We thus opted to use partially matching cards during preswitch, in an effort to reduce conflict and in turn possibly reduce spontaneous labeling, thereby increasing sensitivity to differences between labeling conditions. (An even more extreme step to reduce conflict from the irrelevant dimension is possible: using identical sorting and target cards [e.g., red trucks and blue flowers as sorting and target cards]. However, such cards preclude assessing what rule children are using to sort the cards, hence partially matching cards represent a better option.) Whether these spontaneous-labeling and attention-drawing ideas are correct is irrelevant to the questions of interest in the current study. The partially matching version of the task simply serves as a basis for assessing the key manipulations of labels and novelty. The nuances specific to this version of the task are held constant across conditions, while labels and novelty are varied so their effects can be assessed. We compared performance across three conditions: basic-PPC (so named because it is the most basic version of the task used in this study, and it resembles the partial–partial change version more than the standard version of this task), uninformative labels, and novel stimuli.

## Method

### *Participants*

Seventy-two normally developing 3-year-old children (mean = 36 months 2 days; range = 35 months 24 days to 36 months 7 days) participated in

this experiment and were included in the final analyses, with 24 children in each condition. There were 12 males and 12 females in the basic-PPC condition, 10 males and 14 females in the uninformative-labels condition, and 11 males and 13 females in the novel-stimuli condition. There were 61 Caucasian children, 4 Hispanic/Latino children, 4 children with multiple ethnicity/races, 2 unreported, and 1 African American child. Participants came predominantly from middle-class backgrounds; however, this information was not systematically recorded. Participants were recruited from a departmental participant pool. Children received a small gift for their participation and parents received \$5 to cover travel costs. Additional children participated in the experiment but their data were excluded from analyses for the following reasons. Five children failed preswitch (2 in the uninformative-labels condition and 3 in the novel-stimuli condition); their data were excluded because the measure of interest was children’s ability to switch in postswitch after passing preswitch. In addition, data were excluded for 5 participants from the basic-PPC condition (3 due to parent interference, 1 for refusing to sort the cards, and 1 for not understanding the task), 5 participants from the uninformative-labels condition (3 due to experimental error, 1 due to parent interference, and 1 due to fussiness), and 4 participants from the novel-stimuli condition (2 due to experimental error and 2 due to parent interference).

### *Design*

Participants were tested in one of the three conditions: basic-PPC, uninformative labels, and novel stimuli. Each condition involved two phases: a preswitch and a postswitch phase. The preswitch phase involved two demonstration trials and six test trials. The postswitch phase involved six test trials. Within each condition, participants were counterbalanced for preswitch dimension (shape or color), targets (in the basic-PPC and uninformative-labels conditions, red truck and blue flower, or red flower and blue truck; in the novel stimuli condition [with novel shapes and colors referred to by the labels used in the experiment], red “gub” and blue “dax,” red dax and blue gub, “nust” truck and “flirp” flower, or nust flower and flirp truck], and target locations (right vs. left).

### *Stimuli*

There were two trays (approximately 17 cm × 10.5 cm × 3.5 cm) with a target (a 13 cm × 8 cm card)

attached to each. The trays were constructed by gluing an index cardholder to a plastic bookend, and the targets were attached via Velcro<sup>®</sup> to the bookends. The cards to be sorted were the same size as the targets. In all conditions, each preswitch sorting card matched one target card according to the relevant dimension only, and each postswitch sorting card matched one target card according to one dimension, and the other target card according to the other dimension. In the basic-PPC and uninformative-labels conditions, the target cards were red trucks and blue flowers (or blue trucks and red flowers). When the preswitch rule was shape, the sorting cards were yellow trucks and green flowers (Figure 4, top left). When the preswitch rule was color, the sorting cards were red hearts and blue stars (Figure 4, bottom left). In the postswitch phase, the sorting cards were blue trucks and red flowers (or red trucks and blue flowers; Figure 4, right-hand side). In the novel-stimuli condition, the novel shapes were a "dax" (a cloud-like shape) and a "gub" (a jagged shape), and the novel colors were "flirp" (a purple-gray: Hue = 317, Saturation = 61%, Brightness = 83%) and "nust" (a green-brown: Hue = 52, Saturation = 40%, Brightness = 70%). When the preswitch rule was shape, the target cards were red daxes and blue gubs (or blue daxes and red gubs), the preswitch sorting cards were yellow daxes and green gubs, and the postswitch sorting cards were green gubs and yellow daxes (or red daxes and blue gubs; Figure 5, top row). When the preswitch rule was color, the target cards were flirp trucks and nust flowers (or nust trucks and flirp flowers), the preswitch sorting cards were flirp hearts and nust stars, and the postswitch sorting cards were nust trucks and flirp flowers (or flirp trucks and nust flowers; Figure 5, bottom row).

#### *Procedure*

In all conditions, children sat across a table from the experimenter. There were two trays on the table, and one target was fastened to each tray. The experimenter provided the rules to sort cards by one dimension, demonstrated the rules by sorting one of each type of card, and asked the child to sort the next six cards (preswitch phase). Children were reminded of the rules on each trial, and feedback was provided. Children were then asked to sort six cards according to a different rule, and were reminded of the rules on each trial (postswitch phase). Cards were placed into the trays face down (or the experimenter flipped the child's card over if it was placed face up), so that they were not visible for

subsequent trials. Details and one example for each of the three conditions are provided below.

#### *Basic-PPC Condition*

*Preswitch instructions.* "We are going to play a game. This is the shape game. In the shape game, trucks go here [experimenter waves a yellow truck in front of the red truck tray] and flowers go there [experimenter waves a green flower in front of the blue flower tray]. I'll go first and then it will be your turn."

*Demonstration.* The experimenter placed one card (e.g., green flower) into its tray, and placed the other card (e.g., yellow truck) into the other tray.

*Preswitch trials.* "We are playing the shape game. In the shape game, trucks go here [experimenter waves a yellow truck in front of the red truck tray] and flowers go there [experimenter waves a green flower in front of the blue flower tray]." The experimenter randomly selected and labeled a card by the relevant dimension (e.g., "Here's a truck.") and asked, "Where does this go in the shape game?" Children placed the card into one of the trays. If the sort was correct the experimenter said, "Good job"; if the sort was incorrect the experimenter said, "That was not right. This one goes here." and moved the card. If the children did not sort the card in 5 s, the experimenter prompted them by saying, "Where does that one go?"

*Postswitch instructions.* "Okay, now we're going to switch and play a new game, the color game. We are not going to play the shape game any more. No way. We are going to play the color game and the color game is different. In the color game, red ones go here [experimenter points to the red truck tray] and blue ones go there [experimenter points to the blue flower tray]."

