

ORIGINAL ARTICLE

Roger W. Schvaneveldt · Rebecca L. Gomez

Attention and probabilistic sequence learning

Received: 17 February 1997 / Accepted: 5 January 1998

Abstract Limitations of using fixed sequences of events in studies of learning in the sequential reaction-time task led us to develop a probabilistic version of the task. When sequences occur probabilistically, transitions usually follow a sequence, but with some small probability, events occur out of sequence. This variation on the paradigm provides new evidence associated with manipulations of attentional load. Most notably, single-task learning leads to particularly high error rates on improbable transitions, suggesting anticipation of the sequence. Dual-task learning shows sensitivity to the sequence (by reaction-time differences to probable and improbable transitions), but without inflated errors on improbable transitions. Sensitivity to the sequence and anticipatory errors disappeared when participants transferred from single-task learning to dual-task conditions, suggesting that what is learned with single-task practice cannot be applied under conditions of limited attention. When learners transferred from dual- to single-task conditions, sensitivity of RT to the sequence increased but anticipation errors remained the same, suggesting that attentional load limits performance, but not learning. Qualitative differences in performance result from variations in attentional resources, which may reflect different learning processes.

Introduction

Among the methods used to study implicit learning is the sequential reaction-time (SRT) task (Cleeremans &

McClelland, 1991; Curran & Keele, 1993; Cohen, Ivry, & Keele, 1990; Frensch, Buchner, & Lin, 1994; Lewicki, Hill, & Bizot, 1988; Nissen & Bullemer, 1987; Perruchet & Amorim, 1992; Perruchet, Gallego, & Savy, 1990; Reber & Squire, 1994; Reed & Johnson, 1994; Schvaneveldt & Gomez, 1996; Stadler, 1989, 1992, 1993; Willingham, Nissen, & Bullemer, 1989). In this task, participants are exposed to repeating event sequences. Typically, a target occurs in one of three to six locations on a computer screen as dictated by a pattern sequence. The participant's task is to press a key on the keyboard corresponding to the location of the target as quickly as possible. The sequence is presented continuously over many cycles in a manner that disguises distinguishing features, such as the beginning and ending of the sequence. Learning of the sequential pattern is measured by the disparity in reaction time (RT) for responding to events in the structured sequence in comparison to some control sequence. Practice with a structured sequence results in a greater decrease in RT as compared to practice with randomly generated sequence locations (e.g., Cohen et al., 1990; Nissen & Bullemer, 1987), and when participants transfer from structured to random sequence locations, RTs are slower (e.g., Curran & Keele, 1993). Learning in these tasks is called implicit because most learners are unable to reveal explicit knowledge of the sequence structure.

A number of studies have explored the role attention plays in SRT learning (Cohen et al., 1990; Curran & Keele, 1993; Frensch et al., 1994; Frensch, Lin, & Buchner, in press; Grafton, Hazeltine, & Ivry, 1995; Mayr, 1996; Nissen & Bullemer, 1987; Reed & Johnson, 1994; Stadler, 1995). Researchers often manipulate attention by imposing a secondary task of counting high- or low-pitched tones in a series of tones presented during performance of the SRT task.

In one of the earliest attentional studies, Nissen & Bullemer (1987) exposed participants to a repeating, tentorial sequence in an SRT task. Performance was measured under single-task, divided-attention (dual-task), and random-sequence conditions. Participants in the

R. W. Schvaneveldt (✉)
Psychology Department, 3452, Box 30001,
New Mexico State University,
Las Cruces, NM 88003, USA
Phone: (505) 646-1047; fax: (505) 646-6212;
e-mail: schvan@crl.nmsu.edu

R. L. Gomez
University of Arizona, Tucson, AZ, USA

single-task condition demonstrated learning of the sequence, as evidenced by improved RT relative to participants receiving random sequences. In the dual-task condition, high- and low-pitched tones accompanied the sequence learning task. Participants in this condition counted the number of low-pitched tones. Performance under dual-task conditions was no better than was responding to the random trials, suggesting that SRT learning was dependent on attentional processing.

Cohen et al. (1990) further investigated the relationship between attention and sequence learning by varying the level of complexity in structured sequences. The sequence used by Nissen and Bullemer (1987) consisted of second-order conditionals, so that predicting an item in sequence depended on knowing the prior two locations. Simpler sequences can be constructed using unique associations (or first-order conditionals) in which the location of an item is unambiguously predicted by the preceding item. Cohen et al. tested the hypothesis that more attention should be required for processing sequences with greater complexity by assessing the effect of a secondary task on learning sequences with increasing degrees of complexity; i.e., those consisting entirely of first-order conditionals (e.g., *A-E-B-D-C*), hybrid sequences consisting of both first- and second-order conditionals (e.g., *A-D-B-C-A-B*), and sequences consisting only of second-order conditionals (e.g., *A-B-C-B-A-C*). Participants in the single-task condition learned all three sequence types, but those in the dual-task condition learned only the first-order and hybrid sequences, suggesting that attention was required for learning second-order dependencies, but not for learning sequences of lesser complexity.

Subsequent studies have shown that although learning first-order conditionals is more pronounced than learning second-order conditionals (Frensch et al., 1994), participants can indeed learn second-order sequences under conditions of divided attention (Curran & Keele, 1993; Reed & Johnson, 1994; Frensch et al., 1994, in press; Stadler, 1995). One explanation for more robust learning in single-task compared to dual-task conditions is that attentional load interferes with learning. For instance, Frensch et al. (1994), proposed that the dual task lengthened the interval between the response to a stimulus and the presentation of the next stimulus, thus potentially interfering with the ability to associate consecutive stimuli in short-term memory and, subsequently, interfering with learning. Similarly, Stadler (1995) proposed that a concurrent task disrupts learning by interfering with the organization of the sequence. Finally, Cleeremans and McClelland (1991) successfully modeled the effects of attentional load by adding randomly distributed noise to the input of each unit in a connectionist network of SRT learning, leading them to suggest that attentional load interferes with human learning in a similar manner.

In contrast with the interference view of dual-task learning, Frensch et al. (in press) suggested that the dual task may not disrupt learning, but merely interferes with

the behavioral expression of what is learned. In support of this view, Frensch et al. manipulated practice schedules involving different amounts of single- and dual-task practice. Although single- and dual-task learners showed differences during acquisition, they showed similar degrees of sensitivity to the sequential structure when they were tested under the same conditions (single- or dual-task). This finding is important because it suggests that SRT learning is automatic, involuntary in nature, and robust, even in the face of attentional load. These data also suggest that it is important to distinguish between the degree of learning and the expression of the learning under various conditions of attentional load.

The results reported by Frensch et al. (in press) also lead them to suggest that only one system is involved in SRT learning. In contrast, Curran and Keele (1993) interpreted their findings to suggest the operation of independent, parallel learning modes. Curran and Keele first varied the degree of attentional learning under single-task conditions by manipulating whether participants learned a hybrid sequence intentionally or incidentally. Participants showed differences in learning as a function of how aware they were of the sequence in the single-task condition, but when the dual task was added, the previously observed learning differences disappeared, suggesting that attentional and nonattentional forms of learning might be operating in parallel. This finding was replicated in a second study by transferring single-task, intentional learners to a dual task and comparing them to participants who responded to the sequence under dual-task conditions throughout. Consistent with the first experiment, initial performance differences between the two groups disappeared when the single-task group was asked to divide attention. In a third experiment, participants transferred from initial learning under dual-task conditions to the single task of responding to the sequence. Curran and Keele hypothesized that if there were one type of learning, then participants would show greater disruption effects to random events upon removal of the dual task than before removal. However, if attentional and nonattentional modes of learning were truly distinct, then participants learning under dual-task conditions would continue using this mode after removal of the dual task. Consistent with the hypothesis for dual learning modes, disruption did not increase.

Overall, the findings obtained with the SRT task have been interesting, and their theoretical implications provocative, but one problem with the current state of this methodology (and a potential contributor to the inconsistencies in the literature) is that it provides rather limited information about learning and performance. Most studies have merely shown that attentional load affects the degree of disruption in RT on transfer to random sequences. Because the magnitude of differences may be affected by multiple factors, it is not particularly conclusive simply to show differences on one measure. In contrast, recent research using positron emission

tomography (e.g., Grafton et al., 1995) has been useful for assessing qualitative differences in modes of learning by identifying the brain areas involved in single- and dual-task learning. Contrary to the analyses offered by Curran and Keele (1993), who propose that attentional and nonattentional learning occur in parallel, and to Frensch et al. (in press), who propose that attentional and nonattentional learning are not different, distinct areas of the brain appear to be involved in different conditions of attention and in participants with different degrees of awareness of the sequence. Such findings suggest that qualitative differences may exist between attentional and nonattentional learning. However, there is still much to be learned about the proper interpretation of the neurological findings. It is not at all clear, for example, just what effects performing the secondary task would have on brain activity and how, in turn, that activity would interact with the areas of the brain involved in the SRT task alone. For the purpose of advancing theory, it would be useful to develop additional behavioral measures for assessing qualitative differences in learning. We propose that a probabilistic method for presenting the sequences can provide useful new information about the learning of sequential structure under different attentional conditions. Although the majority of studies have used deterministic sequences in the SRT paradigm, Cleeremans and McClelland (1991) and Stadler (1992) employed probabilistic sequences in their experiments. We propose to extend the use of probabilistic sequences to the study of variations in attentional load in the SRT task.

