6-2008

# Dual Learning Processes in Interactive Skill Acquisition

Wai-Tat Fu
*University of Illinois at Urbana-Champaign*

John R. Anderson
*Carnegie Mellon University*, ja@cmu.edu



Published In
Journal of Experimental Psychology: Applied, 14, 2, 179-191.

# Dual Learning Processes in Interactive Skill Acquisition

Wai-Tat Fu

University of Illinois at Urbana-Champaign

John R. Anderson

Carnegie Mellon University

Abstract

Acquisition of interactive skills involves the use of internal and external cues. Experiment 1 showed that when actions were interdependent, learning was effective with and without external cues in the single-task condition but was effective only with the presence of external cues in the dual-task condition. In the dual-task condition, actions closer to the feedback were learned faster than actions farther away but this difference was reversed in the single-task condition. Experiment 2 tested how knowledge acquired in single and dual-task conditions would transfer to a new reward structure. Results confirmed the two forms of learning mediated by the secondary task: A declarative memory encoding process that simultaneously assigned credits to actions and a reinforcement-learning process that slowly propagated credits backwards from the feedback. The results showed that both forms of learning were engaged during training, but only at the response selection stage, one form of knowledge may dominate over the other depending on the availability of attentional resources.

# Dual Learning Processes in Interactive Skill Acquisition

The acquisition of interactive skills often involves the retention and utilization of internal cues (e.g., episodic memory of previous actions) and the recognition of relevant external cues (e.g., signs on the walls). The utilization of both internal and external cues to inform actions can be commonly found in most human-technology interactions (e.g., Card, Moran, & Newell, 1983; Fu, & Pirolli, 2007), in which the operators have to encode and recall previous actions and their outcomes and to interpret ongoing information displayed on an interface to decide what to do next. A better understanding of the interactive processes involved in the use of these internal and external cues is critical in predicting learning and performance in different designs of human-technology interfaces.

Understanding interactive skills requires the careful study of how internal and external cues are processed during the interplay of cognition and perception (Ballard, Hayhoe, Pook, & Rao, 1997; Fu & Anderson, in press; Fu & Gray, 2000, 2004, 2006; Gray & Fu, 2004; Gray, Sims, Fu, & Schoelles, 2006; Larkin, 1989). Our focus is on one of the most typical situations in which immediate feedback on individual actions is not available, so that the learner has to perform a sequence of actions and gets feedback on their success at the end of the sequence. This kind of situation creates a difficult credit-assignment problem, in which the learner has to assign credits to earlier actions or external cues that are responsible for eventual success. The credit-assignment problem is even more difficult in dynamic environments in which the action outcomes are probabilistic and interdependent, and the environment may change both autonomously and as a result of the actions. In this case, either memory of previous actions or recognition of the correct cues in the external environment is required to properly assign credits to the appropriate actions.

Based on the recent findings from psychological research, we hypothesize that humans exhibit two distinct learning processes to solve this credit assignment problem: a declarative process that encodes internal memory cues for actions and outcomes, and a reinforcement-learning process that encodes the relationship between the external cues and action outcomes (e.g., Daw, Niv, & Dayan, 2005; Grafton, Hazeltine, & Ivry, 1995; Keele, Ivry, Mayr, Hazeltine, & Heuer, 2003; Knowlton, Squire, & Gluck, 1994; Knowlton, Mangels, & Squire, 1996; Packard, & Knowlton, 2002; Poldrack, Clark, Pare-Blagoev, Shohamy, Moyano, Myers, & Gluck, 2001; Squire, 1992; Waldron, & Ashby, 2001). One major source of evidence for this distinction is from studies that show that amnesic patients can perform some tasks (e.g., artificial grammar, sequence, or category learning) despite their lack of declarative access to the events of training (e.g., Foerde, Knowlton, & Poldrack, 2006; Knowlton et al., 1994; Nissen, & Bullemer, 1987), whereas patients with basal ganglia disorders show major impairment to similar tasks (Poldrack et al., 2001). Furthermore, recent findings from neuroscience show that neural activities in the basal ganglia correlate well with the predictions of reinforcement learning when learning occurs in various probabilistic reward structures (e.g., Schultz, Dayan, & Montague, 1997). These two sets of results suggest that there exist two distinct learning systems for probabilistic events.

In this article, we describe results from two experiments designed to tease apart the nature of these two learning systems in the acquisition of interactive skills. We define interactive skills broadly in terms of learning of action sequences in situations that depend critically on the utilization of external cues. Our experiments are designed by bridging together two lines of psychological research: First, learning the sequential nature of actions is related to the research on sequence learning; second, learning the probabilistic relationship among external cues,

previous actions, and their outcomes is related to the research on probability learning and probabilistic classification.

*Sequence Learning*

The dual learning processes have often been investigated through a paradigm called sequence learning (e.g., Cohen, Ivry, & Keele, 1990; Curran, & Keele, 1993; Nissen, & Bullemer, 1987; Willingham, Nissen, & Bullemer, 1989). One typical paradigm is the serial reaction time (SRT) task in which participants have to press a sequence of keys as indicated by a sequence of lights. A certain pattern of button presses recurs regularly and participants give evidence of learning this sequence by pressing the keys for this sequence faster than a random sequence. One common finding is that learning is observed as a facilitation of test performance without concomitant awareness of what is being learned. A number of studies have used a secondary task such as counting of tones to study the effects of diminished attention for sequence learning (e.g., Cohen et al., 1990; Curran, & Keele, 1993; Nissen, & Bullemer, 1987). Cohen et al. (1990) found that when attention is diminished by a secondary task, participants could only learn simple pairwise transitions, but failed to learn higher order hierarchical structures in the sequence. The results show that although sequence learning can occur with diminished attention, its scope is limited to simple cases in which simple associations of stimuli are sufficient to perform the task.

*Probability Learning and Probabilistic Classification*

In a typical probability-learning experiment, participants guess which of the alternatives occurs and then receives feedback on their guesses (e.g., Estes, 1964). One robust finding is that participants often “probability match”; that is, they will choose a particular alternative with the same probability that it is reinforced (e.g., Friedman, Burke, Cole, Keller. Millward, & Estes,

1964). This leads many to propose that probability matching is the result of a habit-learning mechanism that accumulates information about the probabilistic structure of the environment (e.g., Graybiel, 1995; Knowlton et al., 1994). One important characteristic of this kind of habit learning is that information is acquired gradually across many trials, and appears to be independent of declarative memory as amnesic patients could learn to perform in a probabilistic classification task (Knowlton et al., 1994; but see Gallistel, 2005). However, for non-amnesic human participants, it is difficult to determine whether this kind of probabilistic classification is independent of the use of declarative memory. Given that declarative memory is dominant in humans, it has been argued that learners often initially engage in declarative memory encoding in which they seek to remember sequential patterns even when there are none (Yellott, 1969). Researchers argue that true probabilistic trial-by-trial behavior only appears after hundreds of trials – perhaps by then participants give up the idea of explicitly encoding patterns and the habit-learning process becomes dominant (Estes, 2002; Vulkan, 2000). Similarly, recent research on complex category learning has also provided interesting results suggesting the dual learning systems (Allen, & Brooks, 1991; Ashby, Queller, & Berretty, 1999; Waldron, & Ashby, 2001).

## Present Approach

In both the sequence-learning and the probability-learning paradigms, participants do not need to learn from the delayed feedback of a single action as immediate feedback is given. In a typical SRT task there is a sequence of actions but there is a deterministic relationship (given by instructions) between the stimuli and their responses. Participants in the SRT may anticipate the next stimuli but they always get immediate feedback after their responses. In probability learning the stimulus-response relationship is probabilistic but there is a single action after which feedback is received. Neither of these paradigms then directly reflects the complexity of the

credit-assignment problem in interactive tasks in which people often have to learn to sequentially choose actions with probabilistic outcomes and receive feedback only after the whole action sequence is executed. Our studies were designed by combining research from both areas by studying the nature of the learning processes that assign credits to different actions and external cues in a probabilistic sequential choice task. In this task, a sequence of actions was executed before feedback on its correctness was received, and a particular action sequence was correct only with a certain probability. Our goal is to use the novel paradigm to investigate the nature of the dual learning processes in the general context of interactive skill learning when the learner has to choose the right action sequences by utilizing either memory or external cues in the environment.

*The Probabilistic Sequential Choice Task*

A probabilistic sequential choice task was designed in which we predicted different behavioral patterns exhibited by the declarative memory encoding and reinforcement learning processes. At the outset, we would like to stress that although previous studies on the dual learning processes aimed at identifying distinct neural substrates for these two learning processes, our goals were different in that we aimed at finding out the differences in the *nature* of the learning process that is associated with declarative knowledge of task (declarative memory encoding) and the learning process that is associated with tacit knowledge that cannot be verbalized (reinforcement-learning process). In addition, we aimed at finding out how these two learning processes may predict different behavioral outcomes and how they may interact in different situations during the acquisition of interactive skills.

In the task, participants were told that they were in a room and they had to choose one of the two colors presented on the screen to go to the next room. After making two choices,

participants would either reach an exit or a dead-end. Participants were instructed to choose the colors that would lead them to the exit as often as possible. Figure 1 shows an example of the task. In room 1, if they chose "red" they would go to room 2 with probability 0.8 and to room 3 with probability 0.2. (We will explain the objects "computers" and "books" later.) The probabilities were reversed if "blue" was chosen. After the first choice, if participants were in room 2, if they choose "yellow" there was a 0.6 probability of going to an exit and 0.4 probability of going to a dead end. Again, the probabilities were reversed if "green" was chosen. If participants were in room 3, choosing "yellow" would lead to an exit with probability 0.2 and to a dead end with probability 0.8. Choosing "green" would lead to an exit with probability 0.4 and that to a dead end with probability 0.6. Note that if "red" was chosen, "yellow" was more likely to lead to an exit than "green"; but if "blue" was chosen, "green" was more likely than "yellow". The two choices were therefore interdependent: The more likely colors in the second choice were dependent on the first choice.

