Two routes to cognitive flexibility:
Learning and response conflict resolution in the dimensional change card sort task

Michael Ramscar
Department of Psychology
Stanford University

Melody Dye
Department of Psychology
Stanford University

Jessica Witten
Department of Psychology
Stanford University

Joseph Klein
Department of Psychology
Stanford University

Contact information:

Michael Ramscar
Department of Psychology
Stanford
CA 94305

Email: ramscar@stanford.edu
Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Nos. 0547775 and 0624345 to Michael Ramscar. We are grateful to Natasha Kirkham and Sharon Thompson-Schill for discussion of these ideas.
Abstract

There are at least two ways in which response conflict can be handled in the mind: *dynamically*—processing and resolving conflicting responses on-line, and *through discrimination learning*—reducing the amount of response conflict in advance. While children under-four are perfectly capable of discrimination learning, they appear to lack the ability to resolve response conflict on-line. This handicap is frequently demonstrated by their failure in the dimensional change card sort task. Here, we present an analysis of how contextual learning aids children’s performance in the DCCS. We find that under-fours given appropriate training are surprisingly adept at switching between the responses required and discuss the implications for our understanding of prefrontal development and how children learn to use language contextually.
Two routes to cognitive flexibility:

Learning and response conflict resolution in the dimensional change card sort task

Although a three-year-old girl might appear to be just a slightly smaller version of her four-year-old brother, experimenters have revealed numerous differences between them. While her big brother will sail effortlessly through the many tasks devised to test the capabilities of the very young, she will likely fail every one of them. A typical 4-year-old can successfully switch responses and probability match in binary choice tasks, understand false belief and the conflicting dimensions of appearance and reality, and switch easily between competing rules in dimensional change card sorting (DCCS) task. A typical three-year-old, on the other hand, will maximize in binary choice tasks (fixating on the most likely response), fail false belief tasks, be unable to switch from describing the appearance of an object to answering questions about what it really is, and fail to switch from one sorting rule to another in the DCCS, even when the rule is clearly stated (for a review see Ramscar & Gitcho, 2007). This poses both a question—why do children under four fail these tasks?—and a puzzle—given the inflexibility of thought that these various tests reveal, why in the normal course of events is this inflexibility so hard to detect? Why is it that young children appear to be perfectly capable of switching responses and matching their behavior to context once they are outside the lab?

Many proposals have been advanced to explain why young children fail these tasks (e.g., Zelazo, Müller, Frye & Marcovitch, 2003). In what follows, we address both the question and the puzzle by considering the different ways in which the conflict between the potential responses in a task might be resolved, so that an appropriate response can be given in context. We propose that there are at least two ways in which response conflict can be handled in the mind: dynamic response conflict resolution, which enables conflicting response demands to be processed and resolved on-line, and discrimination learning, which shapes the strengths by which responses are evoked by specific contexts, reducing the amount of on-line response conflict that needs to be processed and resolved. We suggest that under-fours are fully capable of discrimination learning, but lack the ability to resolve response conflict on-line. Under-fours are able to match their behavior to context in remarkably subtle and sensitive ways when they have learned to do so. However, if they have not learned to match a response or a behavior to a context, under-fours’ inability to handle on-line response conflict proves their undoing in psychological tests.
At around the age of four, children’s performance on a range of laboratory tests begins to change dramatically. Many of these changes are characterized by a shift from responses dominated by a single factor to more complex behavior that integrates or selects between multiple responses. These changes are consistent with the idea that children develop the cognitive machinery that allows them to monitor and filter their responses in a manner akin to adults over the course of early childhood (Ramscar & Gitcho, 2007). Exactly how and why these developments occur when they do remains an open question. However, converging evidence from behavioral studies, developmental neurobiology and cognitive neuroscience models supports the idea of a domain-general processing shift in early childhood, with notable changes occurring in the fourth year (for a review see Thompson-Schill, Ramscar & Evangelia, in press).

