Attentional and Nonattentional Forms of Sequence Learning

Tim Curran and Steven W. Keele

This research investigated the hypothesis that sequential patterns of behavior can be learned by 2 independent mechanisms. One requires attention to the relation between successive events, whereas the other operates independently of such attention. In 4 experiments, subjects learned visuospatial sequences in a serial reaction time task. The relation between attentional and nonattentional learning was explored by assessing the extent to which learning transferred between conditions with or without distraction. The results suggest that attentional and nonattentional learning operate independently, in parallel, do not share information, and represent sequential information in qualitatively different ways.

A fundamental type of learning in which humans excel is the learning of sequential patterns of behavior. In four experiments, we investigated the hypothesis that humans exhibit two forms of sequential learning. One form of learning requires attention to the relation between successive events in the sequence, not only for acquisition but also for the expression of the learning in performance. We hypothesized that the other type of sequential learning did not require attention to these relations. Furthermore, these two forms of learning are independent of one another, with no communication or sharing of information between them. If subjects perform a series of behavioral acts that occur in a predictable order and under conditions relatively free of distraction, we suppose that attentionally based and nonattentionally based learning of the sequence occur in parallel. If distraction is added during learning, the attentional form is disabled, but the nonattentional one is unmodified. That is, attention is neither necessary nor helpful to the nonattentional form of learning.

This hypothesis is similar to ideas examined by other investigators but is also different in several ways. Nissen and her colleagues (Nissen & Bullemer, 1987; Willingham, Nissen, & Bullemer, 1989; see also Lewicki, Hill, & Bizot, 1988; Stadler, 1989) have suggested that learning of sequences can be either procedural (without awareness) or declarative (with awareness). Both of these, according to Nissen, require attention. We distinguish both of those forms from a third, nonattentional type of learning.

Nissen and Bullemer (1987) used a paradigm, largely copied in our studies, in which visual signals occurred at one of four locations on a screen. Depending on location, subjects pressed one of four keys. Unknown to the subjects, the signals occurred in a particular repeating pattern. With practice, reaction time (RT) to the signals improved. Some of the improvement could be attributed to sequence learning because the RTs following practice were much faster when the signals occurred in the pattern than when they occurred at random.

Nissen and Bullemer (1987) found that adding a secondary task to the primary RT task eliminated sequence learning. The secondary task required subjects to count tones of one pitch and to ignore those of another pitch that sounded within each primary task response–stimulus interval. We presumed that such a task degraded attention to the relation between successive sequence events. When attention between these events was free, subjects were able to learn the sequential structure. Even with attention, however, subjects often were unaware of their learning. Moreover, Nissen and Bullemer found that patients with Korsakoff’s syndrome could learn the sequential structure, as manifested in improved RTs, but that the patients did not report awareness of the sequence. Similarly, scopolamine, an anticholinergic drug, depressed awareness of a sequence but had no effect on sequential learning, as measured by a performance criterion (Nissen, Knopman, & Schacter, 1987). Normal subjects sometimes became explicitly aware of the sequential structure and were able to verbally report it or predict the next item (Nissen & Bullemer, 1987; Willingham et al., 1989). Such explicit knowledge led to much faster, often anticipatory RTs than did learning unaccompanied by expressed awareness. Both the conscious and unconscious forms of learning were blocked by distraction (Nissen & Bullemer, 1987). Therefore, both forms were said to require attention.

In the Nissen paradigm, the visual signals occurred in four different spatial positions. Each position occurred two or three times in a repeating pattern that contained 10 elements. Cohen, Ivry, and Keele (1990) found that if a repeating pattern had at least one unique stimulus position, subjects were able to learn it in the presence of the interpolated tone-counting task. Procedural learning was reflected in improved RTs, even though subjects rarely became aware of...
the pattern. By contrast, if each stimulus element was repeated in different orders in different parts of the pattern (as in Nissen’s experiments), then the distraction task virtually eliminated learning. Cohen et al. suggested that unique items within a sequence allow learning by a simple associative mechanism. However, when items are repeated in different orders in different parts of a sequence, the ambiguity of associations forces subjects to parse the sequence and learn orders within parts. Such parsing is the essence of hierarchic representation. Similarly, Wickelgren (1979) proposed that the associative interference in memory can be overcome by representing information in hierarchic chunks. This idea was elaborated by Keele and Jennings (1992) in explorations of a connectionist simulation of sequential learning developed by Jordan (in press). The interaction between attention and sequence structure in learning (Cohen et al., 1990) suggests that a mechanism that demands attention to the relation between successive events may be specialized for hierarchic representation and that a mechanism that does not require attention to these relations may learn by item–item associations. Therefore, Nissen’s distinction between procedural and declarative learning is not the same as the attentional and nonattentional distinction explored here. Her distinction between procedural and declarative learning was based on the subject’s awareness, and both forms of learning required attention because of the nature of the sequences she used.

The primary focus of our research was on the role of attention in sequence learning, but it remains difficult to untangle attention and awareness. The work of Nissen and colleagues illustrates that under some circumstances, learning requires attention but may or may not involve awareness (Nissen & Bullemer, 1987; Nissen et al., 1987; see also Cohen et al., 1990). Thus, we thought it necessary to attempt to differentiate between the effects of awareness and attention. We examined levels of awareness in order to test the hypothesis that sequence learning would vary with awareness level in an attentional mode but that the effects of awareness would not transfer to a nonattentional mode. When subjects report lack of awareness, concerns are often expressed that more sensitive measures reveal at least partial awareness (e.g., Perruchet & Amorim, 1992). This concern was not that important in our research because we found that differences in expressed awareness affect learning in one situation but not in another. Such an interaction makes moot whether some awareness may remain in a case in which awareness variation has no effect on learning.

When we refer to one form of learning as nonattentional, we do not wish to imply that no attention whatsoever is used on the primary task. Undoubtedly, subjects must in some sense attend to a visual stimulus to make a response. The secondary task used is one in which to-be-counted tones are inserted between a response to one visual signal and the appearance of the next visual signal. It is likely that the effect of the secondary task is to degrade attention to the relation between successive events, but whether attention is completely blocked is not crucial. The main issue is whether variation in attentional availability qualitatively alters the pattern of results. Because of these considerations, we use the terms attentional and nonattentional for convenience to discriminate between two forms of learning that are differentially affected by distraction.