We assumed that the declarative memory system would learn by a goal-directed "tree-searching". Specifically, learning would occur by encoding the first and second choices and their outcomes (exit or dead-end) in declarative memory. If they led to an exit, both choices would be credited, otherwise both would be penalized. With enough experiences, the pair of choices that would more likely lead to the exit would be chosen. The advantage of this strategy was that the credit assignments were fast and efficient, but in the expense of higher memory load. We therefore hypothesized that with the presence of a secondary task, learning by the declarative memory system would be hampered as it would be more difficult to maintain the two choices in declarative memory.

To distinguish between the two learning processes based on behavioral data, the probabilities were chosen such that one of the branches would more likely lead to an exit than the other branch (in Figure 1, the "red" branch would lead to an exit with probability 0.46 while that for the "blue" branch was 0.34). It can be easily shown that no matter which color was chosen in the second choice, the probability that choosing "red" in the first choice would eventually lead to an exit would be higher than that for choosing "blue" (the marginal probabilities were 0.46 and 0.34 for choosing "red" and "blue" respectively). On the other hand, if a color was randomly chosen in the first choice, the probabilities that choosing "yellow" or "green" in the second choice would lead to an exit would be equal (the marginal probabilities were both 0.4). Given that we assumed that both choices were credited at the same time if learning occured through declarative memory encoding, we predicted that in situations where declarative memory encoding was dominant, learning the first choice would be faster than that in the second choice.

Reinforcement learning was assumed to be a learning process that was distinct from declarative memory encoding. The basic prediction of reinforcement learning is that, when feedback is received after a sequence of actions, only the last cue-action association in the sequence will be reinforced (either positively or negatively) initially. As experience of different action outcomes accumulates in a specific probabilistic reward environment, the cue-action associations will gradually become secondary reinforcers such that earlier actions that lead to these cues will be reinforced. This backward propagation mechanism will gradually assign the delayed feedback to earlier cue-action associations (Fu, & Anderson, 2006; Sutton, & Barto, 1998). The major characteristic of this form of reinforcement learning is that learning of actions closer to the feedback will be faster than actions farther away from the feedback. Given that the

learning process does not require active maintenance of previous actions in declarative memory, we hypothesize that it will be less hampered by a secondary task that imposes high working memory load.

The design in this task allowed us to directly pit the two learning processes against each other by comparing which of the two choices would be learned faster in different conditions: When declarative memory encoding was dominant, learning of the first choice would be faster; when reinforcement learning was dominant, learning of the second choice would be faster. We predicted that in the single-task condition, declarative memory encoding would be dominant, and learning the first choice would be faster than the second choice. In the dual-task condition, we predicted that a concurrent secondary task would significantly suppress the declarative memory encoding process but not the reinforcement learning process. Thus, learning of the second choice would be faster than the first choice. In addition, we predicted that reinforcement learning was possible only when the external cues were present to distinguish between the rooms (i.e., "computers" and "books" in room 2 and 3 as shown in Figure 1). When these objects were taken away, reinforcement learning would fail, as the probability of reaching an exit after choosing yellow and green would be the same and thus no learning could occur at the second choice. Given that learning of the first choice depended on the second choice, learning of the first choice would also fail.

Our second prediction on the nature of the two learning processes may seem counterintuitive. We assumed that learning by declarative encoding, as described in Figure 2, was more accessible for recall and eventually lead to top-down, rule-based behavior (e.g., Anderson, 1982). For example, in the example shown in Figure 1, learning by declarative memory encoding would more likely lead to rule-based behavior such as "if in the first room,

choose Red", or "If computers are present, choose Yellow". This declarative rule-based behavior was found to be pervasive in probabilistic classification task, especially when the stimuli were distinguishable by a single dimension (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Ashby, & Gott, 1988). These rules were also found to be optimal to the global properties of the task (Ashby, Maddox, & Bohil, 2002). We predicted that this kind of rule-based behavior would be less affected by short-term fluctuations in its correctness in a probabilistic environment (as indicated by trial-by-trial feedback). In other words, we predicted that learning by declarative memory encoding would lead to fewer changes in their color choices after an error feedback.

In contrast to declarative memory encoding, reinforcement-learning is mostly driven by gradual accumulation of experiences through trial-and-trial feedback to inform the correctness of the choices. It is therefore more likely influenced by local fluctuations of reinforcement rates (i.e., probabilities that a particular choice of colors will lead to an exit averaged over a small number (<10) of trials). Although we did not directly manipulate the distributions of local probabilities, previous studies that aimed at teasing apart the different effects of local and global probabilities have shown that people are indeed sensitive to these local probabilities (e.g., Friedman et al., 1964; Herrnstein, & Prelec, 1991). In fact, in a number of animal studies where the reinforcement learning process was believed to be dominant, sensitivities to local reinforcement rates were found to be much higher than the global reinforcement rates (e.g., Vaughan, 1981). Elsewhere, we have also shown that a model of reinforcement learning does exhibit this property of higher sensitivity to local reinforcement rates (Fu, & Anderson, 2006). We therefore predicted that the reinforcement learning process would lead to more changes in their color choices after an error feedback, compared to learning by declarative memory encoding.

## Experiment 1

### *Method*

#### *Participants*

Fifty-two participants in the Carnegie Mellon University community and 12 participants in the University of Illinois were recruited for the experiment. The age of the participants ranged from 18-32, and approximately half of them were males and the others were females. Four of the participants could not maintain the 2-back task performance at 80% and were excluded. Participants received a base payment of $8 plus a bonus payment of up to $7 depending on performance. Half of the participants were assigned to the single-task group and the other half to the dual-task group; and participants in each group were further divided into the distinct and ambiguous conditions.

#### *Materials and Stimuli*

Before the experiment began, two pairs of colors were randomly chosen from a set of eight colors (red, green, yellow, blue, brown, gray, magenta, and orange) to be used in the two choice sets during the task. One of the colors in each choice set was selected as the more likely color in that set. In the distinct condition, one object was randomly selected from a list of eight objects (torches, spiders, tables, fishes, dishes, books, computers, and cigarettes) and assigned to each of the two choice sets. The colors were presented on a white background in the middle of a 17-in computer monitor. One color was on the left side of the screen and the other was on the right side of the screen, but their locations were randomized across trials. Participants were instructed to hit the left arrow button to select the color on the left and the right arrow button to select the color on the right using a standard US keyboard.

In the dual-task conditions, participants were required to perform a concurrent "2-back" secondary task. A continuous stream of numbers (from 0 to 9) was presented aurally from the speakers. Starting from the third number, participants had to press the control key on the keyboard if the number was identical to the numbers two numbers before. For example, if they heard the numbers 0, 3, 2, 3, and 0, they had to press the control key the second time they heard 3. The numbers were presented once every two seconds. Participants had to maintain their performance at 80% or better at the 2-back task while performing the probabilistic sequential choice task. When participants missed a number or hit a key incorrectly, the word "miss" or "wrong" would appear at the corner of the screen respectively.

*Design*

The experiment was a 2 (distinct vs. ambiguous) x 2 (single vs. dual-task) between-participant design. Participants were randomly assigned to one of the four between-participant conditions. Learning in the two choices (first vs. second) was measured by the choice proportions of the more likely colors in each of the choice sets. Given that we had specific predictions on the learning of the two choices in each between-participant condition, learning of the first and second choices were treated as a within-participant variable in the analysis.

*Procedures*

In the dual-task conditions, participants were given a practice trial with 20 numbers on the 2-back task. If participants did not reach 80% accuracy, they were asked to read the instructions again and repeat the task. If participants failed to reach 80% accuracy the third time, they were dismissed, otherwise they would proceed to the probabilistic sequential choice task.

Participants were given a practice trial of the probabilistic sequential choice task during which the experimenter explained the task and answered any questions that the participants had.

When the task began, the two colors in the first choice set were presented in their colors on the screen. After participants made the first response, the two colors in the second choice set were presented immediately (<50 ms). In the distinct condition, the object associated with the corresponding choice set was presented at the center of the screen. After the second response, a feedback message was presented immediately (<50 ms) on the center of the screen to inform participants as to whether both of the choices had led to success. Participants were instructed that if the feedback was "Exit," both choices made were correct; but if the feedback was "Dead-end," then either one of the choices was wrong or both choices were wrong, and they would not receive any feedback on the correctness of the individual choices in this case. Participants can then hit the spacebar to begin the next trial. Participants were free to take as long as needed to respond and the stimuli and feedback stayed on the screen until a response was made.