*Postswitch trials.* "We are playing the color game. In the color game, red ones go here [experimenter points to the red truck tray] and blue ones go there [experimenter points to the blue flower tray]." The experimenter randomly selected and labeled a card by the relevant dimension (e.g., "Here's a red one.") and asked, "Where does this go in the color game?" Children placed the card into one of the trays.

#### *Uninformative-Labels Condition*

The uninformative-labels condition was identical to the basic-PPC condition with the exception of the preswitch instructions (and corresponding reminders before each trial): "We are going to play a game. This is the sorting game. In the sorting game, these go here [experimenter waves a yellow truck in front of the red truck tray], and these go there

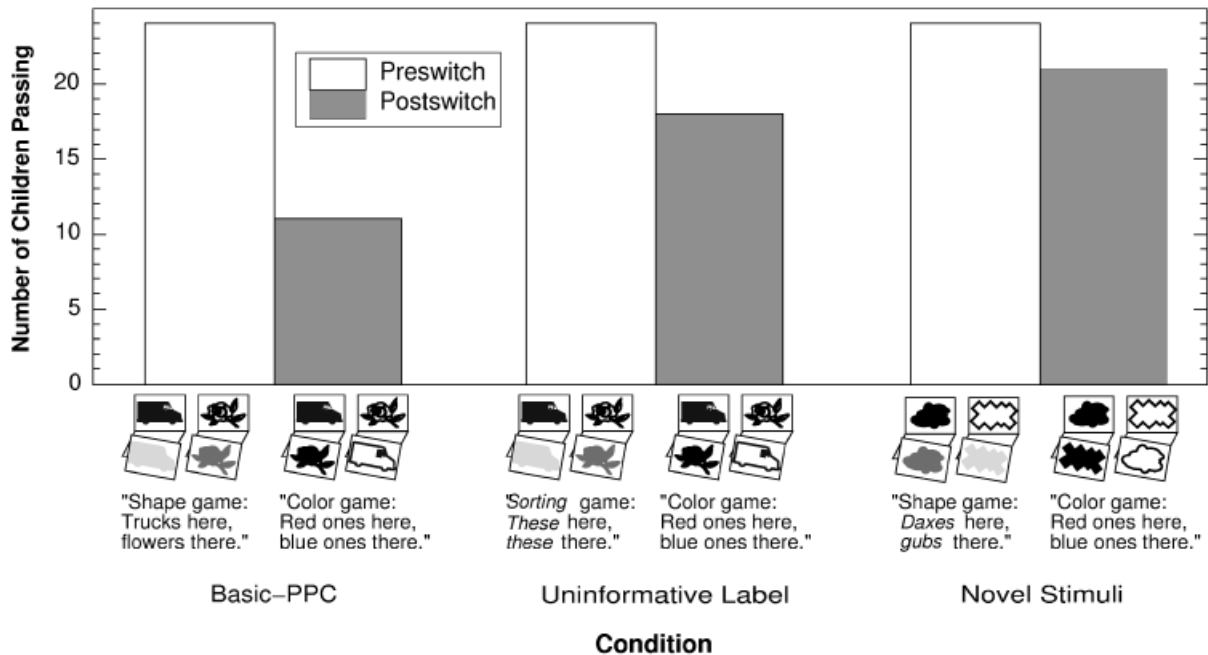


Figure 6. Numbers of children passing preswitch and postswitch in the basic-PPC, uninformative-labels, and novel-stimuli conditions ( $N = 24$  in each condition): Children were significantly better at switching to the new rule in the uninformative-labels and novel-stimuli conditions. Shape-to-color examples are shown for each condition; differences in instructions are italicized.

[experimenter waves a green flower in front of the blue flower tray]. I'll go first and then it will be your turn."<sup>2</sup>

#### Novel-Stimuli Condition

The novel-stimuli condition was identical to the basic-PPC condition with the exception of the stimuli and the preswitch instructions (and corresponding reminders before each trial): "We are going to play a game. This is the shape game. In the shape game, daxes go here [experimenter waves a green dax in front of the red dax tray], and gubs go there [experimenter waves a yellow gub in front of the blue gub tray]. I'll go first and then it will be your turn."

### Analyses and Results

Initial descriptive analyses indicated that the data were nonnormal, consistent with previous findings (Kirkham et al., 2003; Munakata & Yerys, 2001; Zelazo et al., 1996). In preswitch, most children (94.4%) sorted all cards correctly. In postswitch, most children (84.7%) sorted all cards correctly or sorted all cards incorrectly. Thus, as in previous studies, the measure of interest was whether children passed or failed postswitch sorting, and chi-squares were used to analyze the data. And, as in previous studies

(e.g., Munakata & Yerys, 2001), children were classified as passing each sorting phase if they sorted at least four out of the six cards correctly. Other studies have used a stricter criterion of sorting at least five out of the six cards correctly (Kirkham et al., 2003). The pattern of results in the current study is similar across the two criteria; only 3 of the 72 children change classifications with the stricter criterion. It might seem odd to use a categorical measure of performance to test a graded representations framework. However, in this case the data are not conducive to continuous measures, and the categorical measure is sufficient to detect differences predicted by the graded representations framework. An important step for future work will be to develop measures that allow seemingly categorical behaviors to be assessed in a more continuous way (e.g., Diedrich, Thelen, Smith, & Corbetta, 2000; Spencer & Schutte, 2004).

All children passed preswitch, as this was a prerequisite for inclusion in the study. As predicted, the uninformative labels and novel stimuli improved children's postswitch sorting substantially (Figure 6). Only 11 of 24 children passed postswitch in the basic-PPC condition, compared to 18 of 24 children in the uninformative-labels condition,  $\chi^2(1, N = 48) = 4.3$ ,  $p < .05$ , and 21 of 24 children in the novel-stimuli condition,  $\chi^2(1, N = 48) = 7.4$ ,  $p < .01$ .

## Discussion

Consistent with predictions from a graded, competing representations framework, more children successfully switched to sorting cards by a new rule when the first sorting rule involved uninformative labels or novel stimuli. Removing informative labels for the first rule should have weakened children's active representations during that phase. As a result, the latent representations biasing children toward that rule were in turn weaker, making it easier for children to switch to a new rule using active representations of the new rule. And, sorting novel stimuli for the first rule should have tapped into weaker latent representations, also making it easier for children to use active representations to switch to a new rule.