In each of the experiments, we used paradigms developed by Nissen and colleagues, and extended by Cohen et al. (1990), and Keele, Jennings, Jones, and Cohen (1992). Subjects respond with keypresses that correspond to the positions of visual signals. On some occasions the signals occur in repeating sequential patterns that the subjects can learn. Learning is measured by the RT difference between responses to stimuli appearing in sequential versus random orders. Distraction is manipulated by the presence or absence of a tone-counting task that can be superimposed on the primary RT task.

**Experiment 1**

In the first experiment, we examined how variations in the amount of learning acquired under attentional distraction-free conditions would transfer to a situation in which distraction was added. We hypothesized that during initial distraction-free learning, nonattentional learning would occur in parallel with attentional learning but would be uninfluenced by the quality of attentional learning. When distraction is added at a later time, the expression of attentional learning should be largely suppressed, and performance in that period would reflect only the expression of nonattentional learning. Performance in that case should be the same regardless of the original level of attentional learning.

Attentional learning was varied by making the learning task intentional for some subjects and incidental for others. Subjects in the intentional group were told explicitly that the visual signals to which they were responding would occur in a particular order, and the order of the visual signals was then described. We expected that such instructions would immediately lead to short RTs because subjects would be able to explicitly predict the succession of signals (Willingham et al., 1989). Subjects in the incidental group were told nothing about the sequence. With practice under distraction-free conditions, some subjects eventually became aware of the sequence in that they could describe it, but others reported less awareness. We called these two groups, identified on a post hoc basis, “more aware” and “less aware” subjects, respectively. These groups of subjects should differ in the degree to which, when attention is freely available, they use sequential structure in performance. The question was whether differences in explicit knowledge would be communicated to the nonattentional mechanism, as assessed by imposing a distraction task after initial learning.

**Method**

**Subjects.** Fifty-seven subjects participated in the experiment. Subjects were undergraduates and members of the University of Oregon community and were paid $5 each for participating.

**Apparatus.** The stimuli were presented on a Video 100 monitor controlled by an Apple IIe microcomputer.

**Task and design.** The primary task consisted of an X-mark appearing in one of four horizontal positions. Four horizontal line
segments demarcated the four spatial positions, and the X-mark appeared directly above one of the line segments on each trial. Subjects placed the fingers of their right hand on four microswitch keys to the right of the monitor. Responses were made by pressing the key that corresponded to the spatial position of the X-mark. Subjects were encouraged to respond quickly and accurately on each trial. The X-mark remained on the screen until subjects made a response. The next stimulus appeared 200 ms after each response. A total of 120 trials occurred within each block of the experiment. Subjects were allowed to rest as long as desired between blocks.

Each block of trials in this and the subsequent experiments could be one of four types defined by the pattern of X-marks (random or sequence) crossed with the presence or absence of a distraction task. In random blocks, the location of the X-marks was determined randomly for each trial with the constraint that consecutive stimuli never appeared in the same location. In sequence blocks, the X-marks followed a structured pattern of six elements that repeated throughout the block. Equal numbers of subjects were assigned to six different sequences in order to control for any idiosyncratic effect arising from a particular ordering. Each of the sequences had the same basic structure. Two positions occurred only once in the repeating cycle of six elements and the other two were repeated. A repeated item was followed by a different item at its two places of occurrence within a cycle. Designating the four spatial positions as 1, 2, 3, and 4 from left to right, the six sequence versions were as follows: 1-2-3-2-4-3, 1-2-3-1-3-4, 1-4-3-1-3-2, 1-4-2-3-1-2, 1-3-2-4-1-2, and 4-2-3-2-1-3. Each block of trials consisted of 20 repetitions of one of these sequences, for a total of 120 trials. The first stimulus in each block could start anywhere in the sequence (determined randomly), and the beginning and end of the sequence cycles were not marked in any way.

In the dual-task blocks, a tone sounded within each 200-ms response–stimulus interval of the primary task. The tone sounded either 40, 80, or 120 ms (varied randomly) after each primary task response. Each tone was one of two different pitches. Subjects were instructed to keep a silent, running count of the number of high-pitched tones within each block while ignoring the low-pitched tones. After each block, the subjects reported the number of high-pitched tones, and then the correct number was displayed. The number of high-pitched tones varied randomly with the constraint that 60–90 (50%–75%) high-pitched tones must occur within each block. The subject was encouraged to consider the tasks as equally important and not to concentrate too much on one task at the expense of the other.

The four block conditions are abbreviated as follows: single-task, random (sing/rand); single-task, sequence (sing/seq); dual-task, random (dual/rand); and dual-task, sequence (dual/seq). Thus, a series of two dual-task, random blocks followed by five single-task, sequence blocks, then one single-task, random block can be abbreviated as follows: 2(dual/rand)–5(sing/seq)–1(sing/rand).

Procedure. The experiment (12 total blocks) was divided into three separate phases: dual-task practice (2[dual/rand]); single-task learning (4[sing/seq]–1[sing/rand]–1[sing/seq]); and dual-task transfer (2[dual/rand]–1[dual/seq]–1[dual/rand]). Subjects were given the instructions for each phase immediately prior to the first block of the phase.

After identical practice tasks, one third of the subjects were assigned to the intentional learning condition and the other two thirds to the incidental learning condition. Before beginning the first sequence learning block, subjects in the intentional learning group were given a diagram that outlined the pattern that the X-marks would follow. These subjects were told that the X-marks would often follow this pattern in the next phase of the experiment and were given 1 min to study the pattern before beginning. They were not explicitly warned about the random block. Those in the incidental learning group were told nothing about the existence of the repeating pattern.

All subjects were treated identically in the dual-task transfer phase. For each subject, the pattern of X-marks in the sequence blocks was identical to the pattern in the learning phase. The subjects were given no indication that there would or would not be a repeating pattern in this phase.

Between the single-task learning and dual-task transfer phases, subjects in the intentional and incidental learning groups were given a questionnaire to assess their ability to report any explicit knowledge of the sequence. The questionnaire asked subjects if they thought the X-marks ever appeared in any kind of a pattern or if they seemed to appear at random locations. Subjects who thought there was a pattern were asked to describe it.

Results and Discussion

Two criteria (determined preexperimentally) were required in order for each subject’s data to be included in the final analyses. First, the mean RT in the initial two practice blocks had to be under 800 ms. Previous results from our laboratory have shown that removal of extremely slow subjects’ data helped to reduce random error. Because all subjects were treated identically in the practice phase, the procedure did not differentially bias conditions. Second, subjects were eliminated for poor performance on the tone-counting task because they might have paid more attention than desired to the RT task. If a subject’s final tone count was off by more than 10 on two or more of the nonpractice, dual-task blocks, his or her data were eliminated.

