Psychological Research (2008) 72:321–330
DOI 10.1007/s00426-007-0113-7

ORIGINAL ARTICLE

Solving the credit assignment problem: explicit and implicit learning of action sequences with probabilistic outcomes

Wai-Tat Fu · John R. Anderson

Received: 30 April 2006 / Accepted: 20 February 2007 / Published online: 20 April 2007

Abstract In most problem-solving activities, feedback is received at the end of an action sequence. This creates a credit-assignment problem where the learner must associate the feedback with earlier actions, and the interdependencies of actions require the learner to remember past choices of actions. In two studies, we investigated the nature of explicit and implicit learning processes in the credit-assignment problem using a probabilistic sequential choice task with and without a secondary memory task. We found that when explicit learning was dominant, learning was faster to select the better option in their first choices than in the last choices. When implicit reinforcement learning was dominant, learning was faster to select the better option in their last choices than in their first choices. Consistent with the probability-learning and sequence-learning literature, the results show that credit assignment involves two processes: an explicit memory encoding process that requires memory rehearsals and an implicit reinforcement-learning process that propagates credits backwards to previous choices.

Introduction

Some of the most difficult situations in skill learning occur when the learner has to perform a sequence of actions but only receives feedback on their success at the end of the sequence. Outcomes of action sequences are often probabilistic, so that learning has to occur through accumulation of experiences. Learning action sequences with probabilistic outcomes creates a credit-assignment problem, in which the learner has to assign credits to earlier actions that are responsible for eventual success. The credit-assignment problem is even more difficult when the actions are interdependent, and the environment may change both autonomously and as a result of the actions. In two experiments, we study how people learn to solve the credit-assignment problem in a simple but challenging example of such a situation. Our study is developed based on the recent proposal that humans exhibit two distinct processes when learning action sequences with delayed feedback: an explicit memory encoding process that requires attentional resources to encode the actions and their outcomes, and an implicit reinforcement-learning process that requires little attentional resources. The main goal of our experiments is to study the nature of these two processes when people learn to choose action sequences with probabilistic outcomes. We investigate how people solve the credit assignment problem 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 between actions and their outcomes is related to the research on probability learning and probabilistic classification. We will first review research in these two areas. We will then show how the credit assignment problem is related to the theory of reinforcement learning. These ideas are then integrated to guide the design of our two experiments.

Sequence learning

The explicit and implicit learning distinction has often been investigated through a paradigm called sequence learning

W.-T. Fu (✉)
Human Factors Division and Beckman Institute,
University of Illinois at Urbana-Champaign,
405 North Mathews Avenue, Urbana, IL 61801, USA
e-mail: wfu@uiuc.edu

J. R. Anderson
Department of Psychology, Carnegie Mellon University,
Pittsburgh, PA, USA

(e.g., Cleeremans & McClelland, 1991; Cohen, Ivry, & Keele, 1990; Curran & Keele, 1993; Frensch, Buchner, & Lin, 1994; Jimenez, Mendez, & Cleeremans, 1996; Nissen & Bullemer, 1987; Perruchet & Amorim, 1992; Stadler, 1995; Sun, Slusarz, & Terry, 2005; Willingham, Nissen, & Bullemer, 1989). One typical paradigm is the serial reaction time (SRT) task in which subjects have to press a sequence of keys as indicated by a sequence of lights. A certain pattern of button presses recurs regularly and subjects give evidence of learning this sequence by pressing the keys for this sequence faster than a random sequence. Although there have been slightly different definitions to capture the details of the implicit/explicit distinction, the key factor is the idea that implicit learning occurs as a facilitation of test performance without concomitant awareness of what is being learned (e.g., Frensch, 1998; Shanks & St. John, 1994; Reber, 1989; Willingham, 1998). For example, learners in the SRT task who show faster response time in structured sequences are often unable to explicitly verbalize knowledge of the sequence structure.

A number of studies have used a secondary task such as counting of tones to study the effects of diminished attention for implicit learning (e.g., Cohen et al., 1990; Curran & Keele, 1993; Frensch et al., 1994; Nissen & Bullemer, 1987; Reed & Johnson, 1994; Stadler, 1995). Cohen et al. (1990) found that when attention is diminished by a secondary task, subjects could only learn simple pair-wise transitions, but failed to learn higher order hierarchical structures in the sequence. Although, studies show that sequences with higher order hierarchical structures can be learned (e.g., Curran & Keele, 1993; Frensch et al., 1994), the secondary task consistently reduces the amount of learning relative to the single task condition (Stadler, 1995). Researchers have proposed that the secondary task possibly disrupts learning by interfering with the ability to associate stimuli in short-term memory (Frensch et al., 1994; Stadler, 1995), thus diminishing implicit learning of higher-order structures in complex sequences.

An interesting paradigm studying the response-effect mapping in implicit learning was by Ziessler and his collegues (1994, 1998; Ziessler & Nattkemper, 2001; Ziessler, Nattkemper, & Frensch, 2004). In this paradigm, regularities were introduced between a response (a keypress) and its effect on the location of the next stimulus. Ziessler and his colleagues found that subjects learned these regularities implicitly and utilized the acquired implicit knowledge of response-effect regularities to anticipate the next stimulus location to facilitate performance. The studies show that learning and anticipating the outcome of a response could be implicit and does not require explicit encoding of previous action–outcome pair, and that regularities in response-effect relationships are useful in acquisition of sequence knowledge.