In what follows, we describe the neurological and computational bases for this proposal, and present a computational simulation of how discrimination learning and context might affect children’s performance in the dimensional change card sort task (DCCS; Zelazo, 2006) over time. The model suggests that the observed failure of under fours at the DCCS stems from a lack of discrimination learning in the context of the “games” children play in the task, and predicts that if children are exposed to the game contexts in ways that promote discrimination learning, they should then succeed at the task with relative ease. We then present a training study in which three groups of age matched under-fours attempt to complete the DCCS after i) exposure to games that promote discrimination learning, ii) exposure to games that do not promote discrimination learning, and iii) after no prior exposure. Consistent with the model’s predictions, training that promoted discrimination learning enabled children to pass the DCCS with flying colors. Children fared far worse after training that did not promote discrimination learning, and failed as expected when they were given no prior training.

(Figure 1 about here.)

The Dimensional Change Card Sort Task

In the Dimensional Change Card Sort (DCCS) Task, three and four year-old children sort cards with two prominent dimensions—a color and shape—into bins in which these dimensions have been reversed. For example, if a child is holding cards with red stars and blue trucks, the bins will be marked with blue stars and red trucks. If the child is asked to sort by color, the red
stars will go with the red trucks and the blue stars will go with the blue trucks; if the child is asked to sort by shape, the red stars will go with the blue stars, and the red trucks will go with the blue trucks. Importantly, when a child is asked to sort by one dimension—say, shape—switching the sort dimension to color will also switch the correct sort bin; e.g., red stars match to the truck bin when sorted by color, but the star bin when sorted by shape. This is a straightforward task for older children and adults, but not for children under the age of four.

In general, when young children are asked to begin sorting by shape, they can answer questions regarding the rules for correctly sorting either by shape or by color. In addition, after switching from sorting by shape to sorting by color, children can correctly answer questions about how to correctly sort according to the new rule. However, once children are actually handed a card and asked to sort according to the second rule they have learned, their success in the task varies markedly with age. For the most part, 3-year-old children are unsuccessful at this part of the task; they continue to sort the cards according to the first rule (whether it be sorting by shape or color). Around age 4, however, children usually begin to pass the DCCS task without difficulty, successfully matching the cards to the bins before and after the sorting rules are switched (Zelazo, Frye & Rapus, 1996).

Why do three year olds fail this task? We suggest that their poor performance is related to the late development of prefrontal cortex, which appears to be an essential part of the circuit that deals with the demands of on-line response conflict processing in humans (Yeung, Botvinick, & Cohen, 2004). Like many other primates, humans are born with an immature brain. Birth is followed by synaptogensisis (the proliferation of synapses) followed by an extended pruning period (synaptic elimination). Brain development in humans, however, is markedly different from that of other primates. In monkeys, post-natal brain development occurs at a similar rate in all cortical areas (Rakic, Bourgeois, Eckenhoff, Zecevic, & Goldman-Rakic, 1986). In contrast, human cortical development is uneven: synaptogenesis in visual and auditory cortex peaks a few months after birth, while the same developments occur later in prefrontal cortex (Huttenlocher & Dabholkar, 1997; see Thompson-Schill, et al., in press, for a review).

One behavioral consequence of slow prefrontal development is that young children appear unable to engage in behaviors that conflict with prepotent responses, and appear unable to filter their behavior and their learning (Ramscar & Gitcho, 2007; Thompson-Schill et al, in press). The adult ability to select a less well learned, but goal appropriate response is seen in the
Stroop Task, in which subjects have to identify the ink color of a conflicting color word (e.g., if “green” is printed in red ink, red must be identified). Performance in this task involves resolving the conflict between a well-learned, prepotent response (reading) and a contextually appropriate response (ink naming). Adults typically complete the Stroop Task with ease, but young children fail structurally similar tasks, such as the DCCS. In adults, success is facilitated by prefrontal control mechanisms that bias one response over another according to goals or context (Yeung, Botvinick, & Cohen, 2004). This is consistent with a range of evidence suggesting that the prefrontal cortex functions as a dynamic filter, selectively maintaining task-relevant information and discarding task-irrelevant information (Shimamura, 2000).