A total of 13 (7 in the intentional group and 6 in the incidental group) subjects were eliminated on the basis of these criteria. Seven subjects in the intentional group were eliminated, 4 for poor tone counting and 3 for slow RTs. Six subjects in the incidental group were eliminated, 3 for poor tone counting and 3 for slow RTs. Each dropped subject was replaced. The analyses were based on data from 14 subjects in the intentional learning group and 30 in the incidental learning group.

In all of the experiments reported here, errors were analyzed for the RT task. Overall, they averaged less than 5% in each condition of each experiment. In no case did the pattern of errors, whether significant or not, counter the results of the RT analyses. Therefore, we do not report error analyses for any of the experiments. Incorrect responses were omitted from the RT analyses.

Questionnaire. All subjects in the intentional learning group described perfectly the sequence on the questionnaire. Those in the incidental learning group were divided into two subgroups, one designated as being more aware and the other less aware of the sequence. In the following results, those who could write down at least four of the six sequence positions in the correct order were placed in the more aware group (n = 19), and those who knew three or fewer were placed in the less aware group (n = 11). Of the 11 subjects in the less aware group, 7 indicated that they thought the location of the X-marks was determined ran-
random. Changing the exact criterion for this distinction affected the magnitude of the effects but did not qualitatively alter the results.

**Learning phase.** For each subject, median RTs were calculated for each block of trials and the means of medians were then calculated over subjects. Figure 1 shows RTs as a function of blocks of practice. The amount of sequence learning was indexed by the difference in RT on the random block (Block 7) and the average of the preceding and following sequence blocks (Blocks 6 and 8). RT in the random block served as a control for nonspecific practice and fatigue effects. Taking the average of Blocks 6 and 8, rather than either alone, made it more comparable to the random block. Thus, the RT difference between the sequence and random blocks provided an index of RT improvement attributable to sequence learning.

The pattern of results clearly demonstrates that subjects who expressed less awareness—while showing sequential learning under single-task conditions—showed less learning than did subjects who expressed more awareness and less than subjects who were explicitly told the nature of the sequence. Amount of learning was analyzed in a Group (intentional, more aware, less aware) × Block (sequence [mean of Blocks 6 and 8 medians], random [Block 7 median]) analysis of variance (ANOVA). The Group × Block interaction was significant, \( F(2, 41) = 5.50, \) \( MSe = 2,536, \) \( p < .01. \) The mean learning scores for each group, obtained by subtracting the mean of the two sequence blocks from the random block, were as follows: intentional = 210 ms, more aware = 189 ms, and less aware = 118 ms. Post hoc (Newman-Keuls) tests revealed that all of the groups significantly differed in the amount of single-task sequence learning.

**Dual-task phase.** Sequence learning was indexed by the RT difference between the sequence block (Block 11) and the average of the immediately preceding and following random blocks (Blocks 10 and 12). An ANOVA revealed a significant difference between random and sequence, \( F(1, 41) = 51.97, \) \( MSe = 815, \) \( p < .01, \) indicating sequential learning; however, random versus sequence did not interact significantly with group (less aware, more aware, and intentional), \( F(2, 41) = 0.408, \) \( MSe = 815, \) \( p = .67. \) Thus, statistically speaking, all three groups expressed the same sequential knowledge during dual-task conditions, despite the fact that under single-task conditions, the intentional and more aware groups showed much more sequential knowledge than did the less aware group. Groups that differed when attention was freely available no longer differed when distraction was added.

All three groups showed more sequential knowledge under single-task conditions than under dual-task conditions. A post hoc test for the smallest of these differences showed that even for the less aware group, the sequence effect under single-task conditions (random — sequence = 118 ms) was twice that under dual-task conditions (53 ms), \( F(1, 10) = 9.84, \) \( MSe = 2,405, \) \( p < .05. \) Note that this effect varied with the choice of the awareness criterion. When only the 7 subjects who expressed no explicit knowledge (answered "random" on the questionnaire) were analyzed, the magnitude of both the single-task (81 ms) and dual-task (42 ms) sequence learning effects was reduced. Although the single-task learning effect was still nearly twice that of the dual-task effect, the difference between these effects was not significant for this small subset, \( F(1, 6) = 3.067, \) \( MSe = 1,107, \) \( p = .13. \)

The primary conclusion is that variations in single-task learning, caused by awareness differences, were not transferred to dual-task conditions. This was consistent with the hypothesis that attentional and nonattentional learning occur in parallel with no conveyance of attentionally based learning to dual-task conditions. Such a result suggests that awareness affects sequence learning only when attention is fully available. When attention was divided, additional knowledge gained by the aware and intentional groups was not conveyed to the nonattentional mechanism. Similarly, changing the awareness criterion for the less aware groups had a larger effect on single-task learning (118 ms vs. 81 ms) than on the dual-task performance (53 ms vs. 42 ms). Again, this suggests that differences in awareness primarily affect attentional learning.

Interestingly, subjects who were less aware of the sequence still showed a greater sequence effect under conditions of no distraction than when distraction was added. This was consistent with the idea that even for the less aware subjects, some attentionally based learning was blocked during distraction. This conclusion is dampened by the fact that learning under those conditions was statistically equated when a stricter criterion was adopted for less aware subjects. However, even with the stricter criterion, the single-task RT effect was twice that in the dual-task phase. It is likely that the statistical comparison suf-
ferred from a lack of power because of the small number of subjects \((n = 7)\) who met this criterion.

We assumed that any RT improvement exhibited when a sequence was present during the dual-task phase was attributable to memory for the sequence that transferred from the single-task learning phase. However, it is possible that all of the sequential knowledge exhibited in the sequential dual-task block might have actually been learned during that block. We assessed this possibility by examining the RT improvement at multiple points within the transfer blocks. Responses for the last sequence block (Block 11) were divided into four quarters—Q1 (Trials 1–30), Q2 (Trials 31–60), Q3 (Trials 61–90), and Q4 (Trials 91–120)—and a median RT was computed for each quarter for each subject. If the RT improvement in the transfer phase reflects new learning within this phase, then RTs should improve from Q1 to Q4. However, if RT improvement reflects memory from the previous phase, then RTs should be similar in all quarters. This analysis was complicated because RTs tend to increase across a block of trials, perhaps because of fatigue. We reasoned that the normal slowing due to fatigue should be the same in the sequence block (Block 11) as in the surrounding random blocks (Blocks 10 and 12). Therefore, we also computed median RTs for each quarter of Blocks 10 and 12. The Block 10 and Block 12 medians were averaged for each quarter to serve as a control against which the median RTs in the sequence block (Block 11) could be compared.