Participants started with an initial score of 10 points. When an exit was reached, 5 points would be added to the final score; when a dead-end was reached, 1 point would be deducted from the final score. Participants were paid one cent for each point in the total score for the bonus payment. Each participant finished 200 trials. At the end of the experiment, a screen popped up with two questions with an empty box under each of the questions for participants to type in their answers. The first question was: *Please write down any strategy that you used to do this task*. The second question was: *Please write down any of the colors that you believe would most likely lead you to an exit.*

## *Results*

### *Choice Proportions of the More Likely Colors*

Figure 3 shows the mean choice proportions of the more likely colors ("Red" in room 1, "Yellow" in room 2, and "Green" in room 3 as shown in the example in Figure 1) in each 10-trial

block in the single- and dual-task conditions. Given that we had precise predictions on the rate of learning for the first ("Red") and second choice ("Yellow" and "Green" in room 2 and 3 respectively), this was treated as a within-participant variable in the ANOVA analysis. A 2 (first/second choice) x 2 (distinct/ambiguous condition) x 2 (single/dual task) x 10 (block) ANOVA on the choice proportions on the more likely colors showed that the main effects of choice, condition and task were significant ($F(1,56)=5.0$, $MSE=0.72$, $p<0.03$, $\eta_p^2 = 0.17$; $F(1,56)=16.72$, $MSE=3.88$, $p<0.001$, $\eta_p^2 = 0.23$; $F(1,56)=15.38$, $MSE=3.57$, $p <0.001$, $\eta_p^2 = 0.14$ respectively). The choice (first/second) x condition (distinct/ambiguous) interaction was not significant, but the choice x block interaction was significant ($F(9, 1064)=2.29$, $MSE=0.037$, $p<0.01$, $\eta_p^2 = 0.21$). This confirmed that in both ambiguous (no external cues) and distinct (with external cues) conditions, learning was faster in the first choice than the second choice. The three-way interaction choice x condition x task was significant ($F(1,56)=16.72$, $MSE=2.88$, $p<0.001$, $\eta_p^2 = 0.12$). There was no significant difference between the distinct and ambiguous conditions in the single-task condition. This lack of difference indicated that the external cues did not significantly improve performance in the single-task condition, although it could also be caused by a ceiling effect. We tested this possibility in Experiment 2, which was designed to directly test how external cues were utilized in single- and dual-task condition.

Figure 3 shows the sharply different patterns in the dual-task condition. Both the choice x condition and choice x blocks interactions were significant in the dual-task condition ($F(1,56)=4.87$, $MSE=0.70$, $p < 0.05$, $\eta_p^2 = 0.13$; $F(9, 1064)=12.29$, $MSE=0.012$, $p<0.001$, $\eta_p^2 = 0.23$ respectively), confirming that learning in the second choice was faster than the first choice in the dual-task distinct condition. Note that if participants were not aware of the choice dependency and always chose one of the more likely colors in the second choice set (i.e., chose

"yellow" in both room 2 and 3 using the example shown in Figure 1), the choice proportion would have been 80% of the choice proportion of the more likely color in the first choice (i.e., approximately 0.8 x 0.8 = 0.64 in the last 3 blocks). Given that the second choice proportions were higher than 0.64, participants had learned to choose the more likely colors in both room 2 and room 3 – i.e., they had learned the dependency between the choices. None of the choice proportions in the dual-task ambiguous condition was above chance throughout the 20 10-trial blocks, confirming the lack of learning in the ambiguous dual-task condition. The results were consistent with the prediction of the reinforcement learning process: Learning was faster in the second choice than the first choice. However, when external cues were not available in the ambiguous condition, no learning was observed.

The choice x task interaction was significant ($F(1,56)=3.87$, $MSE=1.20$, $p < 0.05$, $\eta_p^2 = 0.2$), confirming the different patterns observed in the single and dual-task conditions. Post-hoc analyses indicated that choice proportions of the more likely colors in the first choice were significantly higher in the single-task than the dual-task condition ($p<0.05$). However, only in the ambiguous condition, the choice proportions of the more likely colors in the second choice were significantly higher in the single-task than the dual-task condition ($p<0.05$). In the distinct condition, the choice proportion of the more likely colors in the second choice was not significantly different between the single- and dual-task conditions.

To summarize, the results matched very well with the predictions of the two learning processes. In the single-task conditions, participants encoded their choices in declarative memory and assigned credits to them simultaneously based on their outcomes. This declarative memory encoding strategy was fast, effective, and was not influenced by the lack of external cues, although it did require higher memory load and attentional resources. Indeed, our results showed

that this learning process was suppressed when participants had to perform the demanding secondary task simultaneously. On the other hand, the reinforcement learning process was effective in the dual-task condition but only with the availability of the external cues.

*Retrospective Verbal Reports*

The written strategies were analyzed by tallying the number of rooms in which they could identify the more likely colors (see Table 1) after the task. In the dual-task condition, most of the participants could not write down the more likely colors in any of the rooms, while participants in the single-task condition could write down the more likely colors in at least two of the rooms. The results were consistent with the proposed dual learning processes in our task. In the single-task conditions, most of the participants encoded the choices and their outcomes in declarative memory and were aware of the choice dependencies in the task. In the dual-task condition, given that the declarative memory encoding of past experiences was suppressed by the secondary task, most participants were not aware of the most likely colors (or at least they were not active enough to be reported retrospectively). Nevertheless, in the dual-task distinct condition, participants increasingly selected the more likely colors, demonstrating that learning did not require effortful maintenance of the choice-outcomes information during the task.

*Choices After an Error Feedback*

The second prediction of the two learning processes concerned how participants would change their choices after feedback. We calculated the choice proportions of the more likely colors according to the feedback they received in the last trial. Our prediction was that access to the declarative knowledge about the more likely colors would likely lead to goal-directed rule-based behavior that was adapted to the global properties of the task. Given that participants learned the fast choice faster than the second choice, we predicted that a declarative rule would

be formed in the first choice earlier than the second choice, and thus participants would *more likely* choose the same color after an error in the first choice than in the second choice. However, this difference would disappear at later blocks after participants learned both choices. In addition, we expected that the declarative memory encoding process would *more likely* lead to the continuance to choose the same colors after an error, because of the goal-driven nature of the rule-based behavior. On the other hand, because of the nature of the reinforcement learning process, we predicted that participants' choices would be more sensitive to the trial-by-trial feedback they received (and thus would *less likely* choose the same colors after an error).

Figure 4 shows the choice proportions of the same colors after an error (reached a dead-end) in the first 5 (Early) and the last 5 (Late) blocks of trials. There was no significant difference between the distinct and ambiguous single-task conditions. In the single-task and the distinct dual-task conditions, participants chose the same colors after an error more often in later trials than in early trials ($t(56)=4.21$, $p < 0.01$). The difference was not significant in the ambiguous dual-task condition, and the choice proportions stayed at chance levels in both early and late trials, confirming the lack of learning in that condition. Most importantly, participants chose the same colors more often after a wrong feedback in the single-task condition than the distinct dual-task condition ($t(56)=3.94$, $p<0.001$), confirming our prediction that the reinforcement learning process was more sensitive to an error feedback than the declarative memory encoding process. Except the ambiguous dual-task condition, the differences between the first and second choice were significant in the early trials but not in the late trials. However, in early trials of the single-task condition, participants switched after an error feedback more often in the second choice, but in the distinct dual-task condition, participants switched more often in the first choice. This pattern was consistent with our findings that the declarative

memory encoding process learned the first choice faster in the single-task condition but the reinforcement learning process learned the second choice faster in the dual-task condition.

*Discussions of Experiment 1*

Results from Experiment 1 have identified some of the distinct properties of the two learning processes. In the single-task condition, in which the behavioral data suggested that the declarative memory encoding process was dominant, learning was faster, knowledge acquired about the task could be verbally reported, and at a later stage of learning, choices of actions were rule-based and were thus less sensitive to trial-by-trial feedback. On the other hand, in the distinct dual-task condition, in which the behavioral data suggested that reinforcement learning was dominant, learning was slower, knowledge acquired was less accessible, and even at a later stage of learning, choices of actions were sensitive to trial-by-trial error feedback.

The advantage of declarative memory encoding is that it facilitates fast development of top-down behavioral rules that guides the selection of actions. Results from Experiment 1 showed that although initially the behavioral rules had to be developed based on local reinforcement information, once they were acquired, their top-down influence on action selection were relatively insensitive to changes in local reinforcement rates. On the other hand, reinforcement learning depends on gradual accumulation and propagation of local rewards and is thus more sensitive to local reinforcement rates. Although reinforcement learning is slow, one possible adaptive function is that it is able to detect changes in local reinforcement rates and behaviorally adapt to those changes. In fact, studies in animal foraging have shown that the ability to detect that a particular food patch is depleting and to switch to another food patch is critical for animal survival and is thus believed to play a critical role in this kind of habit learning (Stephens, & Krebs, 1986).

In Experiment 2, we tested how the two learning processes would adapt to changes in the reward structures differently. In addition, we tested whether the two learning processes utilized external cues differently when adapting to changes. From Experiment 1, one important implication of the dual learning processes was that the external cues played a more important role in the credit assignment process in reinforcement learning than in declarative memory encoding. This was indeed supported by our results in the dual-task condition: participants learned the second choice faster than the first choice with the presence of external cues, and when the external cues were taken away, learning failed. These results provided strong support for the reinforcement learning process. In the single-task condition, the lack of difference between the ambiguous and distinct condition was also consistent with the hypothesis that, in contrast to reinforcement learning, learning by declarative memory encoding did not depend critically on external cues. However, we could not rule out the possibility that the lack of difference was caused by a ceiling effect that limited the potential improvement provided by the external cues in the single-task condition. Experiment 2 was designed to answer these questions.