Although, the majority of the sequence-learning studies have used deterministic sequences, some have used probabilistic sequences in their experiments and have identified interesting properties of implicit learning (Cleeremans & McClelland, 1991; Stadler, 1992; Schvaneveldt & Gomez, 1998). For example, instead of using structured and random sequences, Schvaneveldt and Gomez used a probable (more likely to occur) and an improbable sequence in a SRT paradigm. They found that people were faster at responding to probable than the improbable sequence. The presence of a secondary task did not show significant differences in the magnitude of learning; however, anticipatory errors on improbable sequences were much more frequent in the single task condition. The difference in errors suggests that there are differences in the learning processes under single- and dual-task conditions. In addition, Schvaneveldt and Gomez found that when transferring subjects trained in the single-task condition to the dual-task condition, there was no significance difference in reaction time or error rates between the probable and improbable sequences. Schvaneveldt and Gomez concluded that attentional resources are required to apply what is learned under single-task conditions, suggesting that there may be different modes of learning as elicited by the different levels of attentional demand in the single- and dual-task condition.

Probability learning and probabilistic classification

Although, the sequence-learning studies have provided important information about the distinction between explicit and implicit learning, they are designed to answer questions that are different from ours, namely, how people learn the probabilistic relationship between actions and their outcomes. There have been numerous studies on the learning of the probabilistic relationship between choices and their consequences. The simplest situation is the probability-learning experiment in which subjects guess which of the alternatives occurs and then receives feedback on their guesses (e.g., Estes, 1964). One robust finding is that subjects often "probability match"; that is, they will choose a particular alternative with the same probability that it is reinforced (e.g., Friedman et al., 1964). This leads many to propose that probability matching is the result of an implicit habit-learning mechanism that accumulates information about the probabilistic structure of the environment (e.g., Graybiel, 1995). One important characteristic of this kind of habit learning is that information is acquired gradually across many trials, and seems to be independent of declarative memory as amnesic patients were found to perform normally in a probabilistic classification task (Knowlton, Squire, & Gluck, 1994; but see Gallistel, 2005). However, for non-amnesic human subjects, it is difficult to determine whether this kind of probabilistic classification is independent of the use of declarative memory. Since declarative memory is dominant in humans, it has been argued that learners often initially engage in explicit 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 subjects give up the idea of explicitly encoding patterns and the implicit habit-learning process becomes dominant (Estes, 2002; Vulkan, 2000). Similarly, recent research on complex category learning has also provided interesting results suggesting multiple learning systems (Allen & Brooks, 1991; Ashby, Queller, & Berretty, 1999; Waldron & Ashby, 2001). For example, Waldron and Ashby (2001) showed that although a concurrent Stroop task significantly impaired learning of an explicit rule that distinguished between categories by a single dimension, it did not significantly delay learning of an implicit rule that requires integration of information from multiple dimensions. The results, consistent with those from the sequence-learning studies, have led many to propose the explicit and implicit modes of learning. One common way to distinguish between the two learning modes is to introduce the secondary distractor task to suppress the otherwise dominant explicit learning mode, so that the different properties of the implicit learning mode can be studied.

In both the sequence-learning and the probability-learning paradigms, subjects 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. Subjects 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 reflects the complexity of the credit-assignment problem in situations in which people learn to sequentially choose actions with probabilistic outcomes and receive feedback only after the whole action sequence is executed. Our studies are designed by combining research from both areas by studying how people learn to assign credits to different actions in a probabilistic sequential choice task. In this task, a sequence of actions is executed before feedback on its correctness is received, and a particular action sequence is correct only with a certain probability. Our goal is that the novel paradigm we used will contribute to the understanding of the nature of the explicit and implicit learning modes in the general context of skill learning when the learner has to choose the right action sequences that accomplish a task.

Implicit reinforcement learning and explicit memory encoding

As discussed earlier, previous research in sequence learning and probability learning leads us to expect that people may exhibit an implicit and explicit mode of learning in our probabilistic sequential choice tasks. In addition, as we will discuss next, recent research in neuroscience suggests that the implicit mode of learning is associated with the reinforcement-learning process, and the explicit mode of learning is associated with memory encoding of action sequences and their outcomes (e.g., Daw, Niv, & Dayan, 2005; Grafton, Hazeltine, & Ivry, 1995; Keele, Ivry, Mayr, Hazeltine, & Heuer, 2003; Packard & Knowlton, 2002; Poldrack et al., 2001). For example, Daw et al. (2005) described the implicit learning process as a habit-formation process and the explicit learning process as a goal-directed "tree-searching" memory encoding process, and showed that these two processes together account for a wide set of animal choice behavior. We will show that our task is designed such that these two modes of learning will lead to very different predictions of behavior. However, we will first review recent research in reinforcement learning that serves as the basis of our predictions.