The difference between each subject’s sequence and random RTs for each quarter block was entered into a Group (intentional, more aware, less aware) × Quarter \((Q1, Q2, Q3, Q4)\) ANOVA. This analysis revealed a significant main effect for quarter, \(F(3, 123) = 4.89, M_{S_{e}} = 2.665, p < .01,\) and a significant Group × Quarter interaction, \(F(6, 123) = 2.19, M_{S_{e}} = 2.665, p < .05.\) The difference scores for each quarter of each group were as follows: intentional, \(Q1 = 24\) ms, \(Q2 = 43\) ms, \(Q3 = 38\) ms, and \(Q4 = 52\) ms; more aware, \(Q1 = 49\) ms, \(Q2 = 38\) ms, \(Q3 = 48\) ms, and \(Q4 = 65\) ms; and less aware, \(Q1 = –6\) ms, \(Q2 = 75\) ms, \(Q3 = 70\) ms, and \(Q4 = 73\) ms. Inspection of these scores suggests that the bulk of the within-blocks learning effect was restricted to the less aware group. This impression was supported by analyzing the groups separately. When the effect of learning across the quarters was analyzed for the intentional and more aware groups only, there was no improvement across the quarters, \(F(3, 93) = 1.13, M_{S_{e}} = 2.684, p = .34.\) By contrast, when the less aware group was analyzed alone, there was a significant improvement across the quarters, \(F(3, 30) = 6.56, M_{S_{e}} = 2.608, p < .01.\) In all, the only evidence for learning within the transfer phase could be isolated to the 81-ms improvement of the less aware group from Q1 to Q2.

There are three possible explanations for these results. First, the less aware subjects could have learned the sequence after fewer than 10 cycles in the transfer phase. This possibility was ruled out in a similar experiment by Keele et al. (1992). There, it was shown that if subjects never experienced events in sequence for 10 prior blocks of trials, no learning was exhibited in a single block of dual-task trials with sequence (with 20 cycles of the sequence). Second, the less aware subjects could have been slower to begin to use their memory of the sequence. Third, the results for the less aware subjects might have been spurious. On the basis of the results of Keele et al. (1992) and those of the next two experiments (which show no evidence of single-block learning), we favored the latter two explanations.

Experiment 2

In Experiment 1, initial learning of the sequence occurred under distraction-free conditions in all cases. The second experiment contained two groups that allowed a stronger test of the hypothesis of independent learning mechanisms. One group, like the intentional group in Experiment 1, was given an initial, explicit description of the sequence in which signals would occur. That group then practiced for eight blocks of trials under single-task conditions before being transferred to dual-task conditions. The other group was not told about the presence of a pattern and practiced under dual-task conditions throughout. Previous results by Cohen et al. (1990) using almost identical procedures showed that a dual-task group was able to learn the sequence by a performance criterion but showed little or no awareness of the sequence they had learned. In Experiment 2, after both groups completed an equivalent amount of practice with the sequence, learning was assessed under distraction. We predicted that once dual-task conditions were imposed, both groups would show an equivalent degree of sequence learning.

Method

Subjects. Forty-seven subjects participated in the experiment. The subjects were undergraduates and members of the university community who were paid $5 each for participating.

Apparatus and task. Same as in Experiment 1.

Design and procedure. There were three different phases in the experiment: practice \(2\) \([dual/rand]\); learning \(8\) \([dual/seq]\) or \(8\) \([sing/seq]\); and dual-task performance \(2\) \([dual/rand]-1\) \([dual/seq]-1\) \([dual/rand]\).

Subjects were randomly assigned to two different learning groups. The single-task learning group was given the same instructions as the intentional group in Experiment 1. They were given time to study the repeating pattern before completing eight single-task blocks with the sequence. The dual-task learning group was not informed that the X-marks would follow a repeating pattern. These subjects completed eight dual-task blocks with the sequence.

Before the dual-task performance phase, all subjects were given the questionnaire described in Experiment 1 to assess their reportable, explicit knowledge of the sequence. All subjects were treated identically in the dual-task performance phase.

Results and Discussion

Seventeen subjects were excluded from the analyses because of slow RTs or poor tone counting (as in Experiment 1). Of the 12 subjects removed from dual-task learning
group. 7 failed the tone-counting criterion, 2 were too slow, and 3 failed both criteria. Of the 5 subjects removed from the single-task learning group, 4 failed the tone-counting criterion and 1 was too slow. These subjects were replaced according to group membership until there were 15 subjects in each experimental group.

All of the subjects in the single-task learning group described perfectly the sequence on the questionnaire following the learning phase. Only 2 of the 15 subjects in the dual-task learning group indicated awareness of a repeating pattern. These subjects correctly wrote down only two and three of the six sequence positions, respectively. Thus, the dual-task learning group was predominantly unaware of the presence of a repeating pattern.

The RT results are shown in Figure 2. Again, the first two blocks of trials for both groups involved random events under dual-task conditions. The next eight blocks involved the repeating sequence. The group given intentional learning instructions, and for whom the secondary task was removed, immediately showed a large RT improvement compared with the incidental learning group that retained the dual task. The last four blocks all involved the dual-task condition (2[dual/rand]–[1[dual/seq]–[1[dual/rand]]). Learning was measured as the difference between the sequence block (Block 13) and the average of the immediately preceding and following random blocks (Blocks 12 and 14). An ANOVA on these RTs showed that the single-task learning group responded somewhat more slowly than did the dual-task learning group overall, \( F(1, 28) = 3.16, \ MS_e = 12,809, \ p = .09 \). This would be expected because they had had much less experience performing under dual-task conditions. The measure of sequential learning, however, was based on the difference in RTs between random and sequential conditions. There was a significant main effect for block, \( F(1, 28) = 36.16, \ MS_e = 1,307, \ p < .001 \), indicating that across groups, the subjects learned the sequence as reflected in faster RTs on sequence than the average of the random blocks. Most important, there was no significant interaction between block and group, suggesting equivalent sequential knowledge for the two groups, \( F(1, 28) = 0.48, \ MS_e = 1,307, \ p = .50 \).