## Experiment 2

Experiment 2 was based on the same probabilistic sequential choice task as in Experiment 1, but the focus was on the distinct condition in which external cues were provided. Given that reinforcement learning used the external cues as a secondary reinforcer to pass the credits backwards to the first choice, we predicted that learning would be more sensitive to the external cues than declarative memory encoding. On the other hand, given that declarative memory encoding assigned credits simultaneously to both choices, we predicted that participants would be less sensitive to the external cue that linked the transition from the first to the second room. Using the example shown in Figure 2, we predicted that learning by declarative memory

encoding would assign credits to the color “Red” and “computers-Yellow” simultaneously, whereas reinforcement learning would initially assign credits to “computers-Yellow”, and gradually pass on the credits to “Red” as participants repeatedly observed that the color “Red” led to “computers”. Reinforcement learning would thus be more sensitive to the transition from “Red” to “computers” than declarative memory encoding. The difference implied that if the transitional probabilities from the first choice to room 2 and 3 were changed, learning by declarative memory encoding would be slower to adapt to the new reward structure (as defined by the marginal probabilities of choices) than reinforcement learning. In addition, given that results from Experiment 1 showed that learning by declarative memory encoding was less sensitive to trial-by-trial feedback than reinforcement learning, we predicted that learning by declarative memory encoding would be slower to adapt to a new reward structure.

The main design of Experiment 2 was to transfer participants trained in the same reward structure as in E1 to a new reward structure. Figure 5 shows the structure of the transfer stage. The major change is that the transition probabilities from each of the two colors in the first choice to room 2 and 3 were switched, i.e., the probabilities that choosing “Red” would lead to room 2 and 3 was changed from 0.8 to 0.2 and 0.2 to 0.8 respectively. Similarly the probabilities that choosing “Blue” would lead to room 2 and 3 was changed from 0.2 to 0.8 and 0.8 to 0.2 respectively. The probabilities that Yellow and Green in room 2 and 3 would lead to an exit were increased as shown in Figure 5. With these changes, the probability for reaching the exit by choosing “Red”, “computers-Yellow”, and “books-Green” (the best colors in the training stage) would stay the same in the transfer stage (i.e., both equal 0.56). However, choosing “Blue” (the less likely color in the training stage), “computers-Yellow”, and “books-Green” (the more likely colors in training stage) would increase from 0.44 in the training stage to 0.74 in the transfer

stage. In other words, if participants kept choosing the most likely colors they had learned during training in the transfer condition, their overall rewards would stay the same, but if they switched to the other color in the first choice, the overall rewards would be increased.

Figure 6 illustrates our predictions for the two learning processes in the transfer stage. From results of E1, we predicted that with the presence of external cues, participants would successfully learn to select the most likely colors in both the single- and dual-task condition during the training stage. Given that learning by declarative memory encoding would update both choices simultaneously, the external cue linking the two choices would not be directly updated by the feedback. Given that the overall reinforcement rates from these choices remain the same, we predicted that learning by declarative memory encoding would be slower to switch colors than in the transfer stage. On the other hand, given that the external cue would be critical in passing the credit from the feedback to the first choice during reinforcement learning, the external cue linking the two choices would be directly updated. Using the examples shown in Figure 6, given that reaching "books" was less rewarding than "computers", the more frequent transition from "Red" to "books" would eventually make "Red" less desirable; on the other hand, the more frequent transition from "Blue" to "computers" would eventually make "Blue" more desirable. We therefore predicted that reinforcement learning would lead to faster adaptation (switching colors) to the transfer stage by being more sensitive to the external cues.

## *Method*

### *Participants*

Forty participants in the University of Illinois were recruited for the experiment. The age of the participants ranged from 18-27, and approximately half of them were males and the others were females. Participants received a base payment of $15 for their participation. Half of the

participants were assigned to the single-task training group and the other half to the dual-task training group; and participants in each group were further divided into the single- and dual-task transfer conditions. External cues were presented in all conditions (i.e., the distinct condition in Experiment 1).

*Materials and Stimuli*

The stimuli were the same as in Experiment 1, except the change of probabilities during the transfer condition as specified in Figure 5.

*Design*

The experiment was a 2 (single vs. dual-task training) x 2 (single vs. dual-task transfer) between-participant design. Participants were randomly assigned to one of the four between-participant conditions. Similar to Experiment 1, learning and performance was measured by the choice proportions of the more likely colors in the task.

*Procedures*

Except in the single-to-single condition, all participants were given practice on the 2-back task before they began. All participants were told that there were 300 trials in the whole experiment and they would be informed after they finished every 100 trials. The experimenter then explained the probabilistic sequential choice task to the participants as in Experiment 1. Participants then began to perform the task. When participants finished 100 trials, a screen would inform the participants that they had finished 100 trials and they could then press the spacebar to proceed. Similarly, after they finished 200 trials, a screen would inform participants that they had finished 200 trials. In addition, in the single-to-dual condition, participants were informed that for the next 100 trials they had to perform the 2-back task concurrently with the probabilistic sequential choice task; in the dual-to-single condition, participants were told that they did not

need to do the 2-back task in the next 100 trials. In all conditions, the probabilities were changed as shown in Figure 5 in the last 100 trials.

*Results*

We successfully replicated the findings in Experiment 1 in the training stage (the first 200 trials) and will focus on the results at the transfer stage here. Given that in Experiment 2 we were mostly interested in the first choice, separate 2 (training: single- and dual-task) x 2 (transfer: single- vs. dual-task) x 10 (block) ANOVAs were conducted on the first and second choices. We did not find any differences in the second choices. The mean choice proportions (with standard deviations in parentheses) of the most likely colors were 0.82 (0.11), 0.84 (0.17), 0.82 (0.14), and 0.83 (0.06) for the single-to-single, single-to-dual, dual-to-single, and dual-to-dual respectively. As predicted, participants continued to choose the more likely colors in the second choices. However, an interesting pattern was observed in the first choices. Figure 7 showed the choice proportions of the trained colors (the most likely in the training stage but less likely in the transfer stage) in the transfer stage. The main effect of training was not significant ($p>0.5$) but that of transfer was significant ($F(1, 36)=229.69$, $MSE=1.24$, $p<0.001$, $\eta_p^2 = 0.31$). The interaction between training and transfer was not significant ($p>0.7$). Participants in the single-task transfer condition chose the trained colors more often than those in the dual-task transfer condition. The main effect of blocks and the transfer x blocks interaction was significant ($F(9, 324)=87.2$, $MSE=0.48$, $p<0.001$, $\eta_p^2 = 0.21$ and $F(9, 324)=5.57$, $MSE=0.0031$, $p<0.01$, $\eta_p^2 = 0.19$ respectively), but the training x blocks and the training x transfer x blocks interactions were not significant. Post-hoc analyses showed that in both the single-to-dual and dual-to-dual conditions, the choice proportions started to significantly drop after 20 trials (at the third 10-trial block, $p<0.05$) whereas for both the single-to-single and dual-to-single conditions the choice

proportions did not start to drop until after 60 trials (the seventh 10-trial block, $p<0.05$). The results were consistent with our predictions: participants in the dual-task transfer conditions learned the new reward structure faster than those in the single-task transfer conditions. These results provided strong support for the hypothesis that learning by declarative memory encoding did not depend on the external cues as much as reinforcement learning, and against the ceiling effect explanation of the lack of difference between the ambiguous and distinct single-task condition in Experiment 1. It was also striking to show that diminished attention resources could actually *help* participants to learn the new reward structure faster than in the single-task condition.

*Discussions of Experiment 2*

The overall results confirmed our predictions that reinforcement learning was more sensitive to changes in the new reward structure. The major change in the transfer stage was that, choosing the most likely colors in the first choice in the training stage would lead to more frequent transitions to the comparatively less rewarding room as indicated by the external cues (i.e., more "Red-books" than "Red-computers" transitions in Figure 5). Recognizing this change in transition frequencies was critical in adapting to the new reward structure in the transfer condition. Similar to what we observed from Experiment 1, in the dual-task transfer condition reinforcement learning was dominant. Given that reinforcement learning depended critically on the external cue to pass on the credits to the first choices, participants in the dual-task transfer condition learned to choose the most likely colors faster than those in the single-task condition, in which learning by the declarative memory encoding was found to be dominant.

As shown in the results from Experiment 1, another reason for the faster adaptation to the new reward structure was that the feedback-driven reinforcement learning was more sensitive to

local reinforcement rates than the rule-based behavior developed from the declarative memory encoding learning process. However, the lack of difference in the training conditions seems to suggest that this was not the sole reason for the slower adaptation. The results showed that different forms of learning during training did not make a difference in how well participants learned the new reward structure in the transfer condition. Given that we did not find any difference between the single-to-single and dual-to-single conditions, the declarative rules acquired in the single-task training condition did not seem to make participants less sensitive to the new choice contingencies. On the other hand, when access to declarative knowledge was not available, the tacit knowledge acquired during the dual-task training condition did seem to help guide the selection of colors in the single-task transfer condition. However, the declarative forms of learning in the single-task transfer condition had apparently made participants less sensitive to the external cue that indicated the transition between choices, thus making them less adaptive to the new reward structure.

The lack of difference between the single-to-dual and dual-to-dual conditions was consistent with the notion that learning by declarative memory encoding and reinforcement learning could occur at the same time (e.g., Willingham et al., 1989; Willingham, Salidis, & Gabrieli, 2002) in the single-task condition, but when making color choices declarative knowledge dominated the tacit knowledge gained from reinforcement learning. When participants trained in the single-task condition was switched to the dual task condition, expression of declarative knowledge was suppressed, and the tacit knowledge gained during the training stage became dominant in determining color choices.