Converging evidence have shown that structures in the basal ganglia is closely related to the habit-learning and procedural system in which past response-outcome information is accumulated through experience (e.g., Fu & Anderson, in press; Graybiel, 1995; Schultz, Dayan, & Montague, 1997), and the prefrontal cortex and the medial temporal lobe is related to the declarative system that associates with more reflexive and goal-directed activities. Recent research has also shown that the neural activities in the basal ganglia correlate well with the predictions of reinforcement learning (e.g., Schultz et al., 1997), and are distinct from the activities in the declarative memory system (Packard & Knowlton, 2002; Poldrack et al., 2001). The basic prediction of reinforcement learning (e.g., see Sutton & Barto, 1998) is that when feedback is received after a sequence of actions, only the last action in the sequence will receive feedback but that on later trials its value will then propagate back to early actions. By itself this mechanism cannot learn in cases where success depends on the sequence of actions rather than the individual actions. Memories of previous actions or observations are required to disambiguate the states of the world (e.g., McCallum, 1995). This implies that the cognitive agent needs to explicitly adopt some forms of a memory encoding strategy to retain relevant information in memory for future choices. An explicit memory encoding strategy allows chaining together of action sequence in working memory, so that the outcome of the whole action sequence can be observed. Practically, this strategy implements a "tree-searching" procedure, in which the outcome of a branching set of possible response-outcome situations is tested. The advantage of the explicit memory encoding strategy is that the success of an action sequence can be evaluated and updated with a single feedback. In contrast, reinforcement learning requires the backward propagation of credits to earlier actions, and is thus less efficient. On the other hand, maintaining action sequences in memory puts a high demand on attentional resources. It is therefore expected that a secondary task will significantly hamper the explicit memory encoding strategy. Since the habit-learning mechanism is often thought to be less dependent on attentional resources (e.g., Packard & Knowlton, 2002), it is expected that reinforcement learning is still effective even when attention is diminished by the secondary task.

In two experiments, we study the nature of the learning processes in a probabilistic sequential choice task with and without a secondary distractor task. The sequential choice task is specifically designed to distinguish between explicit and implicit learning processes and we have strong predictions about the outcome: when the implicit reinforcement learning process is dominant, learning will be in the backward direction, i.e., learning of items closer to the feedback will be faster than those farther away. When the explicit memory encoding process is dominant, learning of items will be in the forward direction, i.e., learning of items presented earlier will be faster.

Experiment 1

Subjects were presented with two consecutive choice sets. In each choice set, subjects were asked to choose one of the two colors (one of the boxes in Fig. 1) presented on a computer screen, and the probabilities that the options were correct were independent between the two choices. Subjects were told to imagine that the two colors were on the two sides of a biased coin that was flipped and the side that turned up would be the correct color for that trial. Subjects used the arrow buttons on a standard US keyboard to choose one of the colors. Subjects were instructed to press the left arrow button to select the color on the left side of the screen, and to press the right arrow button to select the color of the right side of the screen. After subjects made the first response, the second choice set were presented immediately (<50 ms). After the second response, a feedback message was presented immediately (<50 ms) on the screen to inform subject as to whether both of the two choices had led to success. Subjects were instructed that if the feedback was "Correct," both choices made were correct; but if the feedback was "Wrong," then either one of the choices was wrong or both choices were wrong, and they would never receive any feedback on the correctness of the individual choices in this case.

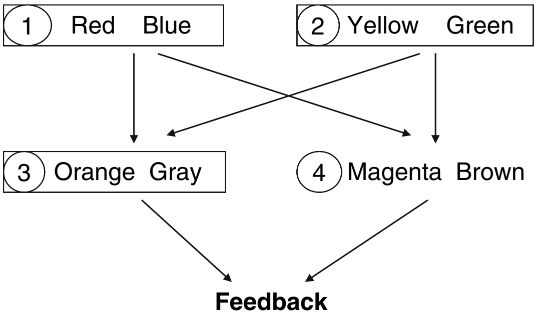

Fig. 1 The probabilistic sequential choice task in Experiments 1 and 2. Each *box* represents a choice set, and the actual colors were randomly selected from a fixed set of colors (*red*, *green*, *yellow*, *blue*, *brown*, *gray*, *magenta*, and *orange*). Four randomly selected choice sets were selected for each subject. Subjects were presented with two choice sets in each trial. The first choice set was randomly selected from (*Red*, *Blue*) or (*Yellow*, *Green*), and the second choice set is randomly selected from (*Orange*, *Gray*) or (*Magenta*, *Brown*). In each choice set, the probability that one of the colors was correct was manipulated differently in the two experiments. See text for details

In this task, four choice sets, each with two colors, were constructed by randomly selecting from eight colors (red**,** green**,** yellow**,** blue**,** brown**,** gray**,** magenta, and orange). In each choice set, one of the colors was randomly selected to be the color that was more likely to be correct. The choice sets were randomly divided into two groups of two. On each trial, the first pair was randomly chosen from the first group (i.e., either 1 or 2 in Fig. 1) and the second pair from the second set (i.e., either 3 or 4 in Fig. 1). Thus, a particular subject might either see the pair red and blue or the pair yellow and green as the first choice and the pair orange and gray or the pair magenta and brown as the second choice. For that subject, one randomly chosen member of each pair would be correct on 80% of the trials and the other on 20%.

When engaged in explicit memory encoding, subjects had to keep previous choices and their outcomes in memory and essentially identify the correct first choices and the correct second choices. Since these choices were equal and independent one might expect equal learning of first and second choice. However, given evidence of strong primacy effects in sequential learning tasks involving explicit memory encoding (e.g., Drewnowski & Murdock, 1980; Ward, 1937), our expectation was that the first choices would tend to be learned first.