The transition from Block 10 (sequence) to Block 11 (random) might also be taken as a reflection of the amount of sequence learning by the dual-task learning group. With this in mind, it is reasonable to ask why the RT difference between these blocks (32 ms) was less than that for Blocks 12–14 (63 ms). Subjects were interrupted between Blocks 10 and 11 for the questionnaire and further instructions. RTs tend to improve after such an interruption (cf. the “reminiscence effect”; Woodworth, 1938), which will reduce the random minus sequence difference. The RTs for Blocks 12–14 were not affected by such an interruption, so this difference yielded a purer measure of sequence learning.

The possibility that the RT effects in the transfer phase were caused by new learning rather than memory from the learning phase was assessed as in Experiment 1 by dividing the transfer blocks into quarters. A median RT was computed for each quarter of each subject’s sequence block (Block 13) trials. These were subtracted from the mean of the medians calculated from each quarter of the random blocks (Blocks 12 and 14) to obtain an estimate of RT improvement attributable to sequential knowledge in each quarter of the sequence block. These difference scores were entered into a Group (single-task learning vs. dual-task learning) \( \times \) Quarter (Q1, Q2, Q3, Q4) ANOVA, which showed no significant effects. Collapsed across both groups, the difference scores for each quarter were Q1 = 48 ms, Q2 = 55 ms, Q3 = 68 ms, and Q4 = 53 ms. Because these RT differences did not change across the transfer block quarters, it is reasonable to infer that the RT advantage is indicative of memory for previous learning rather than of new learning within the transfer phase.

It is instructive to compare Figures 1 and 2. Even though the intentional, single-task learning group of Experiment 2 received considerably more training than did the comparable group of Experiment 1, the sequential knowledge exhibited when the secondary task was introduced was not much different. These results provide strong support for the hypothesis that additional knowledge gained under single-task intentional conditions is not expressed in the nonattentional conditions.

Up to this point we have assumed that RT decrements reflect knowledge of the underlying structure of the spatial sequences. However, it is possible that the decrease reflects the learning of something other than the formal structure of the sequence. For example, Lewicki et al. (1988) observed improvement in a more complex sequence learning experiment, but Perruchet, Gallego, and Savy (1990) argued that the improvement may reflect sensitivity to the relative frequencies of certain target location sequences rather than the acquisition of a complex system of rules. Sensitivity to the relative frequency of certain target positions could also conceivably account for our results. For example, in a sequence such as 1-2-3-2-4-3, the relative frequencies of signals in

---

**Figure 2.** The mean of subjects’ median reaction times plotted for each block of trials in Experiment 2. (Blocks are labeled random [R] or sequence [S]. Separate plots are shown for the subjects who learned with or without distraction. The critical blocks for learning assessment are enclosed in the dashed box.)
Positions 1 and 4 of the visual display are lower than those of Positions 2 and 3. In the random blocks, on the other hand, all positions have the same relative frequency. Could this explain the RT improvements shown in Experiment 2? If so, we would expect that subjects’ RTs would only be faster than random for certain positions in the repeating sequence.

To explore the possibility that something other than the intended sequence was learned by subjects, we analyzed the data from the performance phase of the dual-task learning group of Experiment 2 according to position in the sequence. This group was selected because these subjects showed the least awareness, so they were less likely to learn the complete sequential structure. Three types of sequence positions were defined (using 1-2-3-2-4-3, for example): unique (e.g., 1 and 4); after unique (e.g., 2 following 1, and 3 following 4); and ambiguous (e.g., 3 following 2, and 2 following 3). If subjects have truly learned the structure of the patterns, all three position types should show faster RTs than random events. Mean RT in the three random dual-task blocks was 493 ms. In the sequence block, the mean RTs for each position type were as follows: ambiguous = 457 ms, unique = 434 ms, and after unique = 421 ms. Student Newman-Keuls tests confirmed that RTs were significantly faster for all three types of sequence positions compared with random. If RT improvement was attributable only to sensitivity to the relative frequency of occurrence, only the ambiguous or after ambiguous positions would show improvement. The fact that all sequence positions elicited faster RTs than random events supported our contention that the entire structure of the sequence was learned.

Experiment 3

The first two experiments demonstrated that differing levels of awareness (Experiment 1) and differing availability of attention (Experiment 2) affected initial sequence learning but that these differences were not expressed when subjects were required to transfer this knowledge to a situation requiring dual-task performance. We argued that this asymmetry was due to the parallel acquisition of two forms of sequential knowledge—attentional and nonattentional—under single-task conditions. When the secondary task was added, only the nonattentional form was expressed. Another possibility, however, is that the amount of sequential knowledge measured by the difference between random and sequence events was sensitive to the overall level of RTs. RTs were faster under single-task conditions than under dual-task conditions, and it might have been the change in scale that was responsible for the larger difference in the single-than the dual-task conditions. We examined this problem in Experiment 3.

Here, subjects have initial practice under incidental, dual-task conditions and then transfer to single-task conditions. Presumably, only one form of learning initially occurs: the nonattentional form. When the distraction is subsequently removed, although RTs should speed up, performance should be based only on the nonattentional form of learning, at least until there is sufficient practice to allow attentional learning to occur. Thus, these conditions predict symmetry in which the sequential effect, as assessed by the difference between random and sequential conditions, should be the same under dual- and single-task conditions.

Method

To test this prediction, we needed only a single group. This group of 19 subjects, recruited as before, learned the sequence under dual-task conditions in a manner similar to the dual-task group of Experiment 2. However, there were two changes: following Block 10, which involved dual-task sequence, a block of dual-task random trials was used, followed by another dual-task sequence block. This allowed an assessment of amount of sequence learning under dual-task conditions. Then, in the transfer phase, the secondary task was removed and a test of sequential knowledge was given under single-task conditions. Thus, the conditions of the blocks were as follows: practice (2[dual/rand]); dual-task learning (8[dual/seq]-l[dual/rand]-l[dual/seq]); and single-task transfer (2[sing/rand]-l[sing/seq]-l[sing/rand]).

Results and Discussion

A total of 9 of the 19 subjects were removed from the analyses. Four subjects were removed because of poor tone counting during the nonpractice blocks, 3 because of slow practice RTs, and 2 for failing both criteria.