If under-fours lack (or are deficient in) the ability to dynamically filter responses in accordance with the demands of a context or goal, this may explain why they fail to switch rules in the DCCS. If a card depicts a red star, the color dimension elicits one response (sorting into the color bin) whereas the shape dimension elicits a conflicting response (sorting into the shape bin). Thus in the standard DCCS task, successfully switching rules involves changing from one response associated with a stimulus—a card depicting a colored shape—to an alternative, conflicting response. Since this kind of response conflict processing appears to be the preserve of the frontal areas of the brain (Yeung, Botvinick, & Cohen, 2004), it is possible that the failure of under-fours in the DCCS task—that is, their failure to mediate response conflict—is related to slow prefrontal development.

**Discrimination Learning**

While young children may lack the ability to resolve conflict on-line, discrimination learning may provide another means by which they might still learn to succeed in the DCCS. This is because the “games” associated with each sorting rule provide cues to the appropriate responses, in addition to the cues provided by the shapes and colors on the cards themselves. The “shape game” is an additional cue to the response “sort into the shape bin” and the “color game” is an additional cue to the response “sort into the color bin.” However, since children fail the task despite the presence of these cues, it is clear that in ordinary testing circumstances, the game cues do not provide sufficient extra scaffolding to enable three year olds to pass the DCCS.

To explain why these additional cues might still matter, we need to consider the way that responses that lead to response conflict in the DCCS are learned and discriminated.
Discrimination learning is a process by which information is acquired about the probabilistic relationships between important regularities in the environment (such as objects or events) and the cues that allow them to be predicted (see e.g., Rescorla & Wagner, 1972).

Crucially, the learning process is driven by discrepancies between what is expected and what is actually observed in experience (termed “error-driven” learning). The learned predictive value of cues produces expectations, and any difference between the value of what is expected and what is observed produces further learning. The predictive value associated with cues is strengthened when relevant events (such as events, objects or labels) are under-predicted, and weakened when they are over-predicted (Kamin, 1969; Rescorla & Wagner, 1972). As a result, cues compete for relevance, and the outcome of this competition is shaped both by positive evidence about co-occurrences between cues and predicted events, and negative evidence about non-occurrences of predicted events. This produces patterns of learning that are very different from those that would be expected if learning were shaped by positive evidence alone (a common portrayal of learning): learners discover the predictive structure of the environment, and not just simple patterns of correlations within it.

In considering how discrimination learning might impact DCCS performance, it is worth noting that although young children fail the DCCS, they can often reliably name the dimensions of the cards, even after they have failed to correctly sort by them (Kirkham, Cruess & Diamond, 2003). This finding need not be surprising given that children have far more experience with naming than with sorting. Children regularly hear objects referred to in terms of their shape and color, and as a result, they are likely to have done a great deal of learning about naming these attributes, enabling them to use linguistic contexts to figure out that shape-words go with shape questions and color-words go with color questions. As a result of error-driven learning, repeated exposure to distinctive contexts—such as the particular frame of a question—should diminish competing activations, thereby reducing the degree to which any other responses compete for relevance with a correct response. Over time this kind of contextual discrimination learning should, for example, diminish the degree to which color-words suggest themselves as possible responses to shape-naming questions, enabling children to use context to name the appropriate dimensions of the cards (see Figure 2). If this account is accurate—that experience with naming leads to discrimination learning and diminished response conflict—then we should expect that experience with sorting in the DCCS should produce a similar result.
What might discrimination learning look like in a child playing a series of sorting games? Imagine a child handed a red star card and asked to sort it. No matter which rule she is given—whether to sort by color or by shape—initially, both red AND star will appear as relevant cues in a given sort. If the cards cause a child to expect both dimensions to be relevant, there will be a violation of expectation when only one dimension is used in sorting (i.e., when the inappropriate dimension does not match the sort, such as the shape dimension when a red star is matched with a red truck in a color sort). Given that the child incorrectly predicted the (irrelevant) dimension would match the sort on that trial, she will begin to adjust her expectations accordingly: weakening that dimension’s predictive value and strengthening the predictive value of the dimension that did match the sort. When later the game changes and the child is asked to sort by the other dimension, this will cause problems at the outset since the earlier sort trials will have taught her to ignore the now-relevant dimension.