To study the impact of a secondary task, we introduced a "2-back" task to suppress the otherwise dominant explicit memory encoding process. The secondary task required subjects to listen to a continuous stream of numbers (from 0 to 9) from the speakers. Starting from the third number, subjects had to press the control key on the keyboard if the number is 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 2 s. Subjects had to maintain their performance at 80% or better at the 2-back task while performing the sequential choice task. From earlier discussion, the basic prediction of the implicit process is that actions close to the feedback will acquire value first and then their value will propagate back to early actions. Thus, learning of the second choice will be faster than the first choice in the dual-task condition.

Method

About 60 subjects in the Carnegie Mellon University and University of Illinois community were recruited for the experiment. Subjects received a base payment of $8 plus a bonus payment of up to $7 depending on performance. Half of the subjects were assigned to the single task condition and the other half to the dual task condition. Subjects started with an initial score of ten points. For each correct choice, five points would be added to the final score; for each wrong choice, one point would be deducted from the final score. Subjects completed ten 40-trial blocks. Subjects were paid one cent for each point in the total score for the bonus payment. At the end of the experiment, subjects were asked to write down any strategy they used and whether they were aware of any patterns in the probabilistic sequential choice task.

Results

Figure 3 shows the mean choice proportions of the more likely colors in each 40-trial block. The main effect of condition (single/dual) was significant [$F(1,58) = 29.97$, MSE = 12.51, $P < 0.001$]. Consistent with previous results, subjects in the single-task condition performed better than those in the dual-task condition, showing that the secondary task impairs overall performance. There was a significant effect of block [$F(9,522) = 37.87$, MSE = 0.54, $P < 0.001$] confirming the apparent learning trend in Fig. 2. There was no overall difference between learning of the first and second choice [$F(1,58) = 1.23$, MSE = 0.0074]. However, the interaction between condition and choice was significant [$F(1,58) = 5.15$, MSE = 0.31, $P < 0.05$]. The only other significant interaction was that between condition, block, and choice [$F(9,522) = 3.85$, MSE = 0.0047, $P < 0.001$]. This reflects the fact, apparent in Fig. 3 that the interaction effect between condition and choice only appears in the early blocks. The average performance over the first two blocks of the subjects in the single-task condition was higher for the first choice [$t(28) = 2.44$, $P < 0.05$] while the reverse was true for the first three blocks in the dual-task condition of the not aware group [$t(22) = 2.55$, $P < 0.05$]. All choices were significantly above chance in the last four blocks of trials.

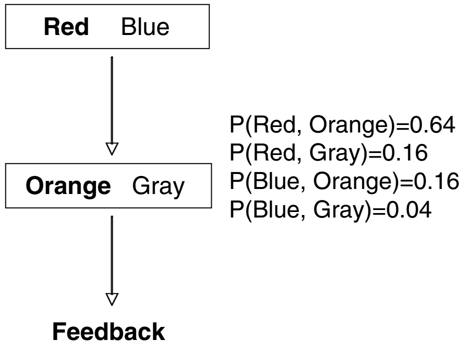

Fig. 2 An example of the choice sets presented in Experiment 1 and the probabilities for each of the two choices made being correct. The *bolded* color is the more likely color in the choice set

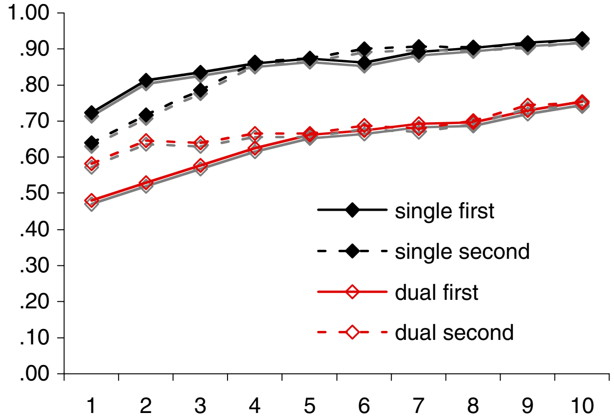

Fig. 3 The mean choice proportions of the more likely colors in Experiment 1 in each 40-trial block

The results were consistent with our predictions of the two learning processes in the task. Because of the primacy effect of sequence memory, subjects in the Single-Task condition learned the first choice faster than the second choice. Consistent with the reinforcement-learning mechanism, learning of the second choice was faster than the first choice.

We also examined the self-reports by the subjects after the experiment to assess whether there were any differences in their knowledge of the task. Our hypothesis was that subjects in the single task condition were more likely to engage in the explicit memory encoding and were therefore more likely to be aware of the structure of the task. Since the dual task required significant working memory resources, explicit memory encoding would be much more difficult. Therefore, in the dual-task condition, subjects were more likely engaged in implicit reinforcement learning.

We counted the number of subjects who could report at least one of the more likely colors in both choices and assigned them to the "aware group" and the rest to the "not aware" group. By this criterion, we found that there were 22 and 7 subjects in the aware group in the single- and dual-task conditions, respectively. This was found be to significantly different from chance [$\chi^2(1) = 15.07$, $P < 0.01$], suggesting that the task knowledge acquired by the subjects were significantly different between the two conditions. However, since we did not assess subject's knowledge during the early part of the task, this post hoc comparison does not allow us to infer whether this difference in task knowledge existed in the early blocks of learning.