The RT results are shown in Figure 3. Blocks 1–12 all involved dual-task conditions. At first, it may be surprising that there was no immediate improvement in RT on Block 13 when the distraction task was first removed. Recall, though, that the last block under dual-task conditions involved a repeating sequence, whereas the first two blocks under single-task conditions involved random events. The improvement caused by removing the secondary task was
cancelled by changing from sequence to random. The main point is that the amount of sequence learning exhibited under single-task conditions was no larger than that under dual-task conditions. Four within-subjects measures were compared: dual/seq (mean of Blocks 10 and 12 medians); dual/rand (Block 11 median); sing/seq (Block 15 median); and sing/rand (mean of Blocks 14 and 16 medians). A Block (random vs. sequence) × Task (dual vs. single) ANOVA showed no significant interaction, $F(1, 9) = 2.51, M_{Sc} = 1.040, p = .15$. The results are consistent with the view that sequential learning occurred under dual-task conditions and that when the distraction was subsequently removed, the only learning that could be exhibited was the nonattentional form.

The question of whether the RT improvement in the transfer phase was attributable to memory from the initial dual-task learning or new learning within the single-task phase was examined as in the previous experiments. A median RT was computed for each quarter of each subject's sequence block (Block 15) trials, and the corresponding RTs from the random blocks (Blocks 14 and 16) served as controls. The RT difference scores were entered into a single factor (Q1, Q2, Q3, Q4) repeated measures ANOVA; no significant differences emerged between the sequence learning scores for each quarter, $F(3, 27) = 1.42, M_{Sc} = 921, p = .26$. The RT difference scores for each quarter were Q1 = 62 ms, Q2 = 67 ms, Q3 = 80 ms, and Q4 = 67 ms. From this, we can infer that any improvement on the sequence block of the single-task phase was transferred from previous learning rather than constituting new learning within that block.

**Experiment 4**

Earlier results by Cohen et al. (1990) suggested that attention to the relation between successive events is necessary for learning some types of sequences but unnecessary for others. Consider a sequence of six events composed of three stimulus positions. If those positions are labeled 1, 2, and 3 from left to right, one such sequence is 1-3-2-3-1-2. Cohen et al. found no evidence that a repeating sequence of this sort (called an ambiguous sequence) was learned in the presence of the tone-counting distraction task. Cohen et al. proposed that the ambiguous sequential associations between adjacent events in this type of sequence prevented learning unless some mechanism kept track of which part of the sequence was being dealt with at the moment. Event 1 is followed by Event 2 at one place and by Event 3 at another place. Likewise, Events 2 and 3 are followed by different events depending on the part of the sequence. This ambiguity can be solved by place keeping (i.e., by a hierarchic code that specifies part and then order within part). Similarly, it has been demonstrated that attention is necessary for learning the positional dependence of grammatical bigrams in artificial grammar (Dienes, Broadbent, & Berry, 1991). Earlier work by Keele and Summers (1976) has also suggested that keeping track of position within a sequence requires attention. Thus, attention seems to play a special role in the acquisition of position-dependent information.
which each signal position occurred three times. The sequence could begin at any point and it recycled once the end was reached. This sequence was longer than ones we have previously used and more comparable to ones used by Nis- 

sen and Bullemer (1987), who also found that distraction blocked learning. If attention enables a place-keeping mechanism, then increasing the sequence length should increase the difficulty of place keeping and should make learning especially dependent on attention. Therefore, we supposed that this more difficult sequence could not be learned under distraction. Also, in the dual-task conditions, subjects were required to count the tones aloud. We hoped this would be more distracting than the silent counting used in the previous experiments.

To summarize, in this experiment subjects first engaged in single-task learning with instructions to intentionally learn the sequence. The sequences were more difficult than in the previous studies. Following single-task learning, the secondary task was added and a test made for a sequence effect on RT. Subjects were divided according to awareness of the sequence during the final test phase. This division turned out to be informative in interpreting the outcome.

Results and Discussion

The data from 4 subjects were removed from the analyses because they did not meet the tone-counting criterion. Because of our inexperience with the new version of the distraction task, we did not establish an RT criterion.

Subjects were divided, on a post hoc basis, into more aware and less aware groups according to their questionnaire responses. We were concerned only with awareness during the dual-task performance phase (all subjects reported substantial awareness during the single-task phase). Eleven of the 25 subjects indicated that the X-marks appeared at randomly determined locations during the dual-task phase. These subjects composed the less aware group. The other 14 subjects indicated that they noticed a pattern during this final phase. These subjects composed the more aware group despite the fact that 2 of them were unable to recall any of the pattern. Thus, the criterion used for calling subjects less aware in this experiment yielded a relatively conservative estimate.

Figure 4 shows the RT results for more aware and less aware subjects separately. As before, the first two blocks of trials were dual-task with random-event orders. The next eight blocks were without the distraction task. Block 11 involved random events, and the contrast of that block with Blocks 10 and 12 allowed an assessment of sequence learning. In the learning phase, there was no difference between subjects who later expressed less awareness during the following dual-task phase and those who expressed more awareness. The mean of median RTs in the single-task random block (Block 11) was compared with the means in the surrounding sequence blocks (Block 10 and 12). An ANOVA showed a significant difference between random

Method

Subjects. Twenty-nine subjects participated in the experiment. The subjects were undergraduates and members of the university community who were paid $5 each for participating.

Apparatus. This was the same as in previous experiments.

Stimuli. Equal numbers of subjects learned one of six different nine-element sequences. The structure of these sequences were identical (A-C-B-C-A-B-A-C), but the spatial positions were varied: 1-3-2-3-1-2-1-3-2, 2-1-3-3-2-3-2-1-3, 3-2-1-3-1-3-2-1-3, 1-2-3-2-1-3-1-3-2, 2-3-1-3-2-1-3-1-3, and 3-1-2-3-2-3-2-1. The total number of stimuli presented in each block was 108. Thus, in sequence blocks, the pattern was repeated 12 times.

Design and procedure. There were three different phases in the experiment: practice (2[dual/rand]); single-task learning (8[sing/seq][1][sing/rand][1][sing/seq]); and dual-task performance (2[dual/rand][1][dual/seq][1][dual/rand]).

At the beginning of the single-task learning phase, subjects were informed that the X-marks would often follow a repeating pattern that they should try to learn in order to enhance the speed and accuracy of responding. They were not told the exact sequence. Next, subjects completed the 10 single-task blocks.