For example, in the color game the red star card is sorted by the color dimension. If the red dimension cues “correct sorting” as matching to color, while star cues “correct sorting” as matching to shape, then over trials in the color game the value of the association between red and “correct sorting” will increase while the value of the association between star and “correct sorting” will decrease (“shape” will be learned about even though it is not used in sorting). Importantly, because the context color game has been introduced, in subsequent color game trials, a conjunctive cue red + color game (e.g., Gluck & Bower, 1988) can compete with red, star (and color game) for associativity to the response of sorting by “color.”

The converse process will occur when the child moves on to the shape game, as the child learns to value star and devalue red in sorting the cards. At the outset, the child will likely make sort errors, since she has spent the previous game learning red to be the most effective cue for sorting cards and must now unlearn that in favor of star. However, over time she will learn that the red cue and the star cues individually prove far less useful—that is, predictive—of correct sorts than the conjunctive cues red + color game or star + shape game. She will thus learn that
which cue is relevant depends on which game is being played, and further, that the conjunction of the relevant cue with its corresponding game is highly predictive of the correct sort.

To formally test these ideas, we simulated the competition between conjunctive cues representing color game + red and shape game + star and the individual cues red and star across repeated DCCS trials using the Rescorla & Wagner (1972) model. In this model, the change in associative strength between a cue i and a response j on trial n is defined as:

$$\Delta V_{ij}^n = \alpha_i \beta_j (\lambda_j - V_{total})$$

where $\Delta V_{ij}$ is the change in associative strength between a set of cues i and response j on a given trial n. $\beta$ is a learning rate parameter for response j, bounded between 0 and 1. $\alpha$ is the saliency of cue i also bounded between 0 and 1. $\lambda_j$ denotes the maximum amount of associative value (total cue value) that response j can support, and $V_{TOTAL}$ is simply the sum of the associative values present on a given trial; i.e., the sum of the current cue values. If there is a discrepancy between $\lambda_j$ (the total possible associative value of an event) and $V_{TOTAL}$ (the sum of current cue values), the saliency of the set of cues $\alpha$ and the learning rate of the event $\beta$ will be multiplied against that discrepancy. The resulting amount will then be added or subtracted from the associative strength of any cues present on that trial. In the simulations, $\lambda_j = 100\%$ (when j was present in a trial) or 0\% (when j was not present in a trial), and $\alpha_i$=0.2 and $\beta_j$=0.3.

The simulation assumes that the appropriate sorting response is always available (i.e., learning is determined by what ought to happen), and that the individual dimensions have been previously learned about as cues (this is represented in Figure 3 as the first 40 trials). We assume further (1) that the individual cues will each predict the correct sort (i.e., red will predict a sort in which color is matched) on half the trials, and (2) that the conjunctive cues will always predict the correct sort (color game + red will also predict a sort in which color is matched, and this conjunctive cue will not be present on trial in which shape is matched). Thus the model was initially trained with color and shape as cues to alternate sorting events, and then the color game was introduced (trial 41). From then on the color game was present on all color sorts, and the shape game (appearing in trial 61) was present on all shape sorts.