We were also interested in how subjects in different conditions would change their choices after feedback. In particular, we calculated the choice proportions of the more likely colors according to the feedback they received in the last trial. Our prediction was that since the explicit "tree-searching" strategy was dominant in the single-task condition, subjects would learn the first choice faster than the second choice during the early trials, and as a consequence, they would be more likely to switch to a different color in the second choice than in the first choice. However, this difference should disappear at later blocks after subjects learned both choices. In addition, since explicit learning would encode in memory the more likely colors, subjects should be more likely to continue choosing the more likely colors even after a "wrong" feedback. On the other hand, since implicit reinforcement learning was mostly driven by the external feedback, we predicted that subjects would be more likely to switch to a different color after a "wrong" feedback.

Figure 4 shows the choice proportions of the more likely colors in the first two (Early) and the last two (Late) blocks of trials in the single- and dual-task conditions. The main effects of condition, blocks (Early/Late), choice (first/second), and feedback (correct/wrong) were significant [$F(1,58) = 26.3$, MSE = 1.87, $P < 0.001$, $F(1,58) = 49.67$, MSE = 3.0, $P < 0.001$, $F(1,58) = 9.95$, MSE = 0.70, $P < 0.01$, and $F(1,58) = 42.98$, MSE = 2.22, $P < 0.001$, respectively]. The interaction between condition and feedback was significant [$F(1,58) = 26.60$, MSE = 1.37, $P < 0.001$]. The difference between condition when the last feedback was "correct" was not significant [$t(58) = 0.60$], but that when the last feedback was "wrong" was significant. Subjects in the dual-task condition switched to a different color when the last feedback was "wrong" significantly more often than those in the single-task condition [$t(58) = 5.24$, $P < 0.001$], confirming our prediction that implicit reinforcement learning was more sensitive to immediate feedback than explicit memory encoding. Indeed, this difference was significant in both early [$t(58) = 2.60$, $P < 0.05$] and late blocks [$t(58) = 6.05$, $P < 0.05$].

The interaction between condition and choice and that between choice and blocks were not significant [$F(1,58) = 2.00$, MSE = 0.14 and $F(1,58) = 0.19$, MSE = 0.001, respectively], but the interaction three-way interaction between condition, choice, and blocks was significant [$F(1,58) = 6.94$, MSE = 0.50, $P < 0.05$]. Only in the single-task condition subjects chose the more likely colors in the first choice more often in the early blocks [$t(58) = 3.69$, $P < 0.05$], although this difference was gone in the late blocks [$t(58) = 0.41$], confirming our prediction that the explicit "tree-searching" strategy would learn the first choice faster than the second choice. No other interaction was significant.

To summarize our results, we found that subjects in the single-task condition performed better and acquired more task knowledge than subjects in the dual-task condition. Most importantly, we found that in the single-task condition, subjects learned in the forward direction but subjects in the dual-task condition learned in the backward direction. We also found that subjects in the dual-task condition were more sensitive to the immediate feedback. These findings were consistent with the predictions that the explicit memory encoding process is dominant in the single-task condition while the implicit reinforcement learning process is dominant in the dual-task condition.

Experiment 2

Since the two choices were independent in Experiment 1, implicit reinforcement learning was possible as learning

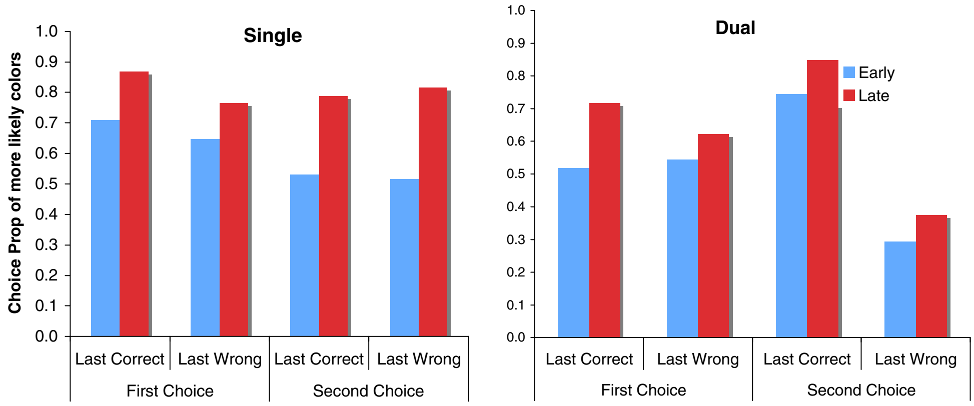

Fig. 4 The choice proportions of the more likely colors broken down by the first/second choice and feedback received in the previous trial (correct/wrong) during the first two (Early) and last two (Late) 40-trial blocks in the single- and dual-task condition of Experiment 1

of the second choice did not depend on memory of the first choice. In Experiment 2, we designed the task such that the choices were dependent. For example, in Fig. 1 when (Red, Blue) are presented in the first choice set, the more likely color would be Orange or Magenta in the second choice; but when (Yellow, Green) are presented as the first choice set, the more likely color would be Gray or Brown. Although, subjects could learn the first color, learning of the second color required memory of the first choice set. For example, given the second choice set of the colors Orange and Gray, the more likely color depends on whether the first choice set was (Red, Blue) or (Yellow, Green). This should not create any problem for subjects who learn this dependency by explicit memory encoding, because the first choice set is already encoded into memory. However, the implicit reinforcement learning process would not be possible to learn this dependency, as reinforcement learning could only learn an earlier choice after a later one was learned. Without memory of the first choice, the colors in the second choice will be equally likely to be correct and thus could not be distinguished. When learning of the second choice fails, propagating credits backward would be impossible (Fig. 5).