The finding that children were more flexible with uninformative labels for the first rule might be somewhat counterintuitive, given that labels often aid children's performance. In many of those cases however, the labels are relevant to the task at hand, rather than to a prior task (e.g., Deak & Bauer, 1996; Gentner & Loewenstein, 2001; Kirkham et al., 2003; Towse et al., 2000). For example, when children are asked to label cards according to the currently relevant dimension during postswitch, they are more likely to successfully sort according to the postswitch rule (Kirkham et al., 2003; Towse et al., 2000). However, in one task, the Flexible Item Selection Task (FIST; Jacques & Zelazo, 2001a), labels relevant to a prior task improve performance on the current task. In this paradigm, children are presented with three cards. Two of the cards match in one way, and two of the cards match in another way. For example, a trio of cards might consist of a purple phone, a pink phone, and a pink fish. Children are first asked to pick two cards that go together, and they might pick the two phones. They are then asked to pick two cards that go together in a different way. At this point, they should pick the two pink cards. This task, like the card-sorting task, requires children to see a single stimulus in two ways; the pink phone must be seen both as a phone and as pink. Four-year-old children generally succeed in picking one pair of cards that go together, but then have difficulty selecting the second pair. Labels in the first task improve performance in the second task. That is, 4-year-old children are better at selecting the second pair if the experimenter labels relevant information in the first pair, or if the children are asked to explain how the two cards go together in the first pair they select (Jacques & Zelazo, 2001b).

What accounts for the discrepancy in labels' helping versus hurting performance? The current

tasks and the FIST differ in various ways. For example, the FIST requires two cards to be picked from three, while the card-sorting task requires a single card to be sorted. And, children must generate for themselves the rule to use in the FIST, whereas they are told what rule to use in the card-sorting task. Perhaps labels for the first rule provide constraints on the kinds of rules that can be considered, so labels help when children must then generate another rule to sort by, as in the FIST. That is, labels during a prior task may strengthen active representations as posited here, making it generally harder to switch to a new task, but these difficulties are outweighed by the advantages of such labels when switching to a new task where the rule options are not specified. Another difference between these studies is the age of the participants: 3-year-olds had more difficulty switching to a new dimension after the first rule was labeled (current findings), while 4-year-olds had less difficulty switching to a new dimension after the first one was labeled (FIST findings). Labels may strengthen active representations as posited here, but developmental changes between 3 and 4 years may influence what children can do with those relatively strong active representations. For example, children may develop better gating mechanisms that influence whether information is maintained or updated in working memory (Frank, Loughry, & O'Reilly, 2001); such mechanisms may act more effectively on strong representations. Further work is needed to evaluate the role of such factors in the effects of labels on flexibility.

The finding that children were more flexible with novel stimuli might also be somewhat counterintuitive, given that novelty often serves to draw children's attention, rather than helping them to attend to other information (e.g., Kersten & Smith, 2002; Ross, 1974). However, in this case, the dominant effect may have been the weaker associations for the novel stimuli, which made it easier for children to switch to a new rule.

Alternatively, one might argue that novel stimuli improved children's switching through the same processes as uninformative labels. That is, although informative labels were employed in the novel-stimuli condition, they may have been ineffective—children never learned, for example, what "dax" meant. As a result, improvements in performance may have been driven solely by the effectively uninformative labels, rather than by the novelty of the stimuli (R. Joseph, personal communication, April 2003). This alternative seems unlikely for two reasons. First, the novel stimuli were repeatedly and consistently named with distinct labels

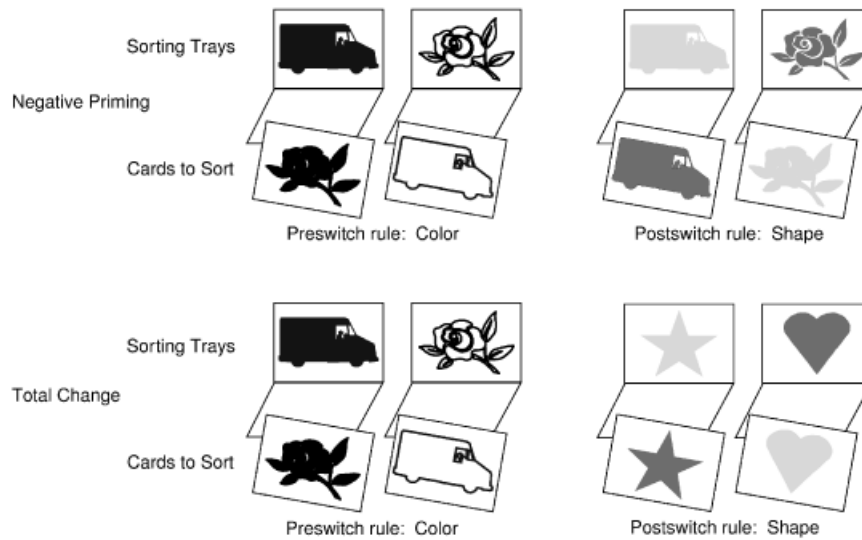


Figure 7. Sample stimuli for negative priming (top row) and total change (bottom row) variants. Preswitch is identical across these variants. In the negative priming variant, only features that were relevant during preswitch change between preswitch and postswitch. In the total change variant, features that were relevant during preswitch and features that will be relevant during postswitch change between preswitch and postswitch. Children perform relatively poorly on the negative priming variant, which is taken as evidence that they inhibited the irrelevant dimension during preswitch; they perform relatively well on the total-change version, which is taken as evidence that they do not perseverate on preswitch dimensions (Zelazo et al., 2003). We consider several alternative explanations in the text.

(a total of 18 times across the instructions and trials for the two stimuli). Children at this age and even much younger can learn novel labels for stimuli in many fewer trials (e.g., Heibeck & Markman, 1987; Markson & Bloom, 1997; Woodward, Markman, & Fitzsimmons, 1994). It thus seems unlikely that the distinct labels in the novel-stimuli condition were uninformative. Second, the game (color vs. shape) was labeled in the novel-stimuli condition but not in the uninformative-labels condition. Thus, *more* information was provided in the novel-stimuli condition. If uninformative labels alone helped children to switch, their performance should have been worse if anything in the novel-stimuli condition relative to the uninformative-labels condition. Instead, these conditions were not significantly different, and the direction of the difference went the other way, with more children switching in the novel-stimuli condition. Thus, the novelty of the stimuli likely contributed to children's greater ease in switching in this condition.

#### *Explaining Additional Findings Within the Graded, Competing Representations Framework*

Other aspects of children's card-sorting behavior have been put forth as challenges to various theories of children's perseveration. In this section, we consider three such findings and how they might be

explained within the graded, competing representations framework. The next section considers alternative theories of perseveration.