In addition to presenting events in fixed sequences, sequential structure can be imposed by manipulating the conditional probabilities of transitions. It has been known since the early work applying information theory to RT tasks (Hick, 1952; Hyman, 1953) that reactions are faster to more probable events and also event transitions with higher conditional probabilities. Viewed from this perspective, fixed sequences have very high ($p = 1.0$) conditional probabilities. There are some important differences between certainty and high probabilities, but there are reasons to prefer sequences with some uncertainty. For instance, learning with probabilistic sequences can be assessed by comparing performance on probable and improbable event transitions. In contrast, testing for learning with fixed sequences requires either: (a) comparing improvement on structured versus random sequences, or (b) assessing disruption in performance caused by switching to a random (or pseudo-random) sequence. Improvement in performance may be particularly insensitive because highly compatible choice-RT tasks often show small improvements with practice, even with completely random sequences (Fitts & Seeger, 1953). Furthermore, the comparison usually involves different groups, resulting in even less sensitivity than a within-subject comparison. The disruption observed on switching to a pseudo-random sequence is probably more sensitive than rate of improvement with practice, but switching sequences of-

fers limited opportunity to observe differences (namely, only after a series of structured blocks), and there are also questions about just what is tested when random sequences are used (cf. Reed & Johnson, 1994). In contrast, the difference in performance on probable and improbable transitions is continuously available and requires no change in the structure of the practiced sequence. Furthermore, the organization of probable and improbable transitions can be manipulated to ensure that they differ only in the structural component of interest (e.g., first- or second-order conditional constraints). A probabilistic sequence also has the advantage of adding a degree of disguise to the sequence. Less obvious sequences should help minimize explicit learning of the sequential structure.

There are some reasons to be concerned about introducing probabilistic sequences into the study of learning in SRT tasks. For example, responses must be based on the imperative stimuli and not simply on knowledge of the sequence, because improbable transitions do not follow the predominant sequence. This change could lead to differences in what is learned in the task. Also, the presence of improbable transitions may disrupt the associations formed between successive events, which may prevent learning or change the structure of learning. Thus, while we can study the issues of learning and attention with probabilistic sequences, the findings will not necessarily generalize to learning and performance with deterministic sequences. However, our goals are to investigate general issues of attention and learning, not simply to investigate the standard SRT paradigm.

Perhaps the greatest advantage of using probabilistic sequences lies in their potential for providing new information about some of the theoretical issues in the study of sequence learning. With fixed sequences, one can only determine whether learning of the sequence has occurred and how much RT is influenced by presenting stimuli in the sequence compared to random or control sequences. In addition to providing more continuous information about RT for probable versus improbable transitions, probabilistic sequences also offer the possibility of examining errors of various kinds. Anticipation errors, operationally defined as when a response appropriate for a probable transition occurs to an improbable transition, may be particularly informative about the nature of what is learned. More generally, errors may provide an additional basis for deciding whether attention produces qualitative differences in learning sequential structure. Willingham et al. (1989) grouped participants according to degree of awareness of the sequence and then examined frequency of anticipatory response times (defined as less than 100 ms). Anticipatory response times in their study were markedly more frequent for participants with full knowledge of the sequence (reaching almost 50% of responses after 400 trials), compared to participants with no knowledge or only partial knowledge of the sequence. The findings of Willingham et al. lead us to predict that more antic-

ipatory errors will occur under conditions of lower attentional load with probabilistic sequences or, conversely, that increasing attentional load should lead to a reduction in anticipation errors.

We designed the present experiments to examine the role of attentional load in SRT learning. As we will show, the use of probabilistic sequences offers useful new evidence about the processes involved in SRT learning and reveals qualitative differences in those processes under different conditions of load. The primary objective of Exps. 1 and 2 was to explore the use of this methodology. In the probabilistic paradigm we might expect to find a greater percentage of anticipation errors in single- than in dual-task conditions, because single-task practice offers more opportunity to attend to the sequence and to develop expectations about the sequence. Experiment 1 investigated the role of attention in learning first-order probabilistic sequences with an 80/20 ratio of probable to improbable transitions. Because a higher ratio was required to produce reliable learning with second-order sequences, Exp. 2 used a 90/10 ratio of probable to improbable transitions to investigate the role of attention in learning second-order probabilistic sequences. Experiment 3 addressed the issue of whether single-task and dual-task learning are qualitatively different and whether attentional load affects learning, the expression of what is learned, or both (cf. Frensch et al., in press). Participants in this experiment received a 90/10 ratio of probable to improbable transitions with second-order conditional sequences, but in addition, were transferred from single- to dual-task conditions or from dual- to single-task conditions. If attentional load merely disrupts the expression of learning, then both RT and errors should show similar patterns of transfer. However, if attentional load results in a different form of learning compared to learning without attentional load, then we would expect to see distinct patterns of transfer reflected in our measures of RT and errors.

Experiment 1

Method

Participants. Sixty undergraduates at New Mexico State University participated in partial fulfillment of requirements for an introductory psychology course. Participants were randomly assigned to one of two conditions: SRT task alone or SRT task plus tone-counting.

Apparatus. Stimuli were presented on PC computers equipped with standard keyboards. Stimuli were displayed in white on a black background. The stimulus was a cursor, 0.40×0.75 cm, centered over one of four 0.15×0.75 cm vertical bars spaced 1.5 cm apart in the lower left-hand corner of the video screen. The target locations are referred to as Locations 1 through 4. The participants made responses with the four fingers of their left hand using the “V”, “C”, “X”, and “Z” keys on the keyboard. The keys corresponded to Locations 1 through 4 from right to left. A sample of the stimulus display is shown in Fig. 1.

Materials. Two first-order conditional sequences were used in Exp. 1. Sequence A was 1-2-4-3 and Sequence B was 1-3-4-2. Each sequence had one occurrence of each of the four locations, and each location was completely determined by the immediately preceding location. One sequence occurred with a probability of 0.80 and the other occurred with a probability of 0.20, depending on which was designated as probable and which was designated improbable for each participant. The probabilistic sequences were implemented by using the most recent event to select the next event. With probability 0.80, the next event would be the event in the *probable* sequence following the just preceding event, and with probability 0.20, the next event would be the event in the *improbable* sequence following the just preceding event. Each block of trials was started by randomly selecting one of the four events to present as the first trial.

Procedure. Half of the participants were assigned to the single-task condition and half were assigned to the dual-task condition. The probable sequence was counterbalanced, so that for half of the participants Sequence A was the probable sequence and for the other half, Sequence B was the probable sequence. In the single-task, or no-tone, condition each trial began with the appearance of the cursor on one of the four vertical bars. The participant's task was to press the key corresponding to the location of the cursor. The cursor remained on the vertical line until the participant pressed the correct key, at which point the cursor disappeared. The onset of the next cursor occurred 417 ms after the correct response.

The dual-task, or tone, condition was identical to the no-tone condition, with the addition of a tone-counting task. A high- or low-pitched tone occurred 17 ms after the response. The pitch of the tone was 800 or 1200 Hz. For each block of trials, the probability of a high-pitched tone was randomly set in the range of 0.40 to 0.60. The participant's task was to count the number of high-pitched tones in the block of trials and report the number at the end of the block.

Participants completed 2 practice blocks containing 8 trials each. During practice, probable and improbable transitions occurred with equal likelihood. Thus, the next trial in sequence was equally likely to be determined by Sequence A as by Sequence B. Participants then completed 32 blocks in which transitions were generated from the probable sequence 80% of the time and from the improbable sequence on 20% of the trials. Participants completed 50 trials in each block.