In the first DCCS game shown in Figure 3, red and the conjunctive cues color game + red gain in associative value while the value of the star cue diminishes. As a result, the model predicts that, initially, red will be the strongest cue to correct sorting (predicting that color should be matched in sorting), even after a rule switch. (Red is initially stronger as a result of
children’s greater prior experience of it, as compared to the conjunctive cue.) If we assume that correct sorting is modeled however, then red will begin to lose cue value after the rule switch, as it will now fail to predict the correct sort. Because of this, even though all of the competing cues co-occur with correct sorts with exactly the same frequency, because the individual cues also co-occur with incorrect sorts, whereas the conjunctive cues do not, over time cue competition will cause the individual cue red to be effectively dissociated from correct sorting in this situation. The error generated by red in shape trials (red will cue a color sort) results in gradual shift of associative value from the individual cue red to the conjunctive cue color game + red, which does not lose value in this way because it is not present on shape trials. As long as the cards are correctly labeled in each context, a child will learn to ignore the unreliable cues, thereby improving response discrimination. This pattern of learning is clearly evident in the simulation (Figure 4), where the erroneous expectations producing red and star cause them to lose out in competition with the conjunctive cues that embody the games as contexts. Because of cue-competition, predictive power, not frequency or simple probability, determines cue value in this context.

(Figure 3 about here.)

(Figure 4 about here.)

**Discrimination Learning, Context, and the DCCS**

The analysis above describes how contextual learning might reduce response conflict in the DCCS. If children are trained to associate sorting by shape with a “shape game” and sorting by color with a “color game” they will eliminate the response-conflict normally associated with the DCCS by learning context-dependent rules. For example, they might learn: “red star card + color game = sort by red,” meaning that the conjunction of the color red and the color-game context will be a perfect predictor of the correct sort location for “red” cards in the color game, and this conjunction will not be present—and will thus not result in error—on shape game trials). However, the simulation also assumes that children’s learning is driven by their awareness of the correct response throughout. Clearly, in sorting, this isn’t usually the case. Young children don’t
switch sort dimensions appropriately, and so it is clear that the process of learning and unlearning will not (and does not) work in life the way it does in the simulation.

However, there is considerable support for generalization in learning about cues (Shepard, 1987). If children were to learn a particular configuration of cues in learning a response to a game context, they might be able to transfer their learning about that context to another response embedded in the same kind of game. This suggests a way in which children’s learning in the DCCS might be successfully enhanced: by providing them with the experience of learning about a “color shape card + color game” context (or “color shape card + shape game” context) while performing a task at which we expect them to succeed.

For example, we know that children can name the appropriate dimensions of the cards in the DCCS before they can sort them. If children were taught to associate naming the appropriate dimensions with the game rules, learning about the high predictive value of certain cue configurations in the game contexts (i.e., learning to associate the game and the game-appropriate dimension with the correct response) might transfer to sorting responses in the same context. If children were to attend to the conjunctive cues they learned playing naming games when playing sorting games, this should reduce response conflict, and enable them to successfully switch sorts in the DCCS task.

Learning with and without Cue Competition

It is important to note that error-driven learning (modeled above) is competitive: the various cues to correct sorts compete for relevance, losing associativity to other, more reliable cues and gaining associativity from other, less reliable cues. In the face of repeated prediction error, it is thus possible to unlearn a cue, even if it is highly correlated with a response. Cue competition allows a learner to improve response discrimination by learning the set of cues that most reliably predicts a given outcome. To illustrate the importance of cue competition to discrimination learning, it is useful to consider a picture of learning in the absence of cue competition.

So far in our discussion of learning in the DCCS, we have assumed that the perceptual features of the cards and the game contexts serve to cue correct sorts. We also suggested that children might learn to name the dimensions of shape and color on the card in a similar way by using the perceptual features of the card and linguistic contexts as cues to appropriate labels,
such as “red” or “star” (we call this Feature to Label-FL-learning; Ramscar et al., 2009). When the cards and contexts serve as cues, FL-learning promotes cue-competition, leading to the learning of the predictive value of the cues associated with the cards (Figure 2).