Method

About 60 subjects in the Carnegie Mellon University and University of Illinois community were recruited for the experiment. Half of the subjects were assigned to the single task condition and the other half to the dual task condition. Subjects completed ten 40-trial blocks. The payment scheme was the same as that in Experiment 1.

Results

Figure 6 shows the mean choice proportions of the more likely colors in each 40-trial block. The main effect of condition was significant [$F(1,58) = 38.39$, MSE = 8.04, $P < 0.001$]. As in Experiment 1, subjects in the single-task condition performed better than those in the dual-task condition. The main effect of block was significant [$F(9,522) = 5.50$, MSE = 0.009, $P < 0.001$], as well as the difference between learning of the first and second choice [$F(1,58) = 24.95$, MSE = 2.34, $P < 0.001$]. However, the interactions between condition and block was significant [$F(9,522) = 2.08$, MSE = 0.003, $P < 0.05$], as well as that between condition and choice [$F(1,58) = 9.89$, MSE = 0.925, $P < 0.01$]. The interaction between choice and block was also significant [$F(9,522) = 2.65$, MSE = 0.003, $P < 0.001$], as well as the three-way interaction between condition, choice, and block [$F(9,522) = 4.10$, MSE = 0.004, $P < 0.001$]. The significant interactions were due to the large differential difference between the choices in the two conditions. In fact, in the single-task condition, the main effects of choice and block were significant [$F(1,29) = 22.89$, MSE = 3.1, $P < 0.001$ and $F(9,261) = 6.19$, MSE = 0.11, $P < 0.001$, respectively]. The choice by block interaction was also significant [$F(9,261) = 4.89$, MSE = 0.007, $P < 0.001$]. On the other hand, the main effects of choice and block were not significant [$F(1,29) = 3.1$, MSE = 0.16 and $F(9,261) = 0.89$, MSE = 0.002], nor their interaction [$F(9,261) = 0.76$, MSE = 0.0003]. There was therefore no significant learning in both choices in the dual-task condition. In fact, none of the choices in the dual-task condition was above chance. In the single-task condition, both choices were significantly above chance in the last five blocks, suggesting subjects learned the dependencies in the single-task condition, and they learned the first choice faster than the second choice.

The dependency between choices in Experiment 2 implied that learning to choose the more likely color in the second choice required memory of the first choice. About 22 subjects in the single-task condition and four subjects in the dual-task condition were aware of the more likely colors in both choices and learned to choose

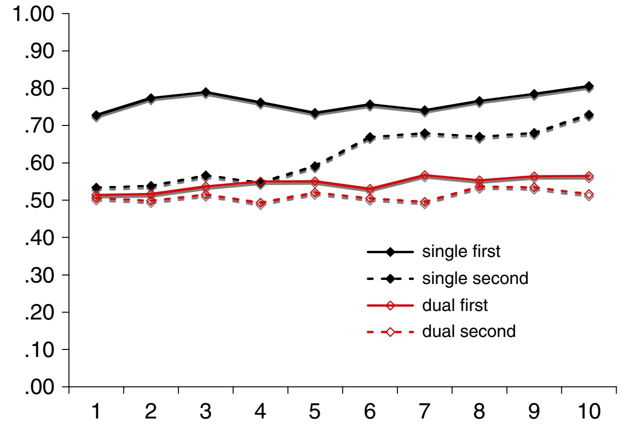

Fig. 6 The mean choice proportions of the more likely colors in Experiment 2 in each 40-trial block

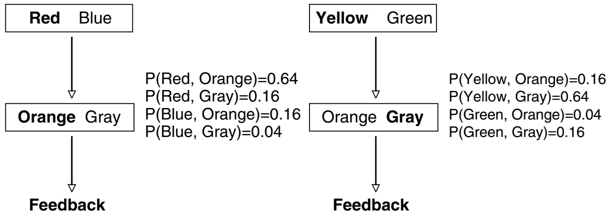

Fig. 5 Two examples of the choice sets presented in Experiment 2 and the probabilities for each of the two choices made being correct in each of the example. The *bolded* color is the more likely color in the choice set. In Experiment 2, the more likely color in the second choice set is dependent on the first choice set

them increasingly often across trials. This suggests that subjects in the aware group did manage to learn the dependency of choices. Similar to the same condition in Experiment 1, learning of the first choice was faster, but unlike the results from Experiment 1, the difference between the first and second choice remained large even after 400 trials. The slower learning of the second choice in Experiment 2 could be explained by its dependency on the first choice: the amount of experience for the most likely color in the second choice was half of those for the first choice.

As in Experiment 1, we analyzed the self-reports given the subjects at the end of the experiment to compare whether there was any difference in the acquisition of task knowledge between the two conditions. Subjects who wrote down at least one of the color combinations that were more likely to be correct were placed in the aware group, otherwise they were put in the not aware group. As a result, 27 and 4 subjects were placed in the aware group in the single and dual-task condition, respectively, the rest were placed in the not aware group. The difference between condition was again significant [$\chi^2(1) = 18.07$, $P < 0.01$].