#### *Negative Priming*

Some researchers have argued that children perseverate in part due to a process of negative priming, whereby children ignore irrelevant rules during preswitch, which leads to difficulty sorting by those rules in postswitch (Zelazo et al., 2003). For example, when children sort by shape in preswitch, they ignore the conflicting information about color; as a result, they have difficulty activating color rules in postswitch. Children's difficulties in particular variants of card-sorting tasks are taken as evidence for negative priming. For example, after sorting red flowers and blue trucks by color, children perseverate with color when asked to sort yellow flowers and green trucks by shape (Zelazo et al., 2003; adapted in Figure 7, top row). The preswitch colors are no longer present in postswitch, therefore children could not perseverate on the specific features that were relevant in preswitch. The argument is that their difficulty must thus reflect a specific problem in attending to shape, as a result of inhibiting it during preswitch.

These findings are interesting and important, but we are not convinced that they indicate that irrelevant information is inhibited during preswitch. At

least two alternatives seem possible. One alternative is that children perseverate on preswitch rules in a way that can be applied to new exemplars. For example, children get stuck on the color dimension after sorting blue and red stimuli by color, so they then sort green and yellow stimuli by color too. Zelazo et al. (2003) tried to argue that children do not get stuck on overall dimensions such as color or shape, based on their performance in the total-change version of the task. In this version, both shapes and colors change between preswitch and postswitch; therefore, any *features* that children might perseverate on are removed (Figure 7, bottom row). Poor performance in postswitch would be interpreted as perseveration on the preswitch *dimension*. Children perform better on this total-change version of the task than on the standard task, which is taken as evidence that children do not perseverate on dimensions (Zelazo et al., 2003). However, a reasonable percent of children fail the total-change version of the task: in Zelazo et al. (2003), 21% of children failed this version in Experiment 7, 37% failed in Experiment 8, and 31% failed in Experiment 9. Thus, children may perseverate on dimensions. Such dimensional perseveration may contribute to their behavior on card-sorting tasks where the features they were attending to change after preswitch, such as negative-priming and total-change versions of the task. However, dimensional perseveration cannot be the only factor at work, given that children perform significantly worse on the negative-priming version of the task than on the total-change version, when both versions should lead to similar levels of dimensional perseveration. This performance difference might reflect benefits from changes between preswitch and postswitch in the features relevant to postswitch. Such changes are present in total change but not in negative priming (Figure 7), and could draw attention to the correct features in postswitch.

A second possibility is that rather than inhibit irrelevant information during preswitch, children activate it to some degree and learn associations that then hinder postswitch performance. For example, when sorting red flowers and blue trucks by color, children also activate information about flowers and trucks, which leads to incorrectly increasing associations between flowers to-sort and truck targets, and between trucks to-sort and flower targets. That is, they learn at some level that flowers go in the truck tray, and trucks go in the flower tray. Children have been shown to learn this form of “opposites” game for information they are attending to (Morton et al., 2003; Yerys & Munakata, 2001). When children are then asked to switch to sorting yellow trucks and

green flowers by shape, they have difficulty because they have learned incorrect associations between trucks and flowers. This would lead to greater difficulty on the negative-priming version of the task (where incorrect associations learned during preswitch must be overcome during postswitch) than on the total-change version (where incorrect associations learned during preswitch are not relevant during postswitch, due to changes in the stimuli between preswitch and postswitch). Similarly, this account may help to explain why children perform better on the partial–partial change version of the task, which the current studies built on, than on the standard version. Specifically, learning incorrect associations with irrelevant features during preswitch would not hurt postswitch performance as much in the partial–partial change version, because the features on the irrelevant dimension change between preswitch and postswitch. For example, incorrectly learning during preswitch that hearts go with truck targets and stars go with flower targets (Figure 4, bottom left) would not impair postswitch performance sorting trucks into the truck tray and flowers into the flower tray.

Thus, the findings that have been put forth as evidence for negative priming in children’s card-sorting are open to alternative interpretations. Similar points have been made in the related adult literature that motivated the negative priming investigations in children. That is, adults also show impairments when previously irrelevant information must then be attended to (Dalrymple-Alford & Budayr, 1966; Neill, 1977; Tipper, 1985). Various researchers have interpreted these effects in terms of a negative priming process (Tipper, 2001), but others have clearly articulated alternative explanations (e.g., MacDonald & Joordens, 2000; MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003; Neill & Valdes, 1992; Park & Kanwisher, 1994).

#### *Intra- Versus Extra-Dimensional Shifts*

Children are better at switching to a second rule in a card-sorting task if the shift is intra-dimensional (e.g., from flowers going in the flowers tray to flowers going in the trucks tray) rather than extra-dimensional (e.g., from flowers going in the flowers tray to red things going in the red tray) (Perner & Lang, 2002, cf. Brooks et al., 2003).

The graded, competing representations framework can explain why intra-dimensional shifts are easier than extra-dimensional shifts, in terms of the nature of the changes required for each. Consider a network with a strong bias from sorting according to

color (Figure 2, right-hand side). The network strongly represents the redness of a red truck. For an intra-dimensional shift (red things now go in the blue tray), the network can build on these strong representations of the relevant features; only the connections from these representations to the output must be changed. Specifically, the network can continue to represent red trucks as red, and only needs to learn to put red things in the blue tray. In contrast, for an extra-dimensional shift (trucks now go in the truck tray), the network cannot build on its representation of color features; the connections to these representations from the input must be changed. Specifically, the network must learn to represent red trucks not as red but as trucks (and to put trucks in the truck tray). For neural networks, the intra-dimensional shift may be easier than the extra-dimensional shift, because neural network learning algorithms can work most effectively when representations are strong and stable (O'Reilly & Munakata, 2000; Rumelhart & McClelland, 1986). This is the case for the intra-dimensional shift (e.g., where the red truck is consistently represented as red), but not for the extra-dimensional shift (e.g., where the representation of the red truck must shift from red to truck). In children, these processes would correspond to finding it easier to change a response to a strongly represented aspect of a stimulus than to change how one views the stimulus.

#### *Target Cards Versus Puppets*

Children are better at switching to a second rule in a card-sorting task if the target cards are replaced by target puppets, with puppets first liking things according to one rule, and then liking things according to a different rule (Perner & Lang, 2002; Towse et al., 2000). Some researchers (Diamond, Carlson, & Beck, in press) claim that this finding challenges our graded memory representations account. They argue that removing target cards increases demands on remembering the correct rules in postswitch, and yet children's performance improves; therefore, memory problems alone cannot explain why children perseverate.