We can contrast this with what might happen if symbolic labels were to serve as a cue to the perceptual features of the cards in learning (Label to Feature-LF-learning). While these two scenarios appear similar, simply reversing the predictive (temporal) order of labels and features can have a dramatic impact on cue competition (for an extended discussion of this effect, see Ramscar et al., 2009). This is because while FL-learning involves predicting an individual label from a complex, multi-dimensional set of cues (the cards and contexts), LF-learning involves just the opposite: predicting a complex, multi-dimensional perceptual item (the card) from an individual cue (symbolic label). In LF-learning, each label predicts a complex outcome, but potentially competing labels never co-occur temporally in a way that would allow them to compete as cues. Because the effects of cue competition become attenuated as the temporal relations between cues vary (Amundson & Miller, 2008), labels cannot compete directly with one another for associative value. As a result, if a label incorrectly predicts an outcome in one context there will be no loss of potential associative value to another label acting as a cue in another context (that is, there can be no cue competition, Ramscar et al., 2009; see Figure 5).

(Figure 5 about here.)

Where there is no cue competition, cue values simply increase when a predicted outcome appears following the cue, and decrease when a predicted outcome fails to appear following the cue. Thus if a symbolic label serves to predict the dimensions of the cards, the cue value of that label will simply track the frequency with which labels and card features co-occur, approximating the conditional probability of a feature given a label (Ramscar et al., 2009; Cheng, 1997; Wasserman, Elek, Chatlosh, & Baker, 1993; treating this kind of correlation learning as the full extent of associative learning is a common misunderstanding, Rescorla, 1988).

To illustrate the difference this makes to discrimination learning, figure 6 depicts an extension of our earlier FL-learning simulation in which naming games alternate between shape and color naming, alongside a simulation of the same games learned LF. As can be seen,
whereas FL-learning results in the unlearning of the individual dimensional cues (red and star), given LF-learning, the associative values between the naming rules and these features continue to rise and fall in relation to the degree to which they are unlearned in one game context, and relearned in the other. Critically, as can be seen in Figure 6, the value of each individual dimension reaches its peak at the point at which a child will be required to switch away from it in the DCCS (e.g. “red” peaks in trial 140, just prior to the beginning of a shape naming game in trial 141; figure 6, right panel). This means that there are still competing activations (ambiguous probabilistic information that points to two or three conflicting outcomes) between the conjunctive and individual cues on switch trials. Thus LF-learning will, in theory, prove far less helpful in mediating response conflict.

(Figure 6 about here.)

LF-training thus allows for our predictions about both the influence of discrimination learning in context, and the mechanisms that give rise to it, to be tested in greater detail. While we expect that naming training in an FL-learning configuration will reduce response conflict in the DCCS, allowing children to switch sort dimensions, we expect that LF-training will result only in the learning of the transitional probabilities between the dimension labels and the cards, and hence result in far less reduction in response conflict in the task.

**Training Experiment**

To test these ideas, we examined the effect of off-line discrimination training on children’s on-line performance in the DCCS.

**Participants**

47 English-speaking children between 3- and 4-years-old (M = 3 years, 6.8 months) participated in this study, with a near even balance between genders (25 girls and 22 boys). Participants were recruited from Stanford and the surrounding community.

**Methods and Materials**

Two groups of children received either Feature-First (FL) or Label-First (LF) training on the cards, before completing standard DCCS tasks (Zelazo, 2006). A control group was tested on the DCCS without training.
In the two training conditions, children were introduced to ‘shape’ and ‘color’ games prior to the DCCS using 12 sorting cards (6 yellow flowers and 6 green boats). In the FL condition children were first told, “In the shape game, we name the different shapes on these cards.” The experimenter then presented the first card to the child and asked the child to label it. After children correctly labeled the first 6 of the 12 cards, the experimenter said, “We’re going to play the color game. In the color game, we are going to say what colors are on these cards.” Children then labeled the remaining 6 cards in the new game.

While children in the FL-condition saw the card and labeled it, children in the LF-condition were asked to say the label before seeing the card. They were told, “In the shape game, we name the different shapes on these cards. The first card is going to be a flower— can you say ‘flower’?” The experimenter showed the card to the child only after the child had repeated the label. The structure of the LF-training was the same as the FL-training: naming 6 cards by one dimension and then switching to the other dimension.