Results and discussion

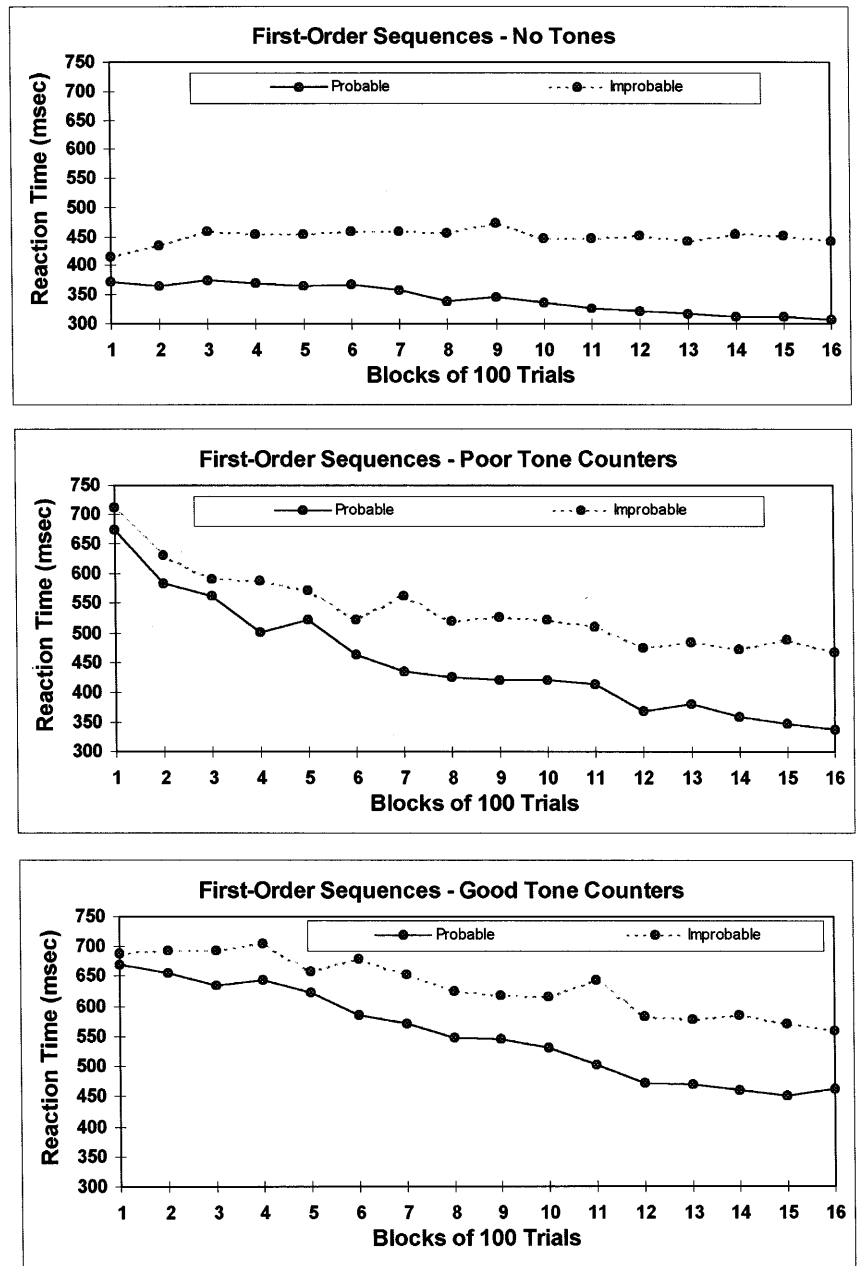
The dependent variables were RT and error rates in the key-pressing task. Mean RTs and error rates were calculated for each block of trials. The data were combined over adjacent blocks for analysis. Thus, instead of reporting results in terms of 32 blocks of 50 trials, the results are reported in terms of 16 blocks of 100 trials.

In previous studies, participants' data have often eliminated for poor tone-counting performance. Poor



Fig. 1 A sample stimulus display showing the cursor in the third position from the left, which calls for pressing the third key from the left

Fig. 2 Exp. 1: Correct reaction time for transitions in first-order conditional sequence structure (80/20 ratio of probable to improbable transitions)



tone-counters may be allocating less attention to the tone-counting task and consequently may spend more resources processing the sequential structure. It would be useful to know how tone-counting accuracy affected performance. Thus, tone-counting performance was treated as a factor in some analyses. Tone-counting scores were obtained by taking the absolute value of the difference between the tone count reported by participants and the actual number of high-pitched tones. Participants were ordered according to their tone-counting scores and were then divided into two groups (good and poor tone-counters), using a median split.

The counterbalancing of probable and improbable sequences was included to control for peculiarities of specific sequences and was of little interest experimen-

tally. However, in order to examine the possibility that different patterns of learning occurred for the two sequences, preliminary analyses were conducted using the specific probable sequence as a factor. Some minor differences emerged as a function of sequence, but the same pattern of results was obtained with both assignments of probable and improbable sequences. Thus, in the interest of focusing on the effects of primary importance, the data were combined across the two assignments.

In reporting the results, we present figures showing means relevant to the analyses and tables listing the effects tested in analyses of variance (ANOVAs). We only discuss the effects of particular interest with the aim of focusing the presentation.

Response-time analyses

Figure 2 shows mean RTs for correct responses to probable and improbable transitions for the no-tone group, the good tone-counters, and the poor tone-counters. It should be noted that with the exceptions of overall differences in speed among the single- and dual-task condition and greater overall improvement across blocks for dual-task compared to single-task learners, the patterns of learning (as indexed by differences in RT for probable and improbable transitions) look very similar among the three groups. Additionally, learning was immediate and rapid. For instance, participants in the single-task condition showed a significant difference in RT to probable and improbable transitions in the first 50 trials, 383 versus 421 ms $t(29) = 4.18, p < .001$.

In order to focus more clearly on the overall effects of practice, the Blocks factor in the reported ANOVAs represents a contrast between the first 200 trials and the last 200 trials. The ANOVA on RT is shown in Table 1, RT column. Of most importance is the Probability \times Blocks interaction, which demonstrated learning of the sequence in terms of greater probability effects later

in practice (130 ms) than earlier in practice (46 ms). The absence of a Probability \times Tasks interaction suggests that single- and dual-task practice leads to comparable learning, as indexed by RT differences to probable and improbable transitions.

Next, RTs were analyzed for good and poor tone-counters (see Table 2 RT column). The interaction of Blocks and Probability reflects learning of the sequence, because there was a smaller effect of probability in the first 200 trials (35 ms) than in the last 200 trials (123 ms). Although good tone-counters were slightly slower than poor tone-counters, there were no significant effects involving tone-counting performance, showing that RTs of good and poor tone-counters were equally influenced by transition probability.

Error analyses

Figure 3 shows mean error rates for probable and improbable transitions for the no-tone group, the good tone-counters, and the poor tone-counters. In contrast to the RT data, in which good and poor tone-counters

Table 1 Analyses of variance for Exp. 1 (first-order conditionals). *T* contrasts single-task with dual-task conditions. *B* contrasts the first 200 with the last 200 trials. *P* contrasts probable with improbable transitions. Values enclosed in parentheses are mean square errors for the preceding *F*s

Source	df	F		
		RT	All errors	Wrong sequence
Between subjects				
Tasks (T)	1	28.22***	0.04	0.60
S/T (error)	56	(71,088)	(117)	(63)
Within Subjects				
Blocks (B)	1	39.15***	27.04***	37.78***
B \times T	1	26.77***	1.33	2.84
BS/T (error)	56	(18,122)	(100)	(56)
Probability (P)	1	264.16***	84.24***	58.96***
P \times T	1	2.53	2.06	2.85
PS/T (error)	56	(1,748)	(50)	(42)
P \times B	1	82.33***	47.70***	45.18***
P \times B \times T	1	0.34	1.43	3.08
PBS/T (error)	56	(1,297)	(39)	(35)