In the practice and dual-task performance phases, subjects were instructed to count high-pitched tones aloud. They were led to believe that their running counts were being tape-recorded in order to encourage compliance with the instructions. In fact, only the final count was actually recorded as in the previous experiments. The number of high-pitched tones was determined randomly with the constraint that 50–80 (46%–74%) occurred in any block.

At the end of the experiment, subjects were given a questionnaire to assess their reportable, explicit knowledge of the sequence. The questionnaire was identical to that described in the first experiment, except that it required a separate response for each phase of the experiment. Thus, subjects reported whether the events in each phase were thought to be random or in sequence and then described the sequence when appropriate.
and sequence, indicating sequence learning, \(F(1, 23) = 50.07, MS_e = 2,865.64, p < .01\), but no interaction with group, \(F(1, 23) = .001\).

On Blocks 13–16, the secondary task was reintroduced. Learning was measured as the RT difference between Block 15 and the average of Blocks 14 and 16. The subjects who expressed less awareness of a sequence during this phase exhibited no sequence effect, although those who expressed more awareness appeared to show a small effect. An ANOVA yielded neither a main effect for block, \(F(1, 23) = 3.83, MS_e = 987, p = .06\), nor a Block \(\times\) Awareness interaction, \(F(1, 23) = 2.49, MS_e = 987, p = .13\). However, inspection of Figure 4 suggests that the aware group might have benefitted from prior exposure to the sequence, whereas the less aware group did not. Post hoc ANOVAs were performed separately on each group to support these impressions. These tests showed a significant effect for the more aware group, \(F(1, 13) = 4.88, MS_e = 1,436, p < .05\), but not for the less aware group, \(F(1, 10) = 0.16, MS_e = 404\).

The results of this experiment are consistent with the view that, when sufficiently long sequences containing multiple, ambiguous associations have been learned, attention is necessary for the expression of this knowledge. Thus, when distraction is added during performance of an already learned sequence, information learned with attention may be inaccessible and sequence knowledge is no longer evident. The mere presence of distraction may not be sufficient to completely retard the use of attention, however. Some subjects report awareness of a sequence, even under dual-task conditions, and they still may use attention for keeping track of location in a sequence task even in the face of distraction. Barring awareness, no sequence effect remains. The reliability of the dual-task sequence learning effect for the more aware group is suspect because of the likelihood that the post hoc ANOVA capitalized on chance. However, because this effect, if real, runs counter to our predictions, we now explore its origins.

One concern with these results centers around the definition of awareness. As many investigators have noted, the concept of awareness is highly subjective and problematic. How is it, for example, that one can be sure that none of those who reported lack of awareness were aware of the sequence at some earlier critical time? How is it that one can be sure that those who reported awareness really were aware? Perruchet and Amorim (1992) have recently argued that a more rigorous assessment of explicit knowledge challenges previous claims of independent procedural and declarative learning systems. Given the particular pattern of results in Experiment 4, these problems are of less concern. Separating subjects with varying ability to explicitly report the sequences allows assessment of the relative contribution of awareness without the need to assume a complete lack of awareness. Although some subjects reporting less awareness might have been partially aware, the major point is that they showed no residual sequence learning once the distraction task was added. This result was qualitatively different for subjects who reported more awareness. Clearly, some factor distinguishes the performance of the two groups. It appears that one group used some process during the dual-task phase that was not used by the other group. (See Schacter, 1987, for a similar discussion of how qualitative performance differences between subjects doing identical tasks can help resolve issues of awareness.)

We have suggested that when a secondary task is added, it retards the use of attentionally based learning. If subjects are aware of a sequence, however, they may be able to strategically use attention. One implication of this view is that subjects who are more aware of the sequence should do less well on the secondary task than subjects who are less aware and thus do not attempt to use attention in preparing for subsequent events. To test this hypothesis, we examined performance on the secondary task. For each dual-task block, a subject was given a tone-counting error score by calculating the absolute value of the difference between the subject’s reported count and the actual number of high-pitched tones. Inspection of the counting errors suggests that the less aware group (mean error = 1.89) devoted more attention to the tone task than did the more aware group (mean error = 3.45). However, a Group (more aware vs. less aware) \(\times\) Block Type (random vs. sequence) ANOVA revealed no significant effects. Unfortunately, the secondary task we used in these studies was ill-suited for such an analysis because performance was near the ceiling for all subjects. To determine more clearly whether subjects who are more aware of the sequence devote some attention to sequence prediction itself would require a more sensitive secondary task, one in which, for example, responses are recorded for each secondary task signal.

Although it may be worth exploring the idea that some subjects trade attention to the tone task with attention to the primary RT task, the important conclusion for our purposes is that use of prior sequential learning is impeded when the distraction task effectively blocks attention to the relation between successive events. This observation suggests that the nonattentional form of learning is incapable of acquiring the relatively long sequences, as used in this experiment, in which no event uniquely predicts the following event.

A second concern involves an apparent conflict with data by Bullemer and Nissen (1990). Those investigators also had allowed subjects to learn an ambiguous sequence under single-task conditions similar to those used in Experiment 4. Subsequently, a distraction task was added. Bullemer and Nissen used both the RT difference and the error-rate difference between sequential and random blocks as a measure of sequential knowledge. As in Experiment 4, they divided subjects according to a post hoc assessment of explicit knowledge of the sequence. With RT as the dependent measure, they found that only subjects with full explicit knowledge of the pattern showed a sequence effect under distraction. However, subjects at all levels of explicit knowledge showed a sequence effect on error rates. Scrutiny of their results makes it unclear, however, whether less aware subjects actually show reliable error-rate differences on sequenced versus random blocks. The error rates in our experiment were all 2% or less in the critical blocks of the dual-task performance phase, making our accuracy data insensitive to such differences if they truly exist.
General Discussion

The primary question addressed by these studies was whether a form of sequential learning exists that requires little or no attention to the relation between sequential events and that is independent of a type of learning that does require attention to such relations. All four experiments were consistent with the idea of independent forms of learning. When knowledge was acquired under distraction-free conditions, performance differences depended on subjects' awareness of the sequence. Subsequently, when distraction was added, subjects continued to exhibit sequential knowledge but to the same degree regardless of prior awareness (Experiment 1). Moreover, the degree of sequential knowledge expressed under dual-task conditions was the same regardless of whether subjects originally learned with or without distraction (Experiment 2). If distraction was present during initial learning and then removed, there was no evident improvement in the expression of sequential knowledge (Experiment 3). We hypothesize that this occurs because there has been no opportunity for attentionally based learning. Altogether, these results suggest two forms of learning that occur independently and in parallel. Moreover, the attentional form of learning seems to require attention for its expression as well as its learning. When distraction was added following such learning, expression of the attentionally dependent knowledge disappeared. An exception may occur if the secondary task was not sufficient to block the deliberate usage of attention and subjects had become aware of the presence of the sequence. Then, some subjects who continued to be aware of the sequence, even when a secondary task was added, expressed sequential knowledge (Experiment 4).