In contrast, we believe that memory accounts can accommodate these findings. The puppet manipulation could affect memory traces in multiple ways, not just the demands for remembering the rules in postswitch. For example, in our graded, competing representations framework, the puppet manipulation would affect the strength of the latent representations at the start of the task. All of the connections between internal representation and

output (Figure 2, left-hand side) would be weaker, because colors and shapes are less associated with puppets than with matching target cards. As a result, an active memory system's representation of the current game could play more of a role in guiding behavior, leading to better switching performance. In children, these processes would correspond to having stronger associations between matching colors than between colors and puppets; children's biases from preswitch would thus be weaker with the puppets, so that their active representations of a new game (e.g., shape) would be better able to compete against those biases.

#### *Alternative Theories*

In what follows, we consider several alternative theories of children's perseveration. Key next steps will be to establish how these accounts explain the current results, to assess the extent to which such explanations are post hoc versus clearly motivated by their theoretical frameworks, and to assess whether such explanations constitute alternatives to our account or different levels of description. We provide some initial speculations about how these theories may or may not accommodate the current findings as well as additional findings, and how such accounts relate to the graded, competing representations account that motivated these studies.

#### *Cognitive Complexity and Control*

Cognitive complexity and control theory (CCC) posits that perseveration results from an inability to consciously reflect on tasks and recognize higher order rules (Frye, Zelazo, & Palfai, 1995; Zelazo et al., 1996; Zelazo & Frye, 1998; Zelazo & Jacques, 1996). For example, to succeed in the card-sorting task, children must reflect on the task and create an if-then rule structure (e.g., if playing the shape game, if the card is a red truck, then it goes in the tray with the blue truck target). According to CCC, children first learn a basic rule (e.g., the card goes in the truck tray) around the age of 2.5 years. Then around age 3, children learn if-then rules (e.g., if the card is a truck, then it goes in the tray with the truck target, and if the card is a flower, then it goes in the tray with the flower target). Finally, at age 4 children learn to use if-if-then rules, which are the lower order if-then rules embedded under a higher order rule that selects the relevant game. Only then can children successfully switch from one rule to another.

CCC theory may have difficulty accommodating the current findings that 3-year-olds switch more easily from one rule to another when the first phase involves uninformative labels or novel stimuli. According to CCC, these children lack the requisite if–if–then rules to successfully switch (e.g., if it is the color game, and if the card is red, then it goes in the tray with the red target). However, each of the current tasks seems to require the same rules, yet children performed better in the uninformative-labels and novel-stimuli conditions than in the basic-PPC condition. Thus, children’s success does not appear to rely on the ability to embed lower order rules under a higher order rule.

Similar arguments have been made against CCC theory using the findings described earlier, regarding intra- versus extra-dimensional shifts and target cards versus puppets (Diamond et al., in press; Perner & Lang, 2002). Each of those variants of the task also seem to require the same if-if-then rule structure, but children are better at intra-dimensional shifts (Perner & Lang, 2002) and with target puppets (Perner & Lang, 2002; Towse et al., 2000). If children fail some card-sorting tasks because they lack the relevant if–if–then rule structures, how can they succeed on other card-sorting tasks that require the same structures?

In response to these kinds of criticisms, CCC theory has been revised, and CCC theorists have argued that children may succeed in some of these task variants without needing to construct higher order rules (Andrews, Halford, Bunch, Bowden, & Jones, 2003; Zelazo et al., 2003). For example, in the case of the target puppets instead of target cards, the pre-switch rules may have been less salient and easier to forget. As a result, children could more easily move on to new rules, without needing to coordinate these new rules with the old (forgotten) rules (Zelazo et al., 2003). The same argument might be applied to the current findings, with children being more likely to forget the preswitch rules when they involve novel stimuli or are described with uninformative labels. Alternatively, one might argue from the CCC framework that labels change children’s biases, so that they can shift from one sorting behavior to another without needing to construct higher order rules (Happaney & Zelazo, 2003); this change in bias might be greatest when the changes in types of labels used in preswitch versus postswitch are most salient, as in our uninformative-labels and novel-stimuli conditions. It is less clear that these kinds of arguments could explain why intra-dimensional shifts are easier than extra-dimensional shifts, since the preswitch rules are more comparable across these variants.

Moreover, one might ask how this salience factor would be implemented within CCC theory. And, once it is implemented, what unique contributions does CCC theory provide relative to other theories? For example, salience might be instantiated in terms of the strength of active or latent representations, as explored in the current account, to explain the findings above within CCC theory. Such factors might also be required for CCC theory to address other apparent challenges, such as the fact that 3-year-olds succeed in other versions of the card-sorting task (Kirkham et al., 2003; Yerys & Munakata, 2001), 4-year-olds fail in certain versions (Kirkham et al., 2003), and 6-year-olds fail other tasks requiring if–if–then rules (Morton & Munakata, 2002b; Morton et al., 2003). One might then ask whether such “additional” factors do the interesting work in terms of explaining behavior, and what independent contributions CCC theory per se provides.

#### *Attentional Inertia*

Alternatively, perseveration may result from an attentional inertia, in which there is a pull to continue attending to previously relevant information, and a difficulty inhibiting this pull when it becomes irrelevant (Kirkham et al., 2003). After repeatedly sorting a red truck according to its color, for example, children adopt the mind-set of seeing the stimulus as red and have great difficulty redirecting their attention to its shape.

Of the existing alternative accounts of perseveration, we view this account as the most compatible with our account. Our account of the strength of underlying representations might be viewed as providing the lower level mechanisms for this attentional inertia account. For example, latent representations that build up after sorting by color lead to the much greater activation of color information over shape information, even when the rule has changed to shape (right-hand side of Figure 2). The primary differences between the attentional inertia and active–latent accounts likely lie in the mechanisms posited to support eventual successful performance, as discussed later in this section.

The attentional inertia theory might accommodate the uninformative label findings by positing that there is less attentional inertia for stimuli under these conditions; hearing a stimulus referred to as “one” as opposed to “a red one” would reduce the tendency to see it as red. As a result, children would be better able to inhibit the pull to see the stimulus in terms of its color when the rule switches to shape. In this case, again, our account might be viewed as providing the

lower level mechanisms for this higher level attentional inertia account (Figure 3b). Weaker inputs lead to weaker active representations, which in turn lead to the formation of weaker latent representations biasing a stimulus to be viewed in a certain way.