## General Discussion

The results from our two experiments provided strong support for the dual learning processes in interactive skill acquisition. Results from Experiment 1 add to existing evidence showing that learning by declarative memory encoding requires memory and attentional resources, whereas reinforcement learning remains effective even with the presence of a demanding secondary task. To our knowledge, our data have provided the most direct evidence showing how these two learning processes contribute to the acquisition of interactive skills when feedback on the correctness of action sequences is delayed. We found that although reinforcement learning was less affected by the secondary task, distinct external cues were necessary to pass credits back from the feedback to earlier actions. This distinguishing feature of reinforcement learning was further supported by results from Experiment 2, which showed that reinforcement learning was more sensitive to the presence of the external cue that indicated the transition between the action and its outcome than declarative memory encoding. Furthermore, the lack of difference between different training conditions is consistent with the idea that in the single-task condition *both forms of learning are engaged* during training. The results are consistent with the idea that only at the stage of action selection, one form of knowledge dominates over the other to determine the response strategy (Willingham et al., 2002). It seems that only with the presence of the secondary task, declarative knowledge is suppressed and action selection is dominated by the tacit knowledge acquired during training.

The current results also have interesting implications to existing theories of skill acquisition. Traditional theories of cognitive skill acquisition propose that skills development starts with some forms of general declarative representations of actions, and through experience action selection and execution speeds up gradually (e.g., Anderson, 1982; Fitts, & Posner, 1967).

According to these theories, skill acquisition always starts with a slow, declarative stage where instructions are interpreted through verbal mediation, and working memory load is high because facts about the skill must be verbally rehearsed. Practice allows declarative knowledge to be slowly "compiled" into procedures that can be executed with little verbal mediation (e.g., Anderson, Bothell, Byrne, Douglass, Lebiere, & Qin, 2004; Logan, 1988; Newell & Rosenbloom, 1981). Our results indicate that in interactive skill acquisition, utilization of external cues can be learned by reinforcement learning that occurs in parallel to the slow declarative-to-procedural process. In fact, our results show that the acquisition of interactive skills involving actions that are cued by external information is possible even with a demanding secondary task without concurrent awareness of the information. It is therefore possible that the reinforcement learning process that we identified reflects a primitive form of learning that is sufficient to learn the associations between external cues and actions without verbal mediation. Indeed, this form of associative learning has a long history in animal research (e.g., Rescorla & Wagner, 1972). As Ohlsson (1993) argued, it is likely that humans' ability to learn preceded language as it is too recent for special purpose brain mechanisms for verbal instructions to have evolved. Nevertheless, we believe that given that learning to associate external cues to actions is an essential component of interactive skills, our results suggest that existing models of skill acquisition may need to be extended to include some forms of reinforcement learning that does not require the gradual declarative-to-procedural progression.

In our simple task, we found that explicit memory encoding is dominant, and reinforcement learning is dominant only when a demanding secondary task is present. We speculate that this may be caused by the simple structure of the task in which declarative memory encoding of previous actions is sufficient to maximize their performance scores. In

many real-world tasks, such as when a person is learning to use a new computer interface, learning will likely involve higher-level goals and may require more complex problem solving steps such as means-ends analysis that may impose a high cognitive load to the person (e.g., Sweller, 1988). It is possible that the higher cognitive load may hamper explicit memory encoding but not the reinforcement learning mechanism as in our dual-task conditions. Our results are consistent with the current wisdom that having good external cues is essential for effective learning in interactive environments (e.g., Larkin & Simon, 1987; Norman, 2002). Our findings that this kind of "display-based" learning (Larkin, 1989) could occur without explicit awareness are also consistent with previous research (e.g., Draper, 1986). Our main contribution is that our results have provided important information for *why* these cues could help learning and *how* these cues are learned during interactive skill acquisition.

References

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120*, 3-19.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*, 369-406.

Anderson, J. R., Bothell, D., Byrne, M., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of cognition. *Psychological Review, 111*(4), 1036-1060.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442–481.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 33–53.

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition, 30*, 666–677.

Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics, 61*, 1178-1199.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences, 20*(4), 723-742.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction.* London: Lawrence Erlbaum Associates.

Cohen, A., Ivry, R., & Keele, S. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 17–30.

Curran, T., & Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology*: *Learning, Memory, and Cognition*, *19*, 189–202.

Daw, N., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8*, 1704-1711.

Draper, S. W. (1986). Display managers as the basis for user-machine communication. In D. A. Norman & S. W. Draper (Eds.), *User centered system design* (pp. 339-352). Hillsdale, NJ: Erlbaum.

Estes, W. K. (1964). Probability learning. In A.W. Melton (Ed.), *Categories of human learning.* New York: Academic Press.

Estes, W. K. (2002). Traps in the route to models of memory and decision. *Psychonomic Bulletin and Review, 9*(1), 3-25.

Fitts P. & Posner, M. (1967). *Learning and skilled performance in human performance*. Belmont CA: Brock-Cole.

Foerde, L., Knowlton, B.J., & Poldrack, R.A. (2006). Distraction modulates the engagement of competing memory systems. *Proceedings of the National Academy of Sciences, 103*, 11778-83.

Friedman, M. P., Burke, C. J., Cole, M., Keller, L., Millward, R. B., & Estes, W. K. (1964). Two-choice behavior under extended training with shifting probabilities of reinforcement. In R. C. Atkinson (Ed.), *Studies in mathematical psychology* (pp. 250-316). Stanford, CA: Stanford University Press.

Fu, W. & Anderson, J. (2006). From recurrent choice to skill learning: A model of reinforcement learning. *Journal of Experimental Psychology: General, 135*(2), 184-206.

Fu, W.-T., Anderson, J. (in press), Solving the Credit Assignment Problem: Explicit and Implicit Learning of Action Sequences with Probabilistic Outcomes. *Psychological Research*.

Fu, W. & Gray, W. D. (2000). Memory versus perceptual-motor tradeoffs in a blocks world task. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, Mahwah, NJ: Lawrence Erlbaum Associates.

Fu, W. & Gray, W. D. (2004). Resolving the paradox of the active user: Stable suboptimal performance in interactive tasks. *Cognitive Science, 28*(6).

Fu, W.-T., Gray, W. D. (2006), Suboptimal Tradeoffs in Information-Seeking. *Cognitive Psychology, 52*, 195-242.

Fu, W., & Pirolli, P. (2007), SNIF-ACT: A Model of Information-Seeking Behavior in the World Wide Web. *Human-Computer Interaction.*

Gallistel, C. R. (2005). Deconstructing the law of effect. *Games and Economic Behavior, 52*, 410-423.

Grafton, S. T., Hazeltine, E., & Ivry, R. (1995). Functional mapping of sequence learning in normal humans. *Journal of Cognitive Neuroscience, 7*, 497-510.

Gray & Fu, (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science, 28*, 359-382.

Gray, W. D., Sims, C., Fu, W., & Schoelles, M. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behavior. *Psychological Review. 113,* 461-482.

Graybiel, A.M. (1995). Building action repertoires: memory and learning functions of the basal ganglia. *Current Opinion in Neurobiology, 5*, 733-741.

Herrnstein, R. J., & Prelec, D. (1991). Melioration: A theory of distributed choice. *Journal of Economic Perspectives, 5*, 137-56.

Keele, S., Ivry, R., Mayr, U., Hazeltine, E., & Heuer, H. (2003). The cognitive and neural architecture of sequence representation. *Psychological Review, 110*, 316-339.

Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science, 273*, 1399-1402.

Knowlton, B. J., Squire, L. R., & Gluck, M. (1994). Probabilistic classification learning in amnesia. *Learning and Memory, 1*, 106-120.

Larkin, J. H. (1989). Display-based problem solving. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 319–341). Hillsdale, NJ: Lawrence Erlbaum Associates.

Larkin, J. & Simon, H. (1987) Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65-99.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492-527.

Newell A. & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. R. Anderson, editor, *Learning and Cognition*. NJ: Hillsdale, Erlbaum.

Nissen, M., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19*, 1–32.

Norman, D. (2002). *The design of everyday things*. New York: Basic Books.

Ohlsson, S. (1993). The interaction between knowledge and practice in the acquisition of cognitive skills. In S. Chipman & A. L. Meyrowitz (Eds.), *Foundations of knowledge*

*acquisition: Cognitive models of complex learning.* Boston, MA: Kluwer Academic Publishers.

Packard, M. & Knowlton, B. (2002). Learning and memory functions of the basal ganglia. *Annual Review of Neuroscience, 25*, 563-593.

Poldrack, R., Clark, J., Pare-Blagoev, E., Shohamy, D., Moyano, J., Myers, C., & Gluck, M. (2001). Interactive memory systems in the human brain. *Nature, 414*, 546-550.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*, 1593-1599.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99*, 195–231.

Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory*. Princeton, NJ: Princeton University Press.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction.* Cambridge, MA: MIT Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12,* 257-285.

Vaughan, W. Jr. (1981). Melioration, matching, and maximization. *Journal of Experimental Analysis of Behavior, 36*, 141-149.

Vulkan N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys, 14*, 101–118.

Waldron, E., & Ashby, G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review, 8*, 168-176.

Willingham, D., Nissen, M., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1047-1060.

Willingham, D. B., Salidis, J. & Gabrieli, J. D. (2002). Direct comparison of neural systems mediating conscious and unconscious skill learning. *Journal of Neurophysiology, 88*, 1451–1460.

Yellott, J. L. (1969). Probability learning with noncontingent success. *Journal of Mathematical Psychology, 6*, 541-575.

## Author Note

The current work is supported by a grant from the Office of Naval Research (N00014-99-1-0097) and a research fund from the Human Factors Division and the Beckman Institute at the University of Illinois. The results on choice proportions in Experiment 1 were previously reported in the Proceedings of the 28th Cognitive Science Conference (2006). We ran 12 more participants in the results reported in this article. The patterns of results were the same with the addition of more participants. In the current article, we provided additional analyses on the choice proportions conditional on an error feedback and we found that with more participants the main effect between early and late trials was significant (it was not significant without the addition participants). The additional participants therefore did not lead to any major changes in patterns of results.