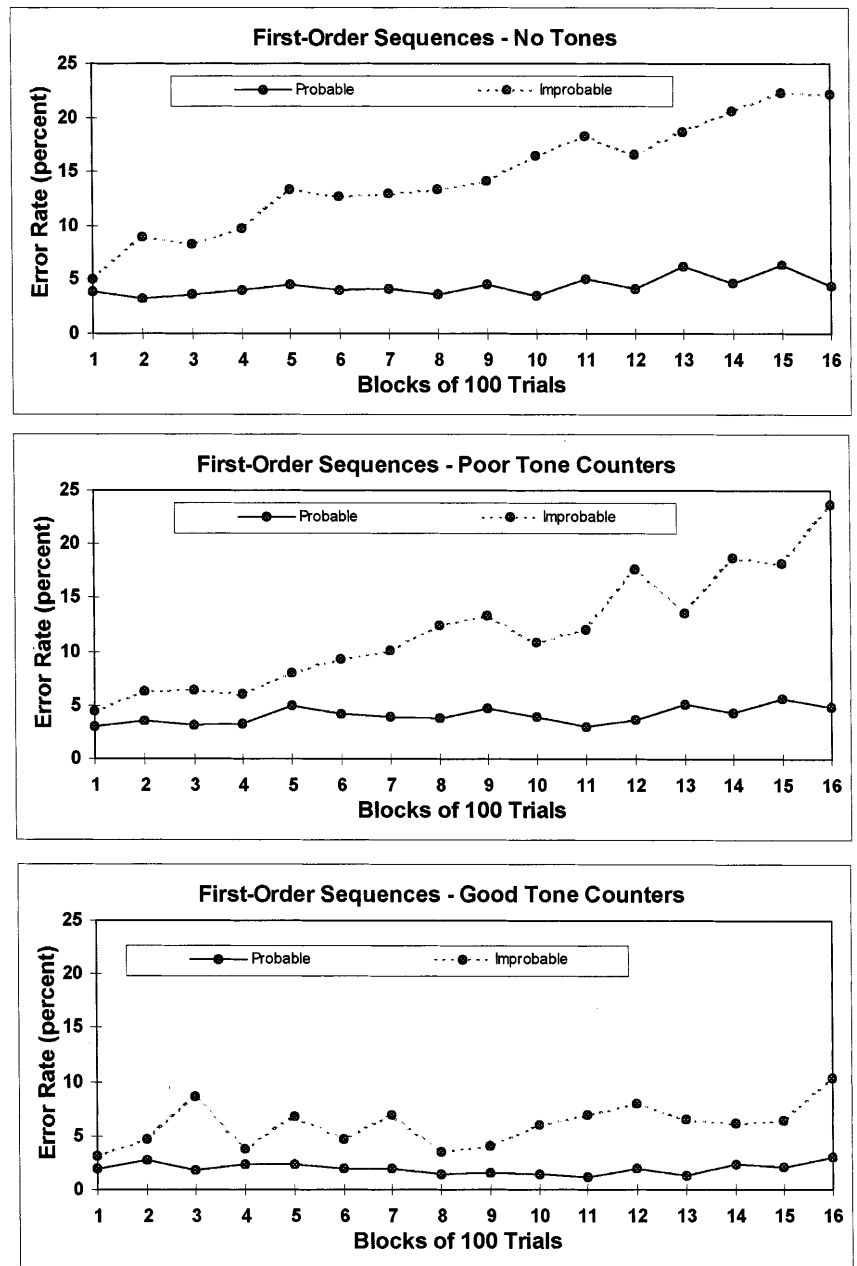
* $p < .05$, ** $p < .01$, *** $p < .001$

Table 2 Analyses of variance for good vs poor tone-counters in Exp 1. *C* contrasts good with poor tone-counters. *B* contrasts the first 200 with the last 200 trials. *P* contrasts probable with improbable transitions. Values enclosed in parentheses are mean square errors for the preceding *F*s

Source	df	F		
		RT	All errors	Wrong sequence
Between subjects				
Counting (C)	1	1.17	6.91*	3.08
S/C (error)	24	(98,703)	(97)	(48)
Within subjects				
Blocks (B)	1	35.61***	14.54***	16.78***
B \times C	1	1.21	4.68*	5.34*
BS/C (error)	24	(32,447)	(66)	(38)
Probability (P)	1	81.20***	28.49***	19.22***
P \times C	1	1.30	6.39*	3.10
PS/C (error)	24	(2,161)	(42)	(37)
P \times B	1	31.28***	15.98***	17.39***
P \times B \times C	1	0.19	4.02	4.20
PBS/C (error)	24	(1,730)	(41)	(31)

* $p < .05$, ** $p < .01$, *** $p < .001$

Fig. 3 Exp. 1: Error rate for transitions in first-order conditional sequence structure (80/20 ratio of probable to improbable transitions)



appeared to learn first-order conditional structure equally well, the error rates showed a different pattern of results. In particular, single-task participants and poor tone-counters showed dramatic increases in anticipation errors. Good tone-counters, on the other hand, showed far fewer anticipation errors. These preliminary observations were supported by statistical analyses of the data.

An ANOVA is shown in Table 1, All errors column. The Probability \times Blocks interaction reflected an increase in the effect of transition probability from 2.8% in the first 200 trials to 14.0% in the last 200 trials.

The observation that considerably more errors occur on improbable transitions strongly suggests that participants frequently make probable responses to improba-

ble transitions, and an additional error analysis confirms this inference. We conducted an analysis of the subset of errors we call “wrong-sequence” errors. Because there were two sequences from which transitions were selected, wrong-sequence errors consists of those responses corresponding to the transition that would have occurred if the other sequence had been selected for that trial. In the case of improbable transitions, the wrong-sequence error consists of the probable response being made to the improbable transition. For probable transitions, a wrong-sequence error consists of the improbable response being made to the probable stimulus. If participants learn to anticipate the sequence, then probable responses to improbable transitions should increase over blocks. An ANOVA for the wrong-

sequence errors (see Table 1, Wrong sequence column) revealed the same pattern of findings observed for all errors, confirming that wrong-sequence errors are a major source of the error effects. Overall, wrong-sequence errors accounted for 40.0% of the total errors on probable transitions and for 65.9% of the total errors on improbable transitions.

Next, error rates for good and poor tone-counters were compared (see Table 2, All errors column). The 3-way interaction shows that the increase in errors with training depends both on the transition probability and the tasks being performed. With probable transitions, the increases are only 1–2%. With improbable transitions, the increase is 3.8% for good tone-counters, but jumps to 18.8% for poor tone-counters. The ANOVA for the wrong-sequence errors (see Table 2, Wrong sequence column) revealed a similar pattern of results, particularly for the 3-way interaction. Poor tone-counters made 15.4% errors to improbable transitions compared to 6.3% for good tone-counters, demonstrating that poor tone-counters were indeed making frequent anticipation errors.

The purpose of Exp. 1 was to examine the use of probabilistic sequences in an SRT task using a sequence with first-order conditional structure. The patterns of learning reflected in the RT results were largely convergent, regardless of whether participants were in the single-task or dual-task condition and regardless of tone-counting accuracy. Examination of the RT results shows few differences in the magnitude of learning, even under conditions of attentional load. The error data, however, suggest greater differences in learning than those observed in the RT data. Namely, errors on improbable transitions were much more frequent when participants were not required to count tones and when they were not very accurate at counting. When participants performed well on tone-counting, they had lower error rates on the improbable transitions, but still showed sensitivity to the probabilities in their RT performance.

The use of probabilistic sequences has an important advantage over studies using fixed sequences. In fixed sequence studies, error rates are not informative because error rates on fixed sequences are very low and tend covary with RT. Errors are informative in the present paradigm, however, because they appear to be an index of how much participants are anticipating the sequence. Examination of the particular errors shows that errors on improbable transitions predominantly consist of the occurrence of the probable response to the improbable stimulus. Thus, the error rate on improbable transitions reflects the extent to which participants were anticipating the sequence locations determined by the more probable sequence.

A question resulting from Exp. 1 is whether the same pattern of results would emerge in cases of more complex learning, such as that involved in learning second-order conditional structure. Previous results are mixed with regard to this issue (e.g., Cohen et al., 1990; Nissen

& Bullemer, 1987; Reed & Johnson, 1994), so it would be informative to examine how manipulations of attention affect the learning of probabilistic sequences with only second-order constraints. This is the objective of Exp. 2.

Experiment 2

Method

Participants. Sixty-four undergraduates at New Mexico State University participated in partial fulfillment of requirements for an introductory psychology course. Participants were randomly assigned to one of two conditions.

Apparatus. The apparatus was identical to that used in Exp. 1.

Materials. Two second-order conditional sequences were taken from Reed and Johnson (1994) for Exp. 2. Sequence A, a 12-item sequence, was 1-2-1-3-4-2-3-1-4-3-2-4. Sequence B was 1-2-3-4-1-3-2-1-4-2-4-3. These sequences are equated with respect to frequency of location (each location occurred three times), number of reversals (one, e.g., 1-2-1, for each sequence), first-order transitions (each locations was preceded once by each of the other three locations), and repetitions (no repetitions in either sequence). The only difference between sequences was in second-order conditional structure. Instead of using an 80/20 probability ratio as in Exp. 1, one sequence occurred with a probability of 0.90 and the other occurred with a 0.10 probability.¹ The probabilistic sequences were implemented by using the two most recent events to select the next event. Thus with probability 0.90, the next event would be the event in the *probable* sequence following the two just preceding events, and with probability 0.10, the next event would be the event in the *improbable* sequence which followed the two preceding events. Each block of trials was started by randomly selecting two of the four events for the first two trials in the block.