However, the attentional inertia account might have predicted the opposite effect with novel stimuli to the one observed, because novelty generally serves to draw attention, which might increase rather than decrease attentional inertia. This would make it more difficult for children to inhibit the pull to see the stimulus in terms of the previously relevant dimension when the rule changed. Instead, children were better able to switch to a new sorting rule after sorting stimuli according to novel features. Alternatively, the attentional inertia account might accommodate the current findings with novel stimuli by positing that there is less attentional inertia for novel stimuli. In this case, again, our account might be viewed as providing the lower level mechanisms for this higher level attentional inertia account (Figure 3c). Yet another possibility is that novel stimuli draw attention, such that less inhibition of the other dimension is required during preswitch; as a result, children more easily switch to that other dimension during postswitch (A. Diamond, personal communication, May 2003; see Perner & Lang, 2002; Zelazo et al., 2003).

The attentional inertia and active-latent accounts may differ primarily in how they explain children's eventual successful switching. The attentional inertia account posits directed inhibitory processes that are distinct from memory processes. For example, a child might switch to sort cards according to their shape by directly inhibiting the tendency to see the cards in terms of their color. This would require using some form of what are essentially "not color" representations. In contrast, within the active-latent and related accounts (e.g., Cohen & Servan-Schreiber, 1992; Goldman-Rakic, 1987), there are no such "not color" representations that are specialized for inhibiting color representations. Instead, stronger active representations of new rules (e.g., shape) provide top-down support for relevant representations (e.g., truck); inhibition falls out of competitive dynamics throughout the system, so that the more support there is for the new rule, the less likely it is that an old rule (e.g., color) will win in this competition. We have documented elsewhere the arguments and evidence in favor of accounts based on top-down support and competitive dynamics over accounts based on directed inhibition (e.g., Munakata, Morton, & O'Reilly, in press; Munakata, Morton, & Stedron, 2003). To our knowledge, predictions

that would serve to distinguish these accounts in the domain of children's flexible card-sorting have not been tested, but we are in the process of doing so (Cepeda & Munakata, 2006; Kharitonova, Chien, Colunga, & Munakata, in preparation).

### *Miscategorization*

According to miscategorization theory, perseveration results from miscategorizing a new situation as an old one and running off the correct response for the old situation (Aguiar & Baillargeon, 2000; Baillargeon & Wang, 2002). Flexible behavior occurs if individuals engage in deeper processing of the situation to recognize it is new, and then determine the correct response rather than running off an old one. For example, in the card-sorting task, after children sort several cards successfully in the pre-switch phase, they may miscategorize the postswitch phase as the same game they were playing before, and run off the now-incorrect responses from that game. From this perspective, children might be more likely to switch to a new sorting rule if the task were altered to reduce their tendency to categorize the task in terms of the first rule, or to alert them more to the change at postswitch.

Miscategorization theory might accommodate the current findings by positing that both of these alterations were implemented in our tasks. First, children's tendency to categorize the task in terms of the first rule may have been reduced by using novel stimuli or uninformative labels, which may have decreased children's certainty about the task. Although children showed ceiling preswitch performance with these alterations, they may have nonetheless been less inclined to categorize the entire task according to this first phase when it involved novel stimuli or uninformative labels. As a result, they would be more likely to engage in deeper processing of the new situation during postswitch, resulting in more flexible behavior. Second, the change to postswitch rules might have been more salient in the uninformative-labels and novel-stimuli conditions relative to the basic-PPC condition. The verbal instructions changed from uninformative to informative or novel to familiar in the former cases, with no such changes in the latter case. As a result, children might have been more alerted to the change at postswitch in the uninformative-labels and novel-stimuli conditions, resulting in more flexible behavior.

Although miscategorization theory might accommodate the current findings in this way, it is not clear whether it would yield a priori predictions about which manipulations should be most effective in im-

proving children's performance. For example, if the current manipulations sufficed for reducing children's miscategorizations, one might predict that changing the previously attended features between preswitch and postswitch would similarly alert children to a change in task. However, this is the "negative priming" version of the task described earlier, in which children still perseverate (Zelazo et al., 2003). Similarly, from the miscategorization perspective, it is not clear why children are better at intra-dimensional switches than at extra-dimensional ones (Perner & Lang, 2002). If anything, extra-dimensional changes should be more salient (and thus more effective in alerting children to a change in task), because they involve the introduction of new terms for describing the relevant dimension and features.

Miscategorization theory can readily accommodate the finding that children generally do much better at switching to new rules in the card-sorting task if they receive negative feedback when they sort cards incorrectly (Bohlmann & Fenson, 2005; Yerys & Munakata, 2001). Such negative feedback would serve to alert children to the fact that the task has changed and deeper processing is required. However, it is less clear how miscategorization theory would explain the fact that many children continue to perseverate in the face of repeated negative feedback, both in the card-sorting task (Yerys & Munakata, 2001) and in searching for hidden toys (Piaget, 1954). Accounting for these behaviors, and other challenges described above, may require factors such as the strength of competing representations considered here.

### Conclusion

This work adds to a growing body of data demonstrating that perseveration is not simply a matter of being unable to inhibit whatever behaviors were previously carried out. The specifics of how those previous behaviors were carried out, what one brings to a task, and the context in which one is asked to switch to a new behavior are all critical in whether one perseverates or responds flexibly to a task (e.g., Aguiar & Baillargeon, 2000; Diamond, Kirkham, & Amso, 2002; Kirkham et al., 2003; Morton et al., 2003; Perner & Lang, 2002; Smith, Thelen, Titzer, & McLin, 1999; Zelazo et al., 2003). Rather than comprising uninteresting details of a simple behavior, these specifics are likely to be critical in informing and evaluating a variety of theories of perseveration, flexibility, and developmental processes that subservise changes in this domain.

In the current case, children's improvements in switching after sorting cards with uninformative labels or novel stimuli are consistent with a graded, competing representations framework. These findings raise questions for other accounts, such as whether they can accommodate these findings in their current form. If so, are the explanations of the current findings consistent with findings from other studies, or do they take the form of post hoc accounts formulated specifically for these data? If additional factors must be incorporated into other accounts to explain the details of perseverative behavior, what additional work is the basic theory doing? And, whatever factors in other accounts might explain the current findings, do these factors constitute true alternatives to the account described here, or different levels of description?

Further work is needed to relate theories of perseveration at different levels of description and to investigate the specifics of these behaviors. Such work should ultimately provide a better understanding of perseveration and flexibility, as well as the nature, organization, and development of representations more generally.