*Table 1*

*Number of participants who wrote down the more likely colors in each of the experimental condition.*

| | Single-Task | | Dual-Task | |
|---|---|---|---|---|
| Rooms | Distinct | Ambiguous | Distinct | Ambiguous |
| All | 12 | 9 | 2 | 1 |
| R1 & R2 | 2 | 5 | 0 | 1 |
| R1 & R3 | 0 | 0 | 0 | 0 |
| R2 & R3 | 0 | 0 | 0 | 0 |
| R1 | 1 | 1 | 2 | 1 |
| R2 | 0 | -- | 1 | -- |
| R3 | 0 | -- | 0 | -- |
| none | 0 | 0 | 10 | 12 |

Note. In the ambiguous condition, participants were not aware of the distinction of room 2 and 3

Figure Captions

*Figure 1.* The probabilistic sequential task use in both experiments. The circled numbers represent room numbers, and the numbers next to the arrows represent transition probabilities. Note that in room 3, regardless of what is chosen, there is a higher probability that it will lead to a dead-end compared to room 2. The actual colors were randomly selected from eight colors (red, green, yellow, blue, brown, gray, magenta, and orange) for each participant.

*Figure 2*. The two hypothesized learning processes and their behavioral predictions in the probabilistic sequential choice task. The colors and objects were based on the example shown in Figure 1. (a) Learning by declarative memory encoding: the first (Red) and second choices (computers-yellow) are encoded in memory as they are presented on the screen. When the feedback is presented (exit), both choices are updated. The size of the arrow next to the box represent the likelihood that this choice will be chosen again in the next trial. Given the design of the task, learning of the first choice will be faster than the second choice (see text for details). Therefore the size of the arrow for Red is larger than that for computers-Yellow (the more likely colors). (b) Reinforcement learning: Initially, all colors are neutral (neither positively or negatively reinforced). Therefore after Red and computers-yellow are presented, no learning occurs (no credit assignment). Only after the positive feedback (Exit) is presented, computers-yellow is assigned the credit of leading to the Exit and is thus positively reinforced. The arrow next to computers-Yellow indicates that the likelihood of choosing this color has increased to a positive level. After several trials of learning, given that computers-Yellow has gained a positive valence, Red will be reinforced for leading to computers-Yellow. After the positive feedback, computers-Yellow is further reinforced (as indicated by the larger size of the arrow next to it). Given the backward propagation of credits, the second choice will be learned faster than the first

choice. Eventually, both kinds of learning will lead to higher choice proportions of the more likely colors (Red and Yellow).

*Figure 3.* Choice proportions of the colors that were more likely to lead to the exit in the single-task and dual-task conditions in each of the 10-trial blocks. Using the example shown in Figure 1, "first" would be the choice proportions of "red", and "second" would be the sum of the choice proportions of "yellow" and "green" in room 2 and room 3 respectively. Error bars represent standard errors.

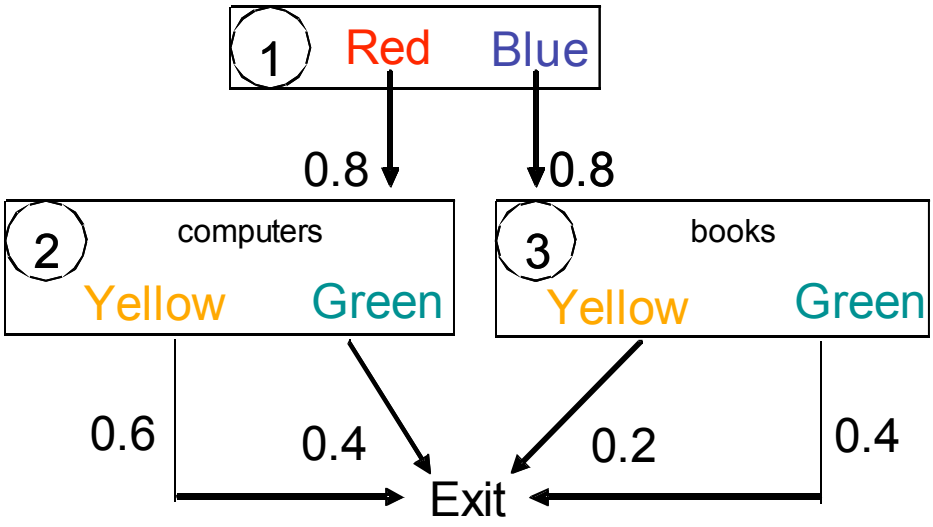
*Figure 4*. The choice proportions of the same colors after an error feedback in the first and second choice, broken down into early (first 5 trials) and late (last 5 trials) in the single-task, distinct dual-task, and ambiguous dual-task conditions. Error bars represent standard errors.

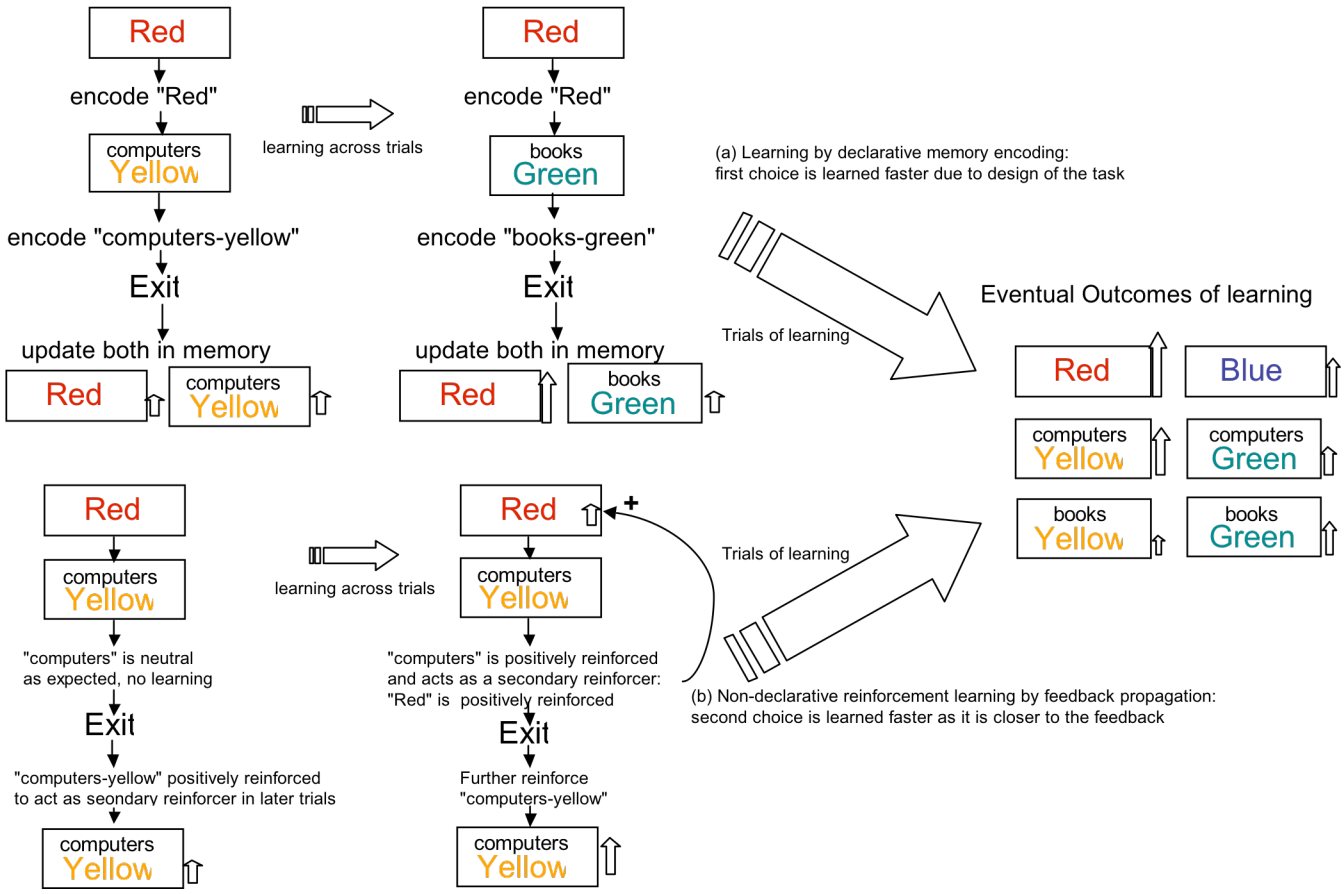
*Figure 5*. The probabilistic sequential task used in the transfer condition of Experiment 2. Compared to that used in the training condition (same as that in Experiment 1, see Figure 1), the transition probabilities from each of the two colors in the first choice to room 2 and 3 are switched, i.e., the probability that choosing "Red" will lead to room 2 and 3 is changed from 0.8 to 0.2 and 0.2 to 0.8 respectively, the same is true for "Blue". In addition the probabilities that Yellow and Green in room 2 and 3 will lead to an exit are increased as shown in the figure. See text for the rationale for these changes.