The two training groups and the control group then completed two standard DCCS tasks. The first DCCS task used the same sort cards as in the training (yellow flowers and green boats), while the second DCCS task used new pictures and colors (blue trucks and red stars). The first testing dimension (shape or color) was counterbalanced across children, and children were required to correctly sort six cards in the pre-switch. Once a child had done this, the sorting dimension was switched. Exactly six cards were sorted in the post-switch test. Before each trial, children were either reminded of the current game’s rules or asked to answer “knowledge questions,” such as, “Where do the flowers go? Where do the boats go?” Children were given no feedback about their sorting of the cards. After the first DCCS task, children completed the second DCCS task with new cards, and the switching dimension was reversed from the first task.

(Figure 7 about here.)

Results

All children in the two training conditions correctly labeled the cards. Children were considered to have “passed” the DCCS task if they sorted at least 5 of the 6 post-switch cards correctly. 69% of the FL-trained children passed the first DCCS task and 75% the second. By
contrast, 33% of the LF-trained children passed the first task and 40% the second. 19% of the control children passed in each test (Figure 7).

Chi-square ($\chi^2$) tests revealed significantly higher passing rates in the FL-condition (11/16 children passed) the first switch as compared to 5/15 in the LF-condition ($\chi^2 [1, N = 31] = 9.7, p = 0.005$; second switch, FL-condition, 12/16 passed; LF-condition 6/15, $\chi^2 [1, N = 31] = 17.0, p = 0.001$). Against the control group (3/16 in each), the comparisons with the FL-condition were, first switch, $\chi^2 [1, N = 33] = 14.9, p = 0.001$; second switch, $\chi^2 [1, N = 33] = 23.7, p = 0.001$.

**General Discussion**

We predicted that if children were exposed to contexts that promoted discrimination learning by playing naming games, they would later be able to transfer this contextual learning and succeed at the DCCS task with relative ease. Consistent with these predictions, children given training that promoted contextual discrimination learning were consistently able to pass the DCCS. When training did not encourage discrimination learning, children’s performance was far worse, while children given the task without prior training failed it. These findings are consistent with our suggestion that the mind has multiple ways of dealing with conflicting response demands, and with evidence that on-line processing of response conflict appears to develop over the course of childhood (Rueda, Rothbart, McCandliss, Saccamanno & Posner, 2005), and may be largely unavailable to children under four (Ramscar & Gitcho, 2007). That under-fours are able to behave in ways that are flexible and contextually appropriate appears to be due to discrimination learning. As the FL-trained children show, under-fours can match their behavior to context in remarkably subtle and sensitive ways, but it may be that they can do so only once discrimination learning has enabled them to do so (for similar effects of training in “theory of mind” tasks, see Slaughter & Gopnik, 1996, Amsterlaw & Wellman, 2006).

Why do under fours lack the ability to dynamically process response conflict? We suggest that the delay of cognitive maturation is due to the value of unsupervised, learning strategies in the mastery of human behavioral conventions. Both linguistic and social knowledge are, in essence, conventional. For symbols (such as words) to have “symbolic value,” their values must be both conventionalized and internalized. In this domain, children’s remarkable inflexibility may actually prove to be to their advantage (Thompson-Schill et al, in press;
Ramscar & Gitcho, 2007; Ramscar et al., 2009). Given a similar set of cues and labels to learn, young human learners will tend to sample the environment in much the same way, making it likely that they will come to have similar expectations regarding the relationship between cues and symbols.