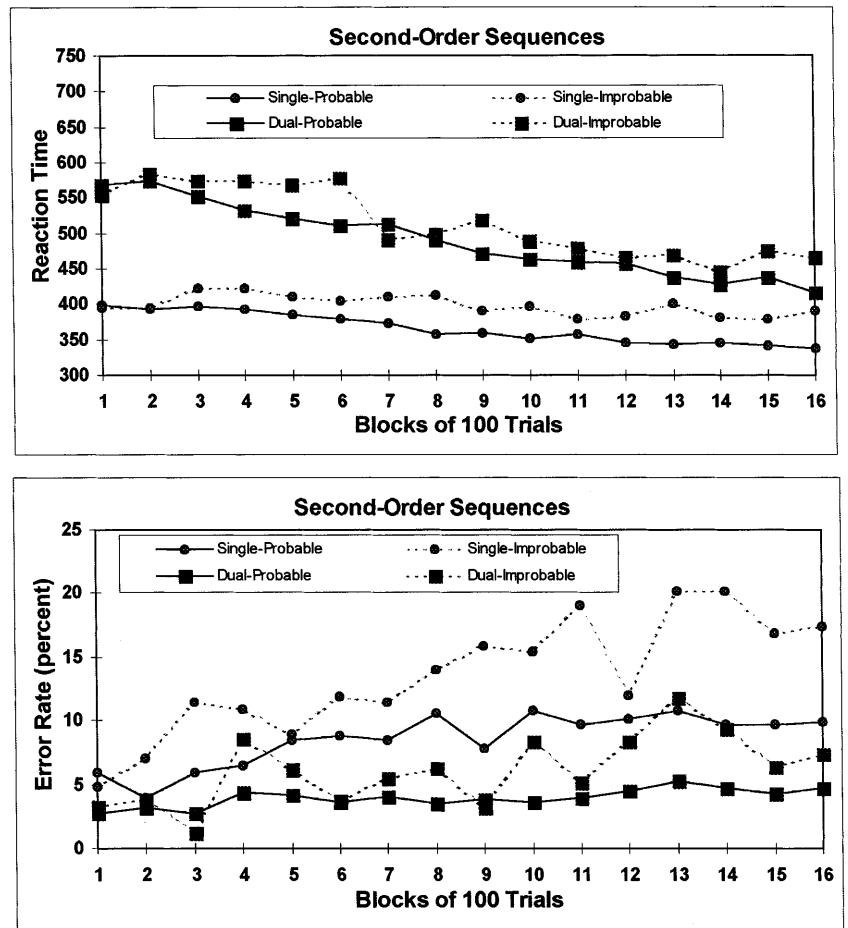
Procedure. As in Exp. 1, half of the participants were assigned to the single-task condition and half were assigned to the dual-task condition. Additionally, the probable sequence was counterbalanced, so that for half of the participants Sequence A was the probable sequence, and for the other half Sequence B was the probable sequence. As in Exp. 1, the 32 experimental blocks were preceded by 2 practice blocks (in which probable and improbable transitions occurred with equal likelihood). Participants completed 50 trials in each experimental block.

Results and discussion

Mean RT and error rates were calculated for each block of trials. The data were combined over adjacent blocks for the purpose of presentation. Thus, the results are analyzed using 16 blocks of 100 trials rather than

¹We first tried using an 80/20 probability ratio with the second-order sequences. The effects of transition probability were present but small in magnitude, and there were no differences in the probability effects as a function of attentional load. In particular, there was no sign of the high rate of anticipation errors observed in Exp. 1. Perhaps an 80/20 probability ratio with second-order conditional structure is not sufficiently salient. As shown in Exp. 2, a 90/10 probability ratio produces markedly different results.

Fig. 4 Exp. 2: Correct reaction time and error rate for transitions in second-order conditional sequence structure (90/10 ratio of probable to improbable transitions)



32 blocks of 50 trials. Good and poor tone-counters were separated in the same manner as in Exp. 1, but, in contrast with that experiment, there were only minor effects associated with tone-counting performance. Apparently, simply counting tones is sufficiently distracting with second-order sequences so that counting accuracy does not come into play. These analyses are not reported, in an effort to streamline the presentation. Preliminary analyses also showed similar patterns of effects for both sequence A and B as the probable sequence; therefore, data were combined across sequence type, eliminating this as a factor.

Response time analyses

Figure 4 shows mean RTs for correct responses to probable and improbable transitions for the single- and the dual-task groups. In contrast to Exp. 1, where single-task participants showed a significant difference in RT to probable and improbable transitions in the first 50 trials, RTs to probable and improbable transitions in Exp. 2 did not diverge markedly until much later in training. However, by the end of training, the magnitude of probability effects appeared similar in the two conditions. Statistical analyses confirmed these observations.

An ANOVA is shown in Table 3, RT column. The Blocks \times Tasks interaction reflected the fact that RTs in the dual-task condition improved much more over training (mean improvement = 122 ms) than RTs in the single task (mean improvement = 33 ms). The Probability \times Blocks interaction reflected the finding that there was no probability effect early in training, but there was a 45-ms effect later in training.

Error analyses

Figure 4 shows mean error rates for probable and improbable trials for the single-task group and for the dual-task group. In contrast to the results of the RT data, which suggested similar magnitudes of learning for both groups, the error data show differences between the groups. An increase in anticipation errors to improbable transitions is particularly evident for single-task learners.

An ANOVA (see Table 3, All errors column) revealed several significant effects. The Blocks \times Probability interaction is of particular interest. The effect of transition probability was greater later in training than earlier (4.9% later vs 0.8% earlier), showing learning of the sequential constraints with practice.

Table 3 Analyses of variance for Exp. 2 (second-order conditionals). *T* contrasts single-task with dual task conditions. *B* contrasts the first 200 with the last 200 trials. *P* contrasts probable with improbable transitions. Values enclosed in parentheses are mean square errors for the preceding *F*s

Source	<i>df</i>	<i>F</i>		
		RT	All errors	Wrong sequences
		Between subjects		
Tasks (T)	1	28.62***	11.21**	8.99**
S/T(error)	60	(38,518)	(140)	(70)
		Within subjects		
Blocks (B)	1	49.09***	46.23***	45.78***
B × T	1	16.18***	13.60**	16.78***
BS/T (error)	60	(7,798)	(37)	(29)
Probability (P)	1	19.51***	14.20***	18.17***
P × T	1	0.03	3.08	4.769*
PS/T (error)	60	(1,511)	(36)	(23)
P × B	1	32.724***	9.22**	16.28***
P × B × T	1	0.02	2.72	7.39
PBS/T (error)	60	(1,027)	(29)	(21)

p* < .05, *p* < .01, ****p* < .001

The wrong-sequence errors (see Table 3, Wrong sequence column) show particularly clear effects. The 3-way interaction of Probability × Blocks × Tasks reflects the finding that single-task learners made increasingly more errors to improbable than to probable transitions (mean difference between early and late trials = 11.2% for improbable transitions and 3.5% for probable transitions), whereas dual-task learners showed much smaller differences (mean difference between early and late trials = 2.6% for improbable transitions and 1.1% for probable transitions). As in Exp. 1, we see a significantly greater increase in anticipation errors in the single-task condition. Overall, wrong-sequence errors account for 53.6% of the errors on probable transitions and for 65.5% of the errors on improbable transitions.

According to the RT results in Exp. 2, dual-task learners are somewhat slower than single-task learners, but there were no interactions involving Probability and Tasks suggesting similar degrees of sequence learning in single- and dual-task conditions. Taken alone, the results of the RT analyses could be taken to suggest that there are no differences in sequence learning as a function of attentional load. However, the error data raise the possibility of a different interpretation. According to these data, single-task learners make substantially more anticipation errors than do the dual-task learners. Thus, group differences occurred in the error data despite the fact that learning was comparable as measured by RT.

The results of our first two experiments suggest differences in learning associated with attention. However, it is still not clear whether single- and dual-task practice lead to the same kind of learning with some associated performance differences, or whether there are multiple modes of SRT learning, (e.g., Curran & Keele, 1993; Mayr, 1996; Nissen & Bullemer, 1987). One method for investigating the similarity of learning under these different practice conditions is to transfer participants between single- and dual-task conditions to determine the extent to which learning under one learning condition is expressed under the other. If attentional load produces

distinct learning modes, then dual-task participants should be faster on transfer to the single task, but should not engage in the anticipatory behavior characteristic of single-task participants. However, if the dual task is merely limiting the learners' resources, then error rates on improbable transitions should increase with the removal of the dual task. Alternately, single-task participants should show some disruption on transfer to the dual task, but the disruption should only be minor and should be temporary, in the case of a unitary learning mode reflecting resource limitations. Major, long-term disruption in RT differences and error rate patterns on transfer to the dual task would be evidence for separate, competing learning modes. In short, if different learning occurs with and without a distracting task, then there should be differential patterns of transfer.

Experiment 3

Method

Participants. Sixty-eight undergraduates at New Mexico State University participated in partial fulfillment of requirements for an introductory psychology course.

Apparatus. The apparatus was identical to that used in Exps. 1 and 2.

Materials. The same two second-order conditional sequences and same probability ratio (90/10) used in Exp. 2 was used in Exp. 3.

Procedure. Half of the participants were assigned to the single-to-dual transfer condition, and half were assigned to the dual-to-single transfer condition. The probable sequence was also counterbalanced, so that for half of the participants Sequence A was the probable sequence, and for the other half, Sequence B was the probable sequence. Participants received 2 practice blocks (one for the dual-task condition and the other for the single-task condition), containing 10 trials each. Probable and improbable transitions occurred with equal likelihood during practice. The practice blocks

were followed by 24 learning blocks (single or dual), then 10 transfer blocks (dual or single respectively). Transitions in the learning and transfer blocks were generated from the primary sequence on 90% of the trials and from the secondary sequence on 10% of the trials. The first 2 trials in each block were chosen randomly, but the remaining trials were determined by the previous 2 trials. Participants completed 50 trials in each block.