Similar to Experiment 1, we calculated the choice proportions of the more likely colors in cases where the feedback from the last trial was "correct" and "wrong" (see Fig. 7). The main effects of condition, choice, and feedback were significant [$F(1,58) = 24.86$, MSE = 1.69, $P < 0.001$, $F(1,58) = 6.87$, MSE = 0.49, $P < 0.05$, and $F(1,58) = 4.92$, MSE = 0.42, $P < 0.05$, respectively], but the main effect of blocks was not significant [$F(1,58) = 1.00$, MSE = 0.009]. Subjects chose the more likely colors more often in the single-task condition, in the first choice, and when the feedback from the previous trial was "correct." The interactions between condition and blocks, condition and choice, and condition and feedback were significant [$F(1,58) = 6.30$, MSE = 0.58, $P < 0.05$, $F(1,58) = 8.74$, MSE = 0.62, $P < 0.05$, and $F(1,58) = 3.86$, MSE = 0.33, $P < 0.05$, respectively]. The three-way interaction among condition, choice, and blocks was also significant [$F(1,58) = 5.10$, MSE = 0.34, $P < 0.05$]. All significant differences were in the single-task condition, and none of the differences in the dual-task condition was significant. As in Experiment 1, subjects in the single-task condition chose the more likely colors more often in the first choice than in the second choice in early blocks [$t(58) = 4.20$, $P < 0.05$] but the difference was not significant in late blocks [$t(58) = 1.33$], confirming the prediction that subjects learned the first choice faster than the second choice during the early blocks, but later the different disappeared when the learned both choices in late blocks. They also tended to switch to a different color more often in the second choice than in the first choice when the feedback from the previous trial was "wrong" during the early blocks, but not in the late blocks. This was consistent with the prediction that in the single-task condition, when subjects learned which colors were more likely to be correct in late blocks, they could ignore the immediate feedback and continue to choose the more likely colors. In the dual task condition, the more likely colors in the second choice depended on the first choice. Without memory of the first choice, immediate feedback appeared random to the subjects when learning the second choice. The implicit reinforcement learning was therefore not effective to guide the selection of colors.

To summarize, the results from Experiment 2 were again consistent with the predictions of the two learning modes proposed earlier. In the single-task condition, explicit memory encoding was dominant, and because of the primacy effect and the sequential dependency, the first choice was learned faster than the second choice (in fact, in Experiment 2, it was impossible to learn the second choice before the first choice). On the other hand, the weak memory trace of the first choice significantly hampered the discovery of the dependency in the dual-task condition. Unlike Experiment 1, reinforcement learning by itself was not sufficient to learn the second choice in the current design because when the first choice was excluded, the chance that either one of the colors in the second choice would lead to a success was the same. As a result, none of the colors would be preferred to the others even after 400 trials. Since learning of the second choice was not possible, the credit could not be propagated back to the first choices. Thus, there was effectively no learning in both choices in the dual-task condition.

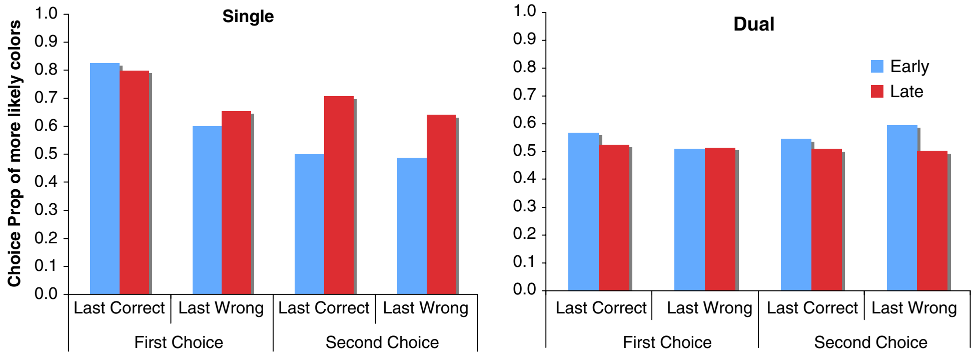

Fig. 7 The choice proportions of the more likely colors broken down by the first/second choice and feedback received in the previous trial (correct/wrong) during the first two (Early) and last two (Late) 40-trial blocks in the single- and dual-task condition of Experiment 2

Discussions

We developed a probabilistic sequential choice task to study how people learn action sequences with probabilistic outcomes with and without a secondary task. As we expected from previous research, we conclude that subjects in our task exhibited two modes of learning by showing that they learned the first choice faster in the single-task condition and learned the second choice faster in the dual-task condition. This finding was consistent with the explicit and implicit modes of learning: an explicit memory encoding mode that requires attentional resources for maintaining past choices and outcomes in memory, and an implicit reinforcement learning mode that is effective even with diminished attention. In Experiment 2, further support was found for the credit-assignment mechanism in this implicit reinforcement learning process–when choices are interdependent, the propagation of credits from the reward to earlier responses will be ineffective, as credits cannot be appropriately assigned to the corresponding choices.