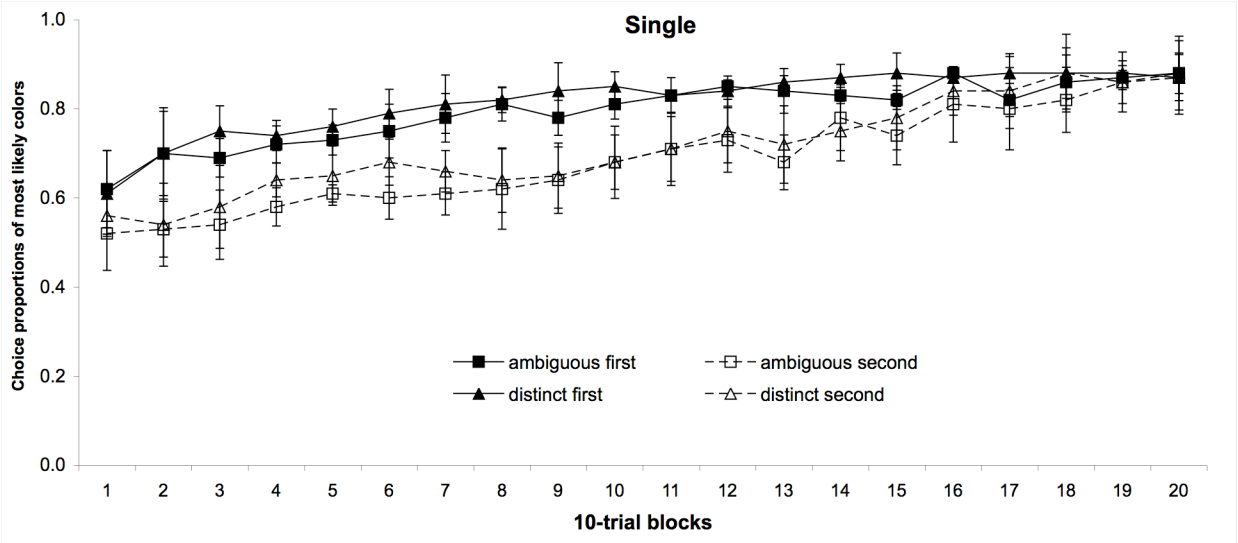
*Figure 6*. The predicted behavior of the two learning processes in the transfer stage in Experiment 2. "Red" and "computers-Yellow" are the most likely colors in the training stage. During the transfer stage, choosing the same trained colors will lead to the same expected reward, but choosing "Blue" and "computers-Yellow" will increase the expected reward. In declarative memory encoding, the two choices are encoded and updated separately and the external cue that between the two choices are not updated directly. The arrow next to each color

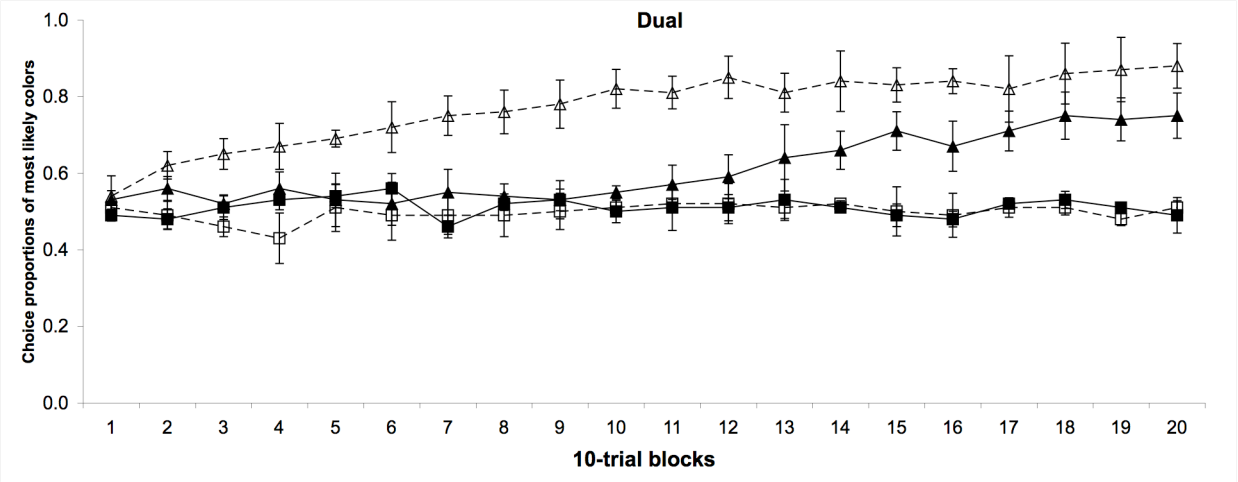
choice indicates the likelihood that it will be chosen in the future. In reinforcement learning, the presentation of “books-Green” (a less likely color in the training stage) after choosing “Red” directly devalued “Red”, thus giving “Blue” a comparative advantage in future choices.

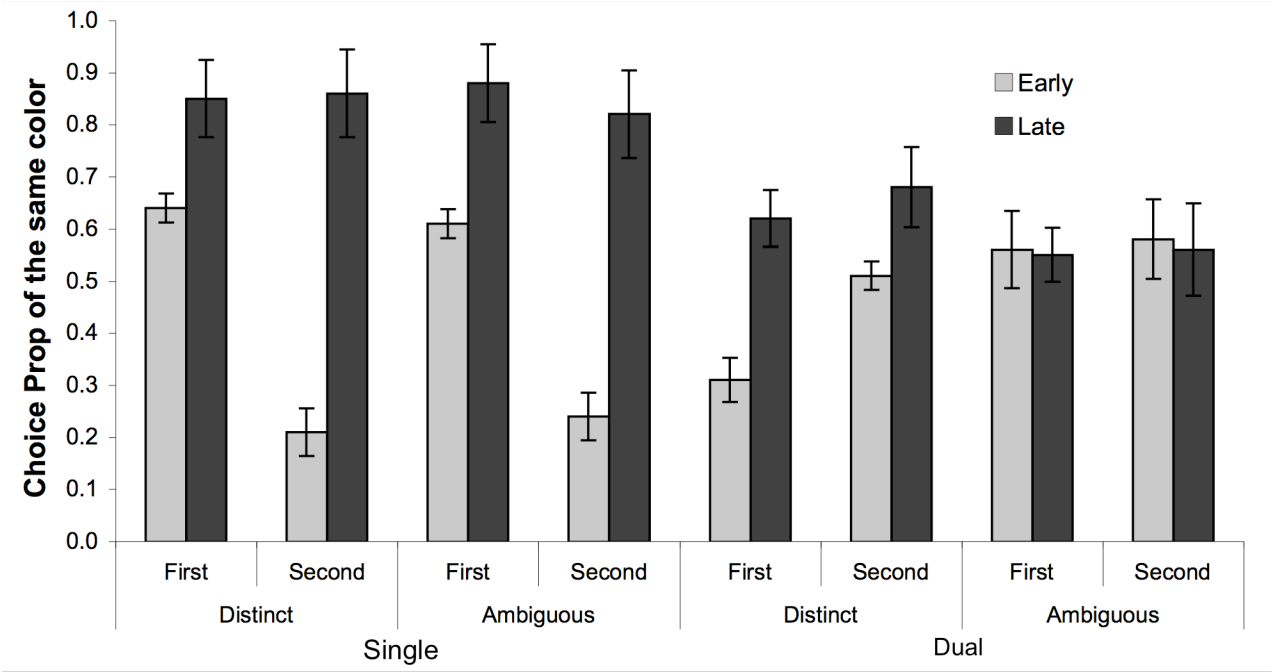
*Figure 7*. Choice proportions of the trained colors in the transfer stage. Note that given the change in reward structures, choosing the trained colors in the transfer stage was less rewarding than switching to choose the other colors in the first choice (see Figure 5). Error bars indicate standard errors.