### 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.
- Andrews, G., Halford, G. S., Bunch, K. M., Bowden, D., & Jones, T. (2003). Theory of mind and relational complexity. *Child Development, 74*, 1476–1499.
- Baillargeon, R., & Wang, S. (2002). Event categorization in infancy. *Trends in Cognitive Sciences, 6*, 85–93.
- Bohlmann, N. L., & Fenson, L. (2005). The effects of feedback on perseverative errors in preschool aged children. *Journal of Cognition and Development, 6*, 119–131.
- Brooks, P. J., Hanauer, J. B., Padowska, B., & Rosman, H. (2003). The role of selective attention in preschoolers' rule use in a novel dimensional card sort. *Cognitive Development, 18*, 195–215.
- Butterworth, G. (1977). Object disappearance and error in Piaget's stage IV task. *Journal of Experimental Child Psychology, 23*, 391–401.
- Casey, B. J. (2001). Disruption of inhibitory control in developmental disorders: A mechanistic model of implicated frontostriatal circuitry. In J. L. McClelland & R. S. Siegler (Eds.), *Mechanisms of cognitive development: Behavioral and neural perspectives. Carnegie Mellon symposium on cognition* (pp. 327–349). Mahwah, NJ: Erlbaum Associates Publishers.
- Cepeda, N. J., & Munakata, Y. (2006). *Why do people perseverate when they seem to know better: Graded working*

- memory, or directed inhibition? Manuscript submitted for publication.
- Chambers, D., & Reisberg, D. (1985). Can mental images be ambiguous? *Journal of Experimental Psychology: Human Perception and Performance*, 3, 317–328.
- Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing model of the Stroop effect. *Psychological Review*, 97, 332–361.
- 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.
- Dalrymple-Alford, E. C., & Budayr, B. (1966). Examination of some aspects of the stroop color-word test. *Perceptual and Motor Skills*, 23, 1211–1214.
- Deak, G. O., & Bauer, P. J. (1996). The dynamics of preschoolers' categorization choices. *Child Development*, 67, 740–767.
- Dehaene, S., & Changeux, J. P. (1989). A simple model of prefrontal cortex function in delayed-response tasks. *Journal of Cognitive Neuroscience*, 1, 244–261.
- Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by infants' performance on A. *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. (2002). A model system for studying the role of dopamine in prefrontal cortex during early development in humans. In M. H. Johnson, Y. Munakata, & R. O. Gilmore (Eds.), *Brain development and cognition: A reader*. Chapter 22 (pp. 441–493). Oxford, UK: Blackwell.
- Diamond, A., Carlson, S. M., & Beck, D. M. (2005). Preschool children's performance in task switching on the dimensional change card sort task: Separating the dimensions aids the ability to switch. *Developmental Neuropsychology*, 28(2), 689–729.
- Diamond, A., & Goldman-Rakic, P. S. (1989). Comparison of human infants and rhesus monkeys on Piaget's A task: Evidence for dependence on dorsolateral prefrontal cortex. *Experimental Brain Research*, 74, 24–40.
- Diamond, A., & Kirkham, N. (2005). Not quite as grown-up as we like to think: Parallels between cognition in childhood and adulthood. *Psychological Science*, 16, 291–297.
- Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology*, 38, 352–362.
- Dias, R., Robbins, T. W., & Roberts, A. C. (1997). Dissociable forms of inhibitory control within prefrontal cortex with an analog of the Wisconsin Card Sort Test: Restriction to novel situations and independence from "on-line" processing. *Journal of Neuroscience*, 17, 9285–9297.
- Diedrich, F., Thelen, E., Smith, L., & Corbetta, D. (2000). Motor memory is a factor in infant perseverative errors. *Developmental Science*, 3, 479–494.
- Dunbar, K., & Sussman, D. (1995). Toward a cognitive account of frontal lobe function: Simulating frontal lobe deficits in normal subjects. *Annals of the New York Academy of Sciences*, 769, 289–304.
- Farah, M. J., Monheit, M. A., & Wallace, M. A. (1991). Unconscious perception of "extinguished" visual stimuli: Reassessing the evidence. *Neuropsychologia*, 29, 949–958.
- 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.
- Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between the frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, and Behavioral Neuroscience*, 1, 137–160.
- Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, 10, 483–527.
- Gentner, D., & Loewenstein, J. (2001). Relational language and relational thought. In J. Byrnes & E. Amsel (Eds.), *Language, literacy, and cognitive development* (pp. 87–120). Mahwah, NJ: Lawrence Erlbaum Associates.
- 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.
- Griffith, E. M., Pennington, B. F., Wehner, Elizabeth A., & Rogers, S. J. (1999). Executive functions in young children with autism. *Child Development*, 70, 817–832.
- Happaney, K., & Zelazo, P. D. (2003). Inhibition as a problem in the psychology of behavior. *Developmental Science*, 6, 468–470.
- Hauser, M. D. (1999). Perseveration, inhibition and the prefrontal cortex: A new look. *Current Opinion in Neurobiology*, 9, 214.
- Heibeck, T., & Markman, E. M. (1987). Word learning in children: An examination of fast mapping. *Child Development*, 1987, 1021–1034.
- Hofstadter, M. C., & Reznick, J. S. (1996). Response modality affects human infant delayed-response performance. *Child Development*, 67, 646–658.
- Hood, B. M., Hauser, M., Anderson, L., & Santos, L. (1999). Gravity biases in a non-human primate. *Developmental Science*, 2, 35–41.
- Jacques, S., & Zelazo, P. D. (2001a). The flexible item selection task (FIST): A measure of executive function in preschoolers. *Developmental Neuropsychology*, 20, 573–591.
- Jacques, S., & Zelazo, P. D. (2001b). *The role of language in the development of cognitive flexibility*. Talk presented at the Biennial Meeting of the Society for Research in Child Development, Minneapolis, MN.
- Jacques, S., & Zelazo, P. D. (2003). Language and the development of cognitive flexibility: Implications for theory of mind. In J. Astington & J. Baird (Eds.), *Why language matters for theory of mind* (pp. 573–591). New York: Oxford University Press.