Our findings are consistent with previous findings that people can learn some sequences even with the presence of a secondary task. When choices are dependent, however, we found that people fail to learn when explicit memory encoding is suppressed by the secondary task. This finding may seem to be different from the findings by Curran and Keele (1993) and Frensch et al. (1994), who found that even complex sequences are learned in the presence of a secondary task. The inconsistency is perhaps due to the inherent difference in the task and how learning is measured: in the typical sequence learning task, people follow a deterministic stimulus-response mapping, receive immediate feedback, and learning is measured by faster response time, but in our task, the mapping is determined by past outcomes, receive delayed feedback, and learning is measured by choice proportions.[1] It is therefore not clear whether our results can be directly compared to previous sequence-learning studies. In addition, in our studies as well as in others, only aggregate results were analyzed. It is possible that the smooth learning curves could arise out of a mixture of different learning functions (e.g., Haider & Frensch, 2002).

Previous studies suggest that the explicit and implicit modes of learning may either compete against each other (e.g., Poldrack et al., 2001), or independent of each other (e.g., Curran & Keele, 1993; Cleeremans, 1997). In fact, it is not necessary that in the current task the implicit learning process we identified be a single, modular process in the basal ganglia; rather it is possible that it may involves different areas such as the association cortex or even parts of the frontal cortex that process and represent the stimuli and outcomes. The current studies obviously were not designed to provide clear answer to this question. This question definitely requires further research to test and examine the interaction of these two modes of learning.

Our measure of awareness requires subjects to self-report their knowledge of the more likely colors in the task. This may lead to the issue of sensitivity of measure (Shanks & St. John, 1994). It is possible that people categorized in the not aware group might have adopted some form of explicit learning but the awareness test was not sensitive enough to detect. At this point of the research, our goal is to manipulate the attentional load by the secondary task to study how people learn probabilistic action sequences. Indeed, we found striking qualitative behavioral differences that correspond to changes in the availability of attentional resources (and the different levels of awareness). In future studies we plan to address how manipulations of attention load and other task variables may influence the two modes of learning to extend our results reported in this article.

The probabilistic sequential choice task used in the experiments, although simple, contains essential components in complex skill learning, in which a sequence of actions are performed before reinforcement on the full course of action is received. Solving the credit-assignment problem is crucial for learning in this kind of situation, as the delayed feedback has to be assigned to earlier actions that are responsible for the desirable or undesirable outcome. The reinforcement-learning process provides a straightforward explanation of how feedback propagates back to earlier actions. Initially, only the action that leads to outcome gets credit or blame. The next time some of that credit/blame propagates back to the previous actions. Eventually, credit/blame can find its way back to critical early actions in a long chain of productions leading to a reward. The effectiveness of this process, however, depends on whether the effects of these actions are independent of each other. When the actions are interdependent, memory of earlier actions are required to ensure the proper assignment of credits for effective cognitive skill learning.

Acknowledgments The current work is supported by a grant from the Office of Naval Research (N00014-99-1-0097). We want to thank Peter Frensch, Roger Schvaneveldt, Jennifer Ferris, and an anonymous reviewer for their useful comments on the manuscript.

References

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

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

[1] The measure of choice proportions is similar to the error measures in the studies by Schvaneveldt and Gomez (1998).

Cleeremans, A. (1997). Sequence learning in a dual-stimulus setting. *Psychological Research, 60*, 72–86.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120*, 235–253.

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.

Drewnowski, A., & Murdock, B. B. (1980). The role of auditory features in memory span for words. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 319–332.

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

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

Frensch, P., Buchner, A., & Lin, J. (1994). Implicit learning of unique and ambiguous transitions in the presence and absence of a secondary task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 567–584.

Frensch, P. (1998). One concept, multiple meanings: On how to define the concept of implicit learning. In: M. A. Stadler, & P. A. Frensch (Eds.), *Handbook of implicit learning* (pp. 47–104). Thousand Oaks: Sage Publications.

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, USA: 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*, 184–206.

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.

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

Haider, H., & Frensch, P. A. (2002). Why aggregated learning follows the power law of practice when individual learning does not: Comment on Rickard (1997, 1999), Delaney et al. (1998), and Palmeri (1999). *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*, 392–406.

Jimenez, L., Mendez, G., & Cleeremans, A. (1996). Direct and indirect measures of implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 948–969.

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

McCallum, A. K. (1995) Reinforcement learning with selective perception and hidden state, Ph.D. Thesis, University of Rochester.

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

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

Perruchet, P., & Amorim, P. A. (1992). Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory, and cognition, 18*, 785–800.

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.

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General 118*, 219–235.

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 585–594.

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

Schvaneveldt, R., & Gomez, R. (1998). Attention and probabilistic sequence learning. *Psychological Research, 61*, 175–190.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17*, 367–447.

Stadler, M. A. (1992). Statistical structure and implicit serial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 318–327.

Stadler, M. A. (1995). The role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 674–685.

Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review, 112*, 159–192.

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

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

Ward, L. B. (1937). Reminiscence and rote learning. *Psychological Monographs, 49*, 64.

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

Willingham, D. (1998). A neuropsychological theory of motor skill learning. *Psychological Review, 105*, 558–584.

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

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

Ziessler, M. (1994). The impact of motor responses on serial pattern learning. *Psychological Research, 57*, 30–41.

Ziessler, M. (1998). Response-effect learning as a major component of implicit serial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 962–978.

Ziessler, M., & Nattkemper, D. (2001). Learning of event sequences is based on response-effect learning: Further evidence from a serial reaction task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 595–613.

Ziessler, M., Nattkemper, D., & Frensch, P. A. (2004). The role of anticipation and intention in the learning of effects of self-performed actions. *Psychological Research, 68*, 163–175.