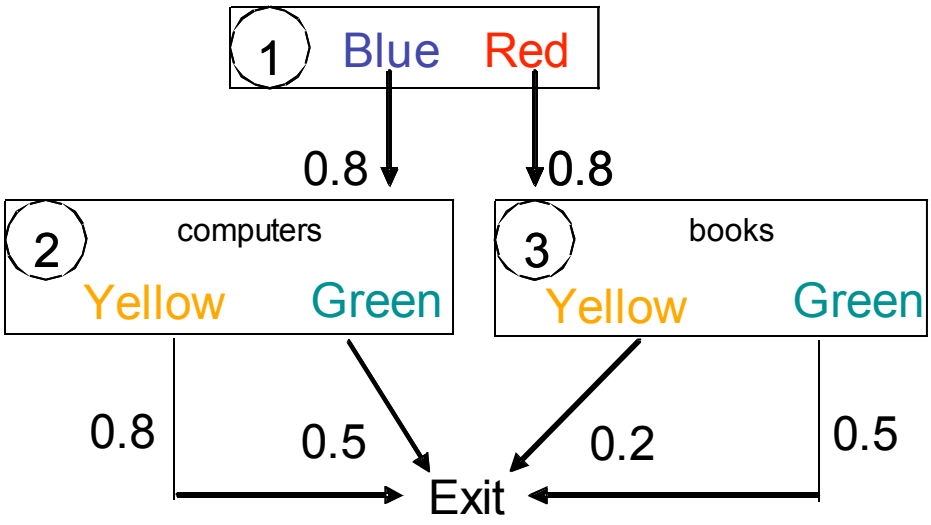

1
Red
Blue
0.8
0.8
2
computers
Yellow
Green
3
books
Yellow
Green
0.6
0.4
0.2
0.4
Exit

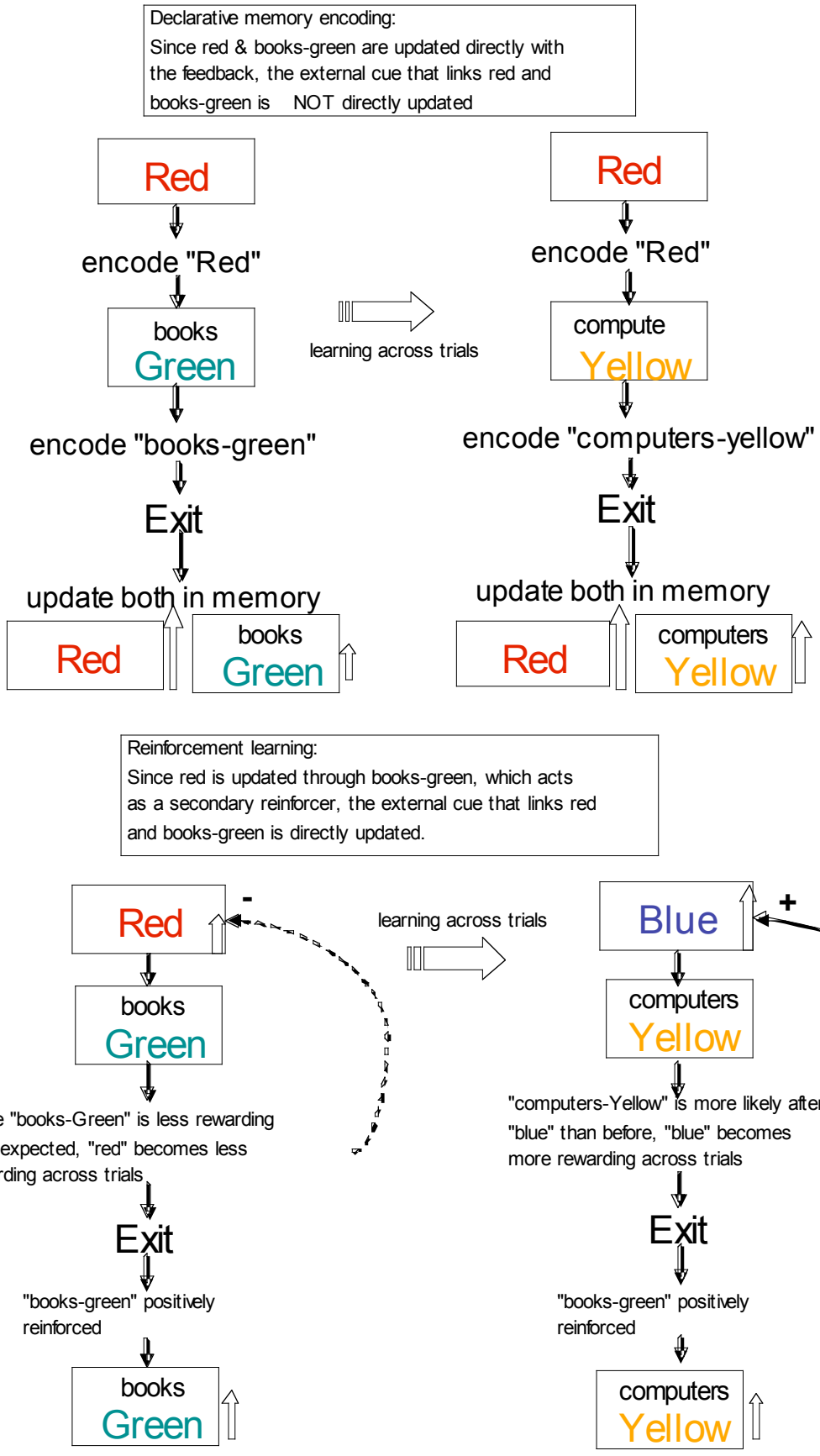

Red
encode "Red"
computers
Yellow
encode "computers-yellow"
Exit
update both in memory
Red
computers
Yellow
learning across trials
Red
encode "Red"
books
Green
encode "books-green"
Exit
update both in memory
Red
books
Green
(a) Learning by declarative memory encoding:
first choice is learned faster due to design of the task
Trials of learning
Eventual Outcomes of learning
Red
Blue
computers
Yellow
computers
Green
books
Yellow
books
Green
Red
computers
Yellow
"computers" is neutral
as expected, no learning
Exit
"computers-yellow" positively reinforced
to act as seondary reinforcer in later trials
computers
Yellow
learning across trials
Red
computers
Yellow
"computers" is positively reinforced
and acts as a secondary reinforcer:
"Red" is positively reinforced
Exit
Further reinforce
"computers-yellow"
computers
Yellow
+
Trials of learning
(b) Non-declarative reinforcement learning by feedback propagation:
second choice is learned faster as it is closer to the feedback

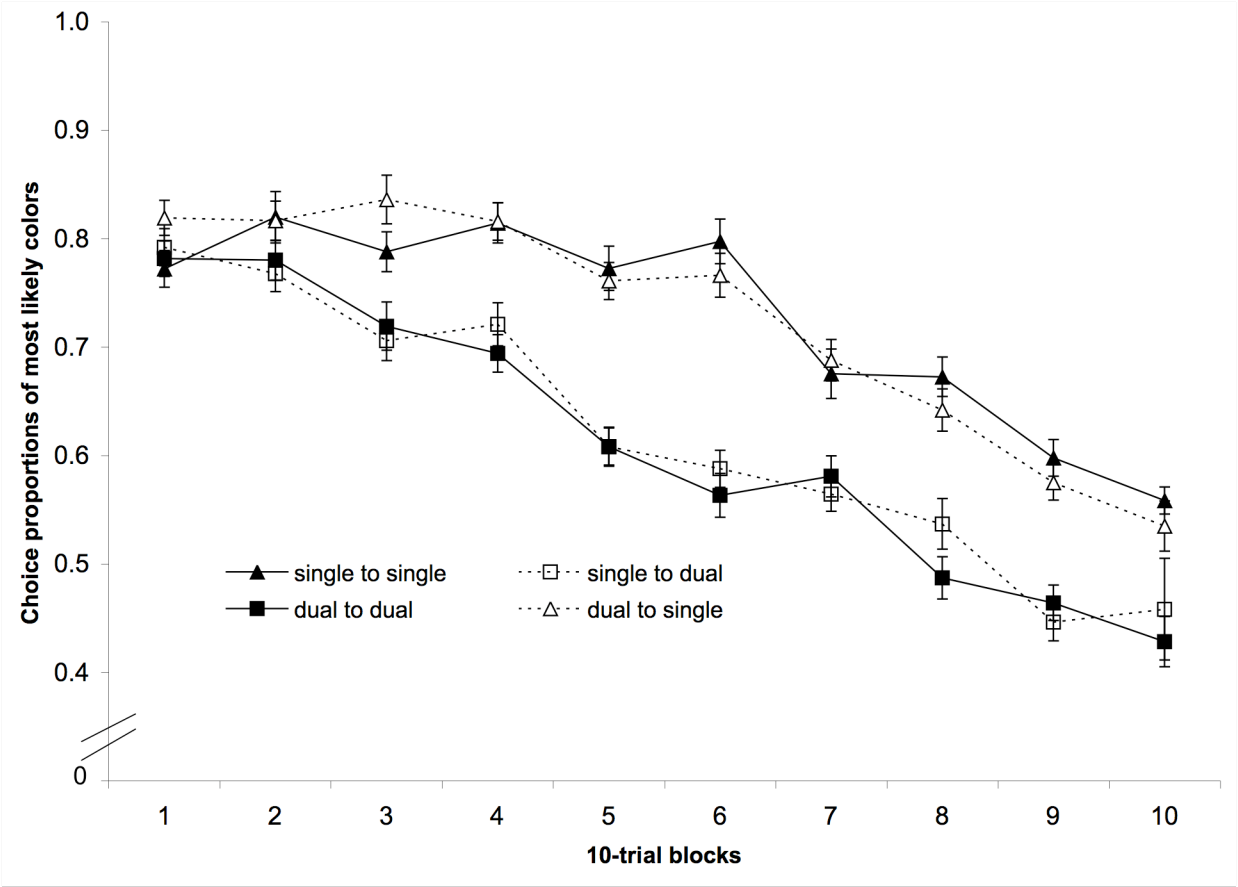

Single
Choice proportions of most likely colors
1.0
0.8
0.6
0.4
0.2
0.0
ambiguous first
ambiguous second
distinct first
distinct second
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
10-trial blocks


Dual
Choice proportions of most likely colors
1.0
0.8
0.6
0.4
0.2
0.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
10-trial blocks


Choice Prop of the same color
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Early
Late
First
Second
First
Second
First
Second
First
Second
Distinct
Ambiguous
Distinct
Ambiguous
Single
Dual


1
Blue
Red
0.8
0.8
2
computers
Yellow
Green
3
books
Yellow
Green
0.8
0.5
0.2
0.5
Exit


Declarative memory encoding:
Since red & books-green are updated directly with the feedback, the external cue that links red and books-green is NOT directly updated
Red
encode "Red"
books
Green
encode "books-green"
Exit
update both in memory
Red
books
Green
learning across trials
Red
encode "Red"
compute
Yellow
encode "computers-yellow"
Exit
update both in memory
Red
computers
Yellow
Reinforcement learning:
Since red is updated through books-green, which acts as a secondary reinforcer, the external cue that links red and books-green is directly updated.
Red
-
books
Green
Since "books-Green" is less rewarding than expected, "red" becomes less rewarding across trials
Exit
"books-green" positively reinforced
books
Green
learning across trials
Blue
+
computers
Yellow
"computers-Yellow" is more likely after "blue" than before, "blue" becomes more rewarding across trials
Exit
"books-green" positively reinforced
computers
Yellow


1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
Choice proportions of most likely colors
single to single
single to dual
dual to dual
dual to single
1
2
3
4
5
6
7
8
9
10
10-trial blocks