Results and discussion

As in the previous experiments, the dependent variables were RTs and error rates in the key-pressing task. Mean RT and error rates were calculated for each block of trials. Data were also combined over adjacent blocks. Thus, the results are analyzed using 17 blocks of 100 trials rather than 34 blocks of 50 trials. Also, because the pattern of results was similar for Sequences A and B, data were combined across sequence type.

Response time analyses

Figure 5 shows mean correct RTs for probable and improbable transitions for single-to-dual and dual-to-single transfer participants. The figure shows all 24

blocks, but the primary analysis was performed on the last 5 blocks before transfer and the 5 blocks after transfer.

An ANOVA is shown in Table 4, RT column. The Order \times Time \times Blocks interaction revealed that single-to-dual task participants improved more after transfer to the dual task than when they were in the single task (mean improvement was approximately 1 ms before and 95 ms after transfer), whereas dual-to-single task participants improved more before transfer to the single task than after transfer (mean improvement was approximately 50 ms before and 5 ms after transfer). The Order \times Time \times Probability interaction (also shown in Table 5) reflects the fact that in the single-to-dual condition, there was a greater effect of transition probability before than after transfer (mean decrease = 44 ms), whereas in the dual-to-single condition the effect of transition probability was greater after than before transfer (mean increase = 25 ms).

Selected tests conducted on performance following transfer confirm the impression conveyed by the means. First, in order to verify that the probability effect found for the dual-to-single task participants was not merely the result of learning after transfer to the single task, an analysis of the effect of probability was conducted on the

Fig. 5 Exp. 3: Correct reaction time and error rate for transitions in second-order conditional sequence structure before and after transfer between dual-to single and single-to-dual task conditions (90/10 probability ratio)

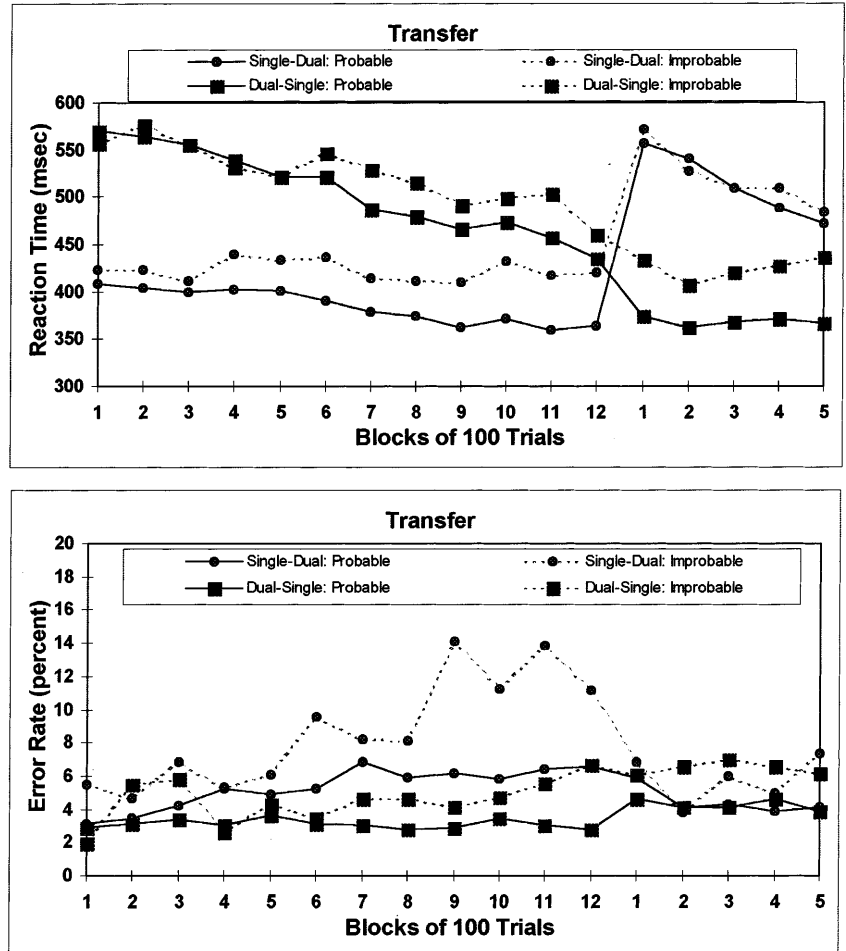


Table 4 Analyses of variance for 5 blocks before and after transfer in Exp. 3. *O* contrasts single-dual with dual-single. *T* contrasts before with after transfer. *B* represents blocks of 100 trials. *P* contrasts probable with improbable transitions. Values enclosed in parentheses are mean square errors for the preceding *F*s

Source	<i>df</i>	<i>F</i>		
		RT	All errors	Wrong sequence
Between subjects				
Order (O)	1	1.00	6.41*	2.47
S/O (error)	64	(101,306)	(284)	(164)
Within subjects				
Time (T)	1	7.10**	9.10**	5.83*
O × T	1	163.40***	37.79***	20.70***
TS/O (error)	64	(22,095)	(59)	(48)
Blocks (B)	4	10.17***	0.31	2.52*
O × B	4	1.85	0.08	0.79
BS/O (error)	256	(4,186)	(43)	(39)
T × B	4	2.19	2.09	1.57
O × T × B	4	10.42***	2.50*	0.53
BTS/O (error)	256	(4,306)	(40)	(34)
Probability (P)	1	95.48***	34.87***	30.05***
O × P	1	3.84	1.63	0.34
PS/O (error)	60	(4,714)	(77)	(92)
T × P	1	3.73	8.79**	6.88*
O × T × P	1	44.04***	9.97**	11.04***
PTS/O	64	(2,365)	(40)	(38)
B × P	4	1.97	0.86	2.53*
O × B × P	4	0.23	0.29	0.76
BPS/O (error)	256	(1,908)	(43)	(38)
T × B × P	4	0.72	0.76	0.80
O × T × B × P	4	1.05	1.67	0.49
TBPS/O (error)	256	(2,063)	(40)	(36)

p* < .05, *p* < .01, ****p* < .001

first 50 trials of single-task performance following transfer. Responses were significantly faster to probable (378 ms) than to improbable (425 ms) transitions, $F(1, 32) = 7.13$, $MSE = 5,266$, $p = .012$. Second, the dual-task performance for the single-to-dual task participants was examined separately in order to verify that no probability effects emerged after transfer. A 5 (blocks) × 2 (probable vs improbable) ANOVA for the dual-task performance of the single-to-dual task participants revealed an effect of Blocks, $F(4, 128) = 9.78$, $MSE = 7,643$, $p < 0.001$ (mean improvement = 85 ms), but no effects of Probability and no interaction involving Probability, even over the course of 500 trials. These analyses of transfer confirm that there is no significant transfer from single-to-dual task performance in the 500 trials following transfer, but the dual-to-single task transfer occurs immediately.

Error analysis

Figure 5 shows mean error rates for probable and improbable transitions for single-to-dual and dual-to-single transfer participants. An ANOVA (see Table 4, All errors column) showed several significant effects. The Order × Time × Blocks interaction reflects the finding that single-to-dual participants made increasingly more errors over blocks before transfer to the dual task than after transfer (mean increase in errors was approximately 5.8% in the single task, and there was virtually no change after transfer), whereas dual-to-single task

participants showed no changes in error rate over blocks. Importantly, the Order × Time × Probability interaction (see Table 5) showed that differences in errors to improbable and probable transitions for single-to-dual task participants were much greater before (5.5%) than after transfer (1.3%), whereas differences in errors to improbable and probable transitions for dual-to-single participants was virtually the same before and after transfer (2.1% and 2.3%, respectively).

As in the previous two experiments, we conducted an analysis on the wrong-sequence errors in order to confirm that participants were making probable responses to improbable transitions, hence engaging in anticipatory behavior. An ANOVA revealed a pattern of results similar to that found with all errors (see Table 4, Wrong sequence column). Most importantly, the Order × Time × Probability interaction mirrored that

Table 5 Improbable minus probable transition performance in Exp. 3

Order	Time	
	Before transfer	After transfer
Mean RT difference (ms)		
Single-to-dual	51	7
Dual-to-single	31	56
Mean error rate difference (%)		
Single-to-dual	5.5	1.3
Dual-to-single	2.1	2.3

found in the overall error analysis, demonstrating that wrong-sequence errors are a major source of the interaction found with all errors. Overall, wrong-sequence errors account for 51.1% of the errors on probable transitions and for 69.9% of the errors on improbable transitions.