- Jacques, S., Zelazo, P. D., Kirkham, N. Z., & Semcesen, T. K. (1999). Rule selection versus rule execution in preschoolers: An error-detection approach. *Developmental Psychology, 35*, 770–780.
- Kersten, A. W., & Smith, L. B. (2002). Attention to novel objects during verb learning. *Child Development, 73*, 93–109.
- Kharitonova, M., Chien, S., Colunga, E., & Munakata, Y. (in preparation). The emergence of human intelligence: An early synergy of flexibility and abstraction
- Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Developmental Science, 6*, 449–476.
- Liittschwager, J. C., & Markman, E. M. (1994). Sixteen- and 24-month-olds' use of mutual exclusivity as a default assumption in second-label learning. *Developmental Psychology, 30*, 955–968.
- MacDonald, P. A., & Joordens, S. (2000). Investigation of a memory-based account of negative priming: Support for selection-feature mismatch. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1478–1496.
- MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. *The Psychology of Learning and Motivation, 43*, 163–214.
- MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology, 14*, 126–135.
- Marcovitch, S., & Zelazo, P. D. (1999). The A-not-B error: Results from a logistic meta-analysis. *Child Development, 70*, 1297–1313.
- Markson, I., & Bloom, P. (1997). Evidence against a dedicated system for word learning in children. *Nature, 385*, 813.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167–202.
- Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science, 263*, 520–522.
- Milner, B. (1963). Effects of different brain lesions on card sorting. *Archives of Neurology, 9*, 90–100.
- Morton, J. B., & Munakata, Y. (2002a). Active versus latent representations: A neural network model of perseveration and dissociation in early childhood. *Developmental Psychobiology, 40*, 255–265.
- Morton, J. B., & Munakata, Y. (2002b). Are you listening? Exploring a knowledge action dissociation in a speech interpretation task. *Developmental Science, 5*, 435–440.
- 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. (2003). Representational inflexibility in children's interpretation of emotion in speech. *Child Development, 74*, 1857–1868.
- Munakata, Y. (1997). Perseverative reaching in infancy: The roles of hidden toys and motor history in the task. *Infant Behavior and Development, 20*, 405–416.
- Munakata, Y. (1998). Infant perseveration and implications for object permanence theories: A PDP model of the task. *Developmental Science, 1*, 161–184.
- Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences, 5*, 309–315.
- Munakata, Y., McClelland, J. L., Johnson, M. J., & Siegler, R. S. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review, 104*, 686–713.
- Munakata, Y., Morton, J. B., & O'Reilly, R. C. (in press). Developmental and computational approaches to variation in working memory. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in working memory*. New York: Oxford University Press.
- Munakata, Y., Morton, J. B., & Stedron, J. M. (2003). The role of prefrontal cortex in perseveration: Developmental and computational explorations. In P. Quinlan (Ed.), *Connectionist models of development* (pp. 83–114). East Sussex: Psychology Press.
- Munakata, Y., & Yerys, B. E. (2001). All together now: When dissociations between knowledge and action disappear. *Psychological Science, 12*, 335–337.
- Neill, W. T. (1977). Inhibitory and facilitatory process in selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 3*, 444–450.
- Neill, W. T., & Valdes, L. A. (1992). The persistence of negative priming: Steady-state or decay? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 565–576.
- O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.
- Park, J., & Kanwisher, N. (1994). Negative priming for spatial locations: Identity mismatching, not distractor inhibition. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 613–623.
- Perner, J., & Lang, B. (2002). What causes 3-year-olds' difficulty on the dimensional change card sorting task? *Infant and Child Development, 11*, 93–105.
- Piaget, J. (1954). *The construction of reality in the child*. New York: Basic Books.
- Roberts, R. J., & Pennington, B. F. (1996). An interactive framework for examining prefrontal cognitive processes. *Developmental Neuropsychology, 12*, 105–126.
- Ross, H. S. (1974). The influence of novelty and complexity on exploratory behavior in 12-month old infants. *Journal of Experimental Child Psychology, 17*, 436–451.
- Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition (Vols. 1 and 2)*. Cambridge, MA: MIT Press.
- Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*, 36–71.
- Smith, L. B., Thelen, E., Titzer, B., & McLin, D. (1999). Knowing in the context of acting: The task dynamics of the A-not-B error. *Psychological Review, 106*, 235–260.

- Spencer, J. P., & Schutte, A. R. (2004). Unifying representations and responses: Perseverative biases arise from a single behavioral system. *Psychological Science*, *15*, 187–193.
- Stedron, J., Sahni, S. D., & Munakata, Y. (2005). Common mechanisms for working memory and attention: The case of perseveration with visible solutions. *Journal of Cognitive Neuroscience*, *17*, 623–631.
- 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.
- 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.
- Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.
- Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *The Quarterly Journal of Experimental Psychology*, *37A*, 571–590.
- Tipper, S. P. (2001). Does negative priming reflect inhibitory mechanisms? a review and integration of conflicting views. *Quarterly Journal of Experimental Psychology*, *54A*, 321–343.
- Towse, J. N., Redbond, J., & Cook, S. (2000). Understanding the dimensional change card sort: Perspectives from task success and failure. *Cognitive Development*, *15*, 347–365.
- 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).
- Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology*, *30*, 553–566.
- Yerys, B. E., & Munakata, Y. (2001). *More flexible than you think: Feedback improves children's switching in a card-sorting task*. Poster presented at the biennial meeting of the Cognitive Development Society, Virginia Beach.
- Zelazo, P. D., & Frye, D. (1998). Cognitive complexity and control: II. The development of executive function in childhood. *Current Directions in Psychological Science*, *7*, 121–126.
- Zelazo, P. D., Frye, D., & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, *11*, 37–63.
- Zelazo, P. D., & Jacques, S. (1996). Children's rule use: Representation, reflection, and cognitive control. In R. Vasta (Ed.), *Annals of child development* (pp. 119–176). London: Jessica Kingsley.
- Zelazo, P. D., Mueller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development*, *68*, 1–137.
- Zelazo, P. D., Reznick, J. S., & Spinazzola, J. (1998). Representational flexibility and response control in a multistep multilocation search task. *Developmental Psychology*, *34*, 203–214.

## ENDNOTES

<sup>1</sup>Manipulating both preswitch and postswitch simultaneously would likely be less informative (e.g., basic stimuli in preswitch and postswitch vs. novel stimuli in preswitch and postswitch). The effects of the manipulations in preswitch and postswitch might cancel one another, so that switching performance would be similar across these conditions despite differences in the underlying dynamics (see Munakata 1997 for a similar point in the context of infants' searching for hidden objects). And if differences were observed across conditions with both preswitch and postswitch manipulated, it might be difficult to isolate the source of such differences.

<sup>2</sup>To ensure that children in the uninformative-labels condition committed to a specific rule during preswitch (as opposed to sorting based on overall similarity), we tested a separate group of children with the same preswitch procedure, followed by a new phase to test whether they were using the intended preswitch rule. In this new phase, conflict cards were introduced, but instructions stayed the same as in preswitch ("These go here, these go there," with the experimenter waving partially matching cards). Eight normally developing 3-year-old children (mean age = 35 months 28 days; range = 35 months 25 days to 36 months 4 days, 4 males and 4 females) participated in the experiment and were included in the final sample. All 8 children passed the preswitch phase and then passed the new phase, sorting at least four of the six conflict cards consistently with how they sorted the preswitch cards. These results demonstrate that the uninformative-labels preswitch procedure led children to commit to a specific rule. We thank Sophie Jacques for raising this issue and suggesting this experiment.