The question of how whole communities come to share the same symbolic conventions has rarely been posed, let alone answered (Wittgenstein, 1953). We would argue that these shared expectations are in large part what linguistic conventions amount to (see Ramscar et al., 2009). If two learners are unable to select what they attend to during the course of learning, then given a similar set of cues and labels to learn, they will acquire similar representations of those cues and labels. Both learners will come to have similar expectations regarding the relationship between cues and symbols. If, on the other hand, those same learners were able to choose what they attended to in learning, the potential for conventions to be learned in this way would decrease dramatically. Thus the less children are able to direct their attention in learning, the more what they learn will be shaped by the statistical regularities of their physical, social and linguistic environments (see also Singleton & Newport, 2004; Hudson, Kam & Newport, 2005). In contrast, adults struggle to master new linguistic conventions (Johnson & Newport, 1989). This may reflect an inevitable handicap that adults’ increased ability to selectively respond or attend to the world imposes on convention learning. If learners are able to choose what they attended to in learning, the potential for conventions to arise in the way we describe above will decrease dramatically. The greater the variety there is in what adults focus on, the less conventionality there will be in what they learn.

It is clear that flexible, controlled cognition is important for thinking in unconventional ways. There is also evidence that it is effortful, and that our capacity for such thought is limited (see e.g., Yeung et al., 2004). We suggest that the lack of this ability in very young children reflects the way that an inflexible, uncontrolled learning process optimizes convention learning (Thompson-Schill et al, in press; Ramscar & Gitcho, 2007). Further, the ability of under-fours to “hide” their cognitive inflexibility (such that it requires ingenious psychological experiments to be revealed), and the speed with which children in our experiment were able to utilize contextual cues to succeed at the DCCS may have important implications for our understanding of “everyday” cognitive processing in adults. The evidence we present here suggests that young children make extensive use of context in their interactions with the world: it may be that much...
of adult cognition works in the same way, and is thus far more contextually embedded than is often supposed (see also Fodor, 2000).
References


Figure Captions

*Figure 1.* The DCCS task. Cards can be sorted by shape (in which case, the star is sorted into the left bin) or color (in which case, the star is sorted to the right bin).

*Figure 2.* Initially a child learning to name might expect that a card with a red star on it might be correctly labeled as “red” or “star.” However, if she consistently hears “red” in context A—which might correspond to the question, “what color is it?”—the child’s expectations will weaken the association between the card and “star” in this context. The converse will be true if she hears “star” in context B.

*Figure 3.* Rescorla-Wagner simulation of cue competition in two DCCS trials. Each line represents the association between a given cue and the correct response. Trials 1-40 represent assumed prior learning about *red* and *star* as relevant dimensions, while trials 41-60 represent the color game and trials 61 onwards the shape game. The erroneous expectations that *star* produces in color game trials cause it to be unlearned as a cue, resulting in *red* initially being a far more active cue; this pattern reverses over the course of the shape game. Note that, for modeling purposes, the 6 trials in each DCCS game correspond to 20 trials in the simulation.

*Figure 4.* Rescorla-Wagner simulation of cue competition in six DCCS games. Each line represents the association between a given cue and the correct response. Each peak can be taken to represent a rule switch. Notice how the individual cues lose associative value to the conjunctive cues over trials.

*Figure 5.* When labels precede the cards as discrete events, there may be no opportunity for cue competition.

*Figure 6.* Rescorla-Wagner simulations of cue competition in naming games in FL-Learning (left panel) and LF-Learning (right panel). Each line represents the association between a given cue and the correct response. Trials 1-40 represent prior learning about *red* and *star* as relevant dimensions. Trials 41-60 represent a color naming game, trials 61-80 a shape naming game, trials 81-100 a color naming game, etc. FL-learning results in the unlearning of the individual dimensional cues (*red* and *star*), whereas in LF-learning, the associative values of these cues continue to rise and fall as they are unlearned in one naming game context and relearned in the other.

*Figure 7.* Percentage of children passing the first and second DCCS tasks by condition.
Figure 1
Figure 2
Figure 3
Figure 4
Figure 5
Figure 6
Figure 7

This figure shows the percentage of children switching rules successfully across different conditions: Control, LF-Trained, and FL-Trained. The bars represent the performance in DCCS Game 1 and DCCS Game 2.