Again, selected tests conducted on performance following transfer confirm the impression conveyed by the means. First, a 5 (blocks) \times 2 (probable vs improbable) ANOVA for the dual-task performance of the single-to-dual participants showed no significant effects in the 500 trials following transfer. This result and result of the RT analysis suggest that there is no reliable transfer from single to dual task conditions. Second, an analysis of the effect of probability was conducted on the first 50 trials of single-task performance following transfer from the dual to the single task. Even though the RT analyses showed that dual-to-single task participants responded faster to probable than improbable transitions in the first 50 trials immediately following transfer (47 ms), error rates to probable (4.8%) versus improbable (5.4%) transitions were not significantly different, $F(1, 32) < 1$.

We hypothesized that if the dual task were merely suppressing the expression of learning (Frensch et al., in press), then patterns of transfer to the single task should be similar for RT and errors. However, if different learning modes occur with and without a distracting task, then there should be differential patterns of transfer. Consistent with the latter hypothesis, dual-task participants became faster upon transfer to the single task, and the difference in RT to improbable and probable transitions increased. At the same time, there was no cost of increased error rates on improbable transitions, a cost which appears to be characteristic of the learning acquired under single-task practice. In contrast, participants in the single-to-dual task condition showed substantial disruption in RT, and the difference in performance on improbable and probable transitions disappeared. The fact that probability effects failed to appear in the 500 transfer trials suggests that participants were unable to apply the knowledge acquired in single-task practice when the secondary task was added.

General discussion

We developed a probabilistic version of sequence learning to provide additional information about the learning of sequential structure under varying conditions of attentional load. As we expected on the basis of earlier research, people learn about the probabilistic structure of sequences, as evidenced by faster RTs and lower error rates to probable as compared to improbable sequential transitions. Although probabilistic sequences with only first-order conditional constraints are learned more quickly and the effects of transition probability are larger, sequences with second-order constraints are also learned both with and without a secondary task. Our

findings are in agreement with those of Curran and Keele (1993), Reed and Johnson (1994), and Frensch et al. (in press), who found that the more complex sequences are learned even in the presence of a secondary task. The evidence now clearly favors this conclusion, in contrast to the suggestion from earlier studies (Nissen & Bullemer, 1987, and Cohen et al., 1990) in which there was no learning of second-order sequences under dual-task learning conditions.

With our probabilistic sequences, there are some notable differences in the learning that results from practice on the SRT task alone, compared with practice on the SRT task while a secondary tone-counting task is also being performed. Single-task practice leads to particularly high error rates on improbable transitions. Single-task learners frequently give the probable response to improbable transitions, indicating overt anticipations of the probable transitions. Probabilistic sequences are particularly informative because the RT data alone do not show substantial differences between single- and dual-task practice conditions. The difference in RT to probable and improbable transitions is larger under single-task conditions with second-order conditional sequences, but the difference is present under both dual- and single-task practice conditions. The error data give a clearer indication that there are differences in the learning that occurs under single- and dual-task conditions.

Perhaps the most persuasive evidence of different learning modes under single- and dual-task practice comes from Exp. 3, where the practice conditions were switched after extensive practice. The transfer data suggest that the smaller effect of transition probability on RT under dual-task conditions is due to limitations in performance instead of learning, because the effect of transition probability on RT immediately increases to near the level of single-task conditions when the secondary task is removed. This conclusion is consistent with that of Frensch et al. (in press), who presented evidence that dual-task conditions affect the expression of learning much more than the magnitude of the learning. They argue that the participants they analysed in both dual- and single-task learning conditions were implicit learners, because participants with explicit knowledge were discarded. Thus, their findings lead them to the conclusion that there is only one mode of *implicit* learning under dual- and single-task practice conditions. In their tests of explicit knowledge, Frensch et al. did find that training under single task conditions led to greater amounts of explicit knowledge.

In our studies, we did not test participants for explicit knowledge for reasons we discuss later, but with our methods we find that distinct modes of learning result from single- versus dual-task practice. There are two sources of evidence for this claim. First single- and dual-task practice lead to clearly different effects on error rates, particularly the frequency of anticipatory errors. Learning under dual-task conditions produces larger effects of transition probability on RT upon transfer to single-task conditions, but this effect on RT is not seen

in the error rates. In other words, the participants who learned under dual-task conditions do not show an increase in anticipatory errors when they are transferred to single-task conditions, despite the enhanced effect of transition probability on RT. Thus, even though single-task and dual-task learners show similar RT performance under single-task conditions, the RT similarity is countered by differences in the effect of transition probability on error rates. This pattern of findings shows a clear advantage for the use of probabilistic sequences, because the error data are critical to seeing differences between single- and dual-task practice.

The second source of evidence for distinct learning modes comes from the asymmetry in transfer. Training under single-task conditions leads to a form of learning that apparently cannot be applied under dual-task conditions. It appears that attentional resources are required to apply what is learned under single-task conditions, and our participants failed to show any effect of transition probability on either RT or error rates over the whole 500 trials of transfer to the dual-task condition. In fact, their data look very much like that of the participants trained under dual-task conditions at the very beginning of practice.

The result of Exp. 3 are inconsistent with the various proposals calling on interference as an explanation of the effects of a secondary task on learning sequential structure (Cleeremans & McClelland, 1991; Frensch et al., 1994; Stadler, 1995). Most directly relevant to interference hypotheses is our conclusion that dual-task practice does not, in fact, interfere with learning but instead produces performance limitations which may reduce the influence of the sequence on RT. Frensch et al. (in press) reach the same conclusion using deterministic sequences.

According to yet another view, Curran and Keele (1993) suggest that attentional and nonattentional forms of learning occur independently and in parallel. They argue that nonattentional learning should occur whether attention is devoted to the learning or not, because nonattentional learning is largely automatic in nature. Although our results support the claim of independence, they suggest that rather than operating in parallel, the two modes of learning compete with one another. In particular, there is no transfer of transition probability effects in the single-to dual transfer condition, implying that the learning observed with single-task training cannot be applied under attentional load. If a nonattentional form of learning occurred automatically during the single-task practice, this form of learning should have led to some sensitivity to the sequential structure when a secondary task was added, but our data show no reliable effects of transition probability during the 500 trials following the transfer. Further, the effects of event probabilities that are transferred in dual-to-single task transfer are demonstrably different than the effects observed under single-task learning. In particular, following dual-task learning, single-task performance does not exhibit the inflated error rate on improbable transitions

that regularly occurs following single-task learning. Thus, our findings suggest that quite distinct modes of learning occur under single- and dual-practice.

If some learning occurs automatically, as many have suggested, it may seem puzzling that the learning occurring under single-task conditions does not produce some "automatic" effects in addition to those observed with more "intentional" learning. Interestingly, the available data from neuro-imaging studies appear to show that quite distinct areas of the brain are most active under single- and dual-task conditions (Grafton et al., 1995). Such findings are consistent with distinct forms of learning, but the proposals are still quite speculative, as more work need to be conducted in order to interpret brain activity in situations as complex as those involving dual-task practice.

One interpretation of our results is that single-task learning encourages intentional control of responding. Intentional control requires attention for establishing the specific intentions and for applying them in performance. The benefit of intentional control is realized in the ability to respond quickly to anticipated events. The cost of such anticipation is the high error rate on unexpected (improbable) events, but then, these events rarely occur. In contrast, dual-task learning promotes the development of perceptual-motor skill, which reflects the structure of the task and situation, including the structure of the sequence. Such skill is developed through practice and does not require full attentional resources. This perceptual-motor skill comes to incorporate the structure of the sequence, but it is more stimulus driven, leading to less overt anticipation than does intentional control.

Although it is understandable that sufficient attentional load might block intentional control, it is less obvious why, in the absence of attentional load, intentional control and the perceptual-motor skill do not both develop. In our transfer study, there was no sensitivity to the sequential structure on transferring from single- to dual-task conditions. Why should perceptual-motor control fail to develop at the same time as intentional control? From one perspective, the same SRT task is being performed, so the opportunity for developing the same perceptual-motor skill would seem to be present. A possible explanation is that intentional control of behavior bypasses the source of control in a perceptual-motor skill. According to this view, perceptual-motor control would not develop when responses are under intentional control because perceptual-motor learning occurs in the control of responses. Presumably, continuing to practice would eventually lead to development of perceptual-motor skill, but perhaps only after intentional control drops out. This is clearly a speculative hypothesis at this point, and further research is needed to test and extend the proposal. We might also look to some of the ongoing work on the neural basis of sequence learning in cognitive neuroscience (Curran, 1997; Grafton et al., 1995; Keele & Curran, in press) to help evaluate this hypothesis. Comparing active brain areas

for participants trained under single-versus dual-task conditions and tested under single-task conditions would be particularly informative using probabilistic sequences because the testing conditions are identical, yet our findings show clear differences in error rates on improbable transitions.

At this stage of the research, there are certainly many other possible explanations for the pattern of data we observed. Two particular alternatives should be considered. First, it is possible that single- and dual-task learning lead to different criteria for responding in the task. Single-task learning leads to faster responses at the cost of more errors. This alternative could explain why more errors overall and more errors on improbable events occur under single-task training. However, the criterion explanation alone does not handle the transfer data particularly well, because dual-task training leads to responses just as fast as single-task training without increased error rates when the tone-counting task is removed. It also fails to explain why there is no transfer of sequence probability effects from single-task training to dual-task performance. A second hypothesis might help with the transfer findings. Suppose that dual-task performance interferes with the expression of sequence knowledge because learning to cope with two tasks takes time and interference results until then. If this hypothesis is true, our data suggest that it takes a lot of time (over 500 trials) to adapt to the dual task, because no evidence of sensitivity to the sequence is revealed over that many trials after transferring to the dual-task conditions following single-task learning. In one sense, our hypothesis about intentional control versus perceptual-motor control asserts that there is interference in applying the knowledge acquired under single-task learning to dual-task performance. We speculate that the kind of knowledge acquired under single-task training cannot be applied under dual-task conditions. Still, additional research will be required to examine these (and other) alternatives as our research progresses.

It is important to emphasize that we suspect a rather complex relationship among attentional resources, task complexity, and modes of learning. For attentional resources to matter, the task to be learned must be at an appropriate level of complexity. If the sequence is too easy (Exp. 1), limited distraction does not block allocation of sufficient attention to prevent the development of intentional control (as with poor tone-counters who had a high rate of anticipatory errors). If the sequence is too difficult (as when we tested second-order conditional sequences with an 80/20 probability ratio and found limited learning and few anticipation errors), the full resources of attention may not be sufficient to develop intentional control of behavior. The complexity of probabilistic sequences is also likely to be important in determining the nature of learning (as in the sequences generated from artificial grammars as used by Cleeremans & McClelland, 1991). Only with an appropriate level of task and sequence difficulty does the relationship

between availability of attentional resources and modes of learning become clear.

Our data also show clearly that the amount of practice required to learn about sequential structure varies both with the complexity of the sequence and the attentional load. For the simpler, first-order sequences used in Exp. 1, sensitivity to transition probability emerged within the first 100–200 trials, regardless of attentional load. With the more complex, second-order sequences, it was necessary to increase the probability ratio to 90/10 to obtain marked effects of transition probability even with no secondary task, and with the 90/10 ratio it may take more than 500 trials of practice before the probability effect is clearly established. It is difficult to generalize these estimates to learning fixed sequences, but perhaps some of the inconsistencies among previous studies are due to differing amounts of practice. For example, Nissen and Bullemer (1987) trained people for 400 trials and found no learning of the more complex sequences with dual-task practice.

One final issue concerns the extent to which the high rates of anticipatory errors observed in our studies reflect explicit learning of the sequence. Anticipatory behavior in the present experiments may reflect awareness, but we have no evidence which directly addresses this question. We chose not to use direct tests of explicit learning for several reasons. First, there are serious difficulties associated with devising measures which meet both information (a test of the appropriate knowledge) and sensitivity (a test with sensitivity comparable to the indirect tests) criteria (Shanks & St. John, 1994). Second, it is even more difficult to characterize the appropriate knowledge with probabilistic sequences because there is much greater variety in the material seen by participants. Third, concluding that some participants have no explicit knowledge involves accepting the null hypothesis, which is a risky experimental practice, and it could easily result in erroneously identifying explicit learning as implicit. Fourth, using different tasks to presumably reflect implicit as opposed to explicit learning is suspect, because performance in most tasks is probably influenced by both implicit and explicit processes. For example, Jacoby (1991) has proposed methods for dissociating the contributions of implicit and explicit processes in memory tasks. In addition Buchner, Steffens, Erdfelder, & Rothkegel (1997) have also proposed a model to assess the contributions of such processes in sequence learning tasks. Unfortunately, the process dissociation approach also has its limitations (cf. Curran & Hintzman, 1995, 1997; Dodson & Johnson 1996). With all the difficulties in assessing explicit knowledge, using tests of explicit knowledge runs the risk of encouraging flawed inferences. In our research program, we prefer to use experimental methods to manipulate the degree to which participants can be conscious of particular aspects of a task and to look for qualitative changes in performance that may reflect differences in awareness (cf. Cheesman & Merikle, 1985; Marcel, 1983). In the present studies, we manipulated the attentional resources available for

learning about sequential structure and found interesting qualitative differences in performance. In future studies we plan to address how manipulations of explicit training affect learning and performance in the probabilistic version of the SRT task to extend the manipulation of attentional load in the studies reported here.

Acknowledgements This research was supported by Grant No. F41624-96-1-0003 from the Armstrong Laboratory of the United States Air Force and by the Computing Research Laboratory at New Mexico State University.

We thank Axel Buchner, Asher Cohen, Tim Curran, Peter Frensch, Peder Johnson, Steven Keele, James Neely, and Michael Stadler for helpful comments and suggestions on various drafts of this article.

References

- Buchner, A., Steffens, M. C., Erdfelder, E., & Rothkegel, R. (1997). A multinomial model to assess fluency and recollection in a sequence learning task. *The Quarterly Journal of Experimental Psychology*, *50*, 631–663.
- Cheesman, J., & Merikle, P. M. (1985). Word recognition and consciousness. In D. Besner, T. Waller, & G. Mackinnon (Eds.), *Reading research: Advances in theory and practice* (Vol.5). New York: Academic Press.
- Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, *120*, 235–253.
- Cohen, A., Ivry, R. I., & Keele, S. W. (1990). Attention and structure in sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 17–30.
- Curran, T. (1997). Implicit sequence learning from a cognitive neuroscience perspective: What, how, and where? In M. Stadler & P. Frensch (Eds.), *Handbook of implicit learning*. Beverly Hills, CA: Sage.
- Curran, T., & Keele, S. W. (1993). Attentional and nonattentional forms of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 189–202.
- Curran, T., & Hintzman, D. L. (1995). Violations of the independence assumption in process dissociation. *Journal of Experimental Psychology: Learning Memory, and Cognition* *21*, 531–547.
- Curran, T., & Hintzman, D. L. (1997). Consequences and causes of correlations in process dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 496–504.
- Dodson, C. S., & Johnson, M. K. (1996). Some problems with the process-dissociation approach to memory. *Journal of Experimental Psychology: General*, *125*, 189–202.
- Fitts, P. M., & Seeger, C.M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, *46*, 199–210.
- Frensch, P. A., Buchner, A., & Lin, J. (1994). Implicit learning of unique and ambiguous serial transitions in the presence and absence of a distractor task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 567–584.
- Frensch, P. A., Lin, J., & Buchner, A. (in press). Learning versus behavioral expression of the learned: The effects of a secondary tone-counting task on implicit learning in the serial reaction task. *Psychological Research*.
- Grafton, S. T., Hazeltine, E., & Ivry, R. (1995). Functional mapping of sequence learning in normal humans. *Journal of Cognitive Neuroscience*, *7*, 497–510.
- Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*, 11–26.
- Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *45*, 188–196.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional use of memory. *Journal of Memory and Language*, *30*, 513–541.
- Keele, S. W., & Curran, T. (in press). On the modularity of sequence learning systems in humans. In E. Covey (Ed.), *Neural representation of temporal patterns*. New York: Plenum.
- Lewicki, P., Hill, T., & Bizot, E. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. *Cognitive Psychology*, *20*, 24–37.
- Marcel, A. J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology*, *15*, 197–237.
- Mayr, U. (1996). Spatial attention and implicit sequence learning: Evidence for independent learning of spatial and nonspatial sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 350–364.
- Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*, 1–32.
- Perruchet, P., & Amorim, M. 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.
- Perruchet, P., Gallego, J., & Savy, I. (1990). A critical reappraisal of the evidence for unconscious abstraction of deterministic rules in complex experimental situations. *Cognitive Psychology*, *22*, 493–516.
- Reber, P. J., & Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learning & Memory*, *1*, 217–229.
- 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.
- Schvaneveldt, R. W., & Gomez, R. L. (1996, November). *Attention and modes of learning: Evidence from probabilistic sequences*. Paper presented at the 37th annual meeting of the Psychonomic Society, Chicago, IL.
- Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*, 367–447.
- Stadler, M. A. (1989). On learning complex procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *10*, 581–598.
- Stadler M. A. (1992). Statistical structure and implicit serial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 318–327.
- Stadler M. A. (1993). Implicit serial learning: Questions inspired by Hebb (1961). *Memory & Cognition*, *21*, 819–827.
- Stadler M. A. (1995). Role of attention in implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 674–685.
- Willingham, D. B., Nissen, M. J., & Bullemer, P. (1989). On the development of procedural knowledge. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *15*, 1047–1060.