Relation to Other Dichotomous Distinctions of Learning

Our conclusion for two distinct forms of learning seems on the surface much like other distinctions that have been made between procedural and declarative learning and memory (e.g., Squire, 1987; Willingham et al., 1989), between implicit and explicit learning and memory (e.g., Reber, 1989; Schacter, 1987), and between selective and nonselective learning (Berry & Broadbent, 1988; Hayes & Broadbent, 1988). Our distinction between attentional and nonattentional learning is not, however, identical to any of these. A central concern of these other distinctions involves whether learning and memory are controlled by or available to consciousness (however, Matthews et al., 1989, posited that implicit learning is accessible to awareness), whereas our distinction depended on the effects of distraction rather than on awareness.

It might be suggested that our distinction between attentional and nonattentional learning merely redescribes forms of learning that are identical to those previously identified as conscious and unconscious. Theories suggesting that consciousness is the product of an attentional mechanism (e.g., Norman & Shallice, 1986; Posner & Rothbart, 1991) invite such a view. However, numerous dissociations between attention and awareness have been found in other sequence learning experiments using various measures of attention and awareness (Bullember & Nissen, 1990; Cohen et al., 1990; Nissen & Bullember, 1987; Nissen et al., 1987). Most of these studies have used a distraction task, similar to that used in these experiments, that limits attention to the relation between successive events. It has been shown that completely ambiguous sequences cannot be learned with such distraction, but awareness has been demonstrated to be unnecessary when (a) normal subjects reported no awareness (Cohen et al., 1990; Nissen & Bullember, 1987; Willingham et al., 1989); (b) normal subjects were unable to explicitly predict the successive sequence events (Cohen et al., 1990; Nissen & Bullember, 1987; Willingham et al., 1989); (c) patients with Korsakoff's syndrome reported no awareness of the sequence (Nissen & Bullember, 1987); and (d) awareness was depressed in normal subjects by scopolamine (Nissen et al., 1987). Converging evidence has been provided by a study that used the cost of responding to out-of-sequence events as a measure of attentional allocation (Bullember & Nissen, 1990). Significant costs were found only in the absence of distraction, so it was likely that the two methods of assessing attention were measuring related processes. Importantly, cost did not vary with different levels of reported awareness. These results have led Nissen's group to argue that two types of learning, declarative and procedural, exist that differ in their availability to awareness but both depend on attention.

Willingham et al. (1989) have argued for parallel and independent development of procedural knowledge and declarative knowledge, with the latter allowing reports of awareness and explicit description of a learned sequence. However, the sequence they used was one that required attention in order to be learned. It was of a type that Cohen et al. (1990) called ambiguous, in which each event occurred more than once in a sequence and could not be uniquely predicted by its predecessor. In the experiments reported here, with the exception of Experiment 4, we used sequences that subjects could learn under distraction. It was the learning under distraction, we argued, that was independent of attentionally based learning. Together with the Willingham et al. (1989) claim for the existence of two independent forms of attentional learning, our results suggest three different types of sequence learning: nonattentional, attentional with awareness (Willingham et al., 1989, declarative), and attentional without awareness (Willingham et al., 1989, procedural).

Attention and awareness were difficult to untangle in the present experiments. Experiment 1 demonstrated that awareness exerts its effects only within the attentional domain. Also, learning was less evident when distraction was added, even for less aware subjects. This may reflect attentional learning under single-task conditions even in the absence of awareness, but this conclusion may be questionable because the size of the learning with and without distraction was statistically equated when a stricter criterion was adopted for less aware subjects. However, the statistical power to detect this difference might have been compromised by the reduced sample size.
The relation between attention and awareness became even more complicated in Experiment 4. It appears that awareness of the sequence might have allowed some subjects to overcome the effects of distraction. Another possibility is that the secondary task ineffectively blocked attention and that the opportunity for attentional allocation allowed some subjects to regain awareness of the sequence. We believe that either of these conclusions points to the ineffectiveness of the distraction task for completely blocking attention. The only other alternative—one that we feel is unlikely given the results of Experiment 1—is that the effects of awareness can operate without attention. We cannot presently solve these puzzles. The conclusion we wish to draw concerning the relation between attention and awareness is that awareness enhances only an attentional form of sequence learning. Our data are suggestive, but not conclusive, regarding two forms of attentional learning, one form requiring and the other not requiring awareness.

The role of attention has been central to the distinction between selective and nonselective modes of learning (Berry & Broadbent, 1988; Hayes & Broadbent, 1988), with the former being subject to interference from distraction and the latter insensitive to distraction. This distinction is similar to our distinction between attentional and nonattentional learning. We have argued that the two forms are independent, whereas Hayes and Broadbent suggested that they compete with one another, with the success of one being at the expense of the other. It is likely, though, that this competition occurs at the level of response selection rather than at the level of learning. In the procedure used by Hayes and Broadbent, subjects had to make one and only one choice of response per trial. Prior to mastery, the choice was often wrong. In this situation, if nonattentional learning favors one response and attentional learning another, then there is competition for responses, even though the two knowledge sources may be independent. In the paradigm we used, there was no response competition. The visual signal dictated the response. Attentional and nonattentional knowledge sources presumably prime the response. This may explain why our results indicate independent knowledge sources.

**Nature of Attentional and Nonattentional Learning**

An important issue concerns whether the attentional and nonattentional forms of learning differ in the kinds of sequential knowledge they acquire. Hayes and Broadbent (1988) suggested that their nonselective mode learns all contingencies between successive responses. The selective mode essentially operates on a hypothesis that predicts the next response to the exclusion of other responses. These two forms of learning are reminiscent of an old idea in concept learning: that some forms of concept discovery proceed by hypothesis (e.g., Levine, 1966) and other forms proceed by accretion of knowledge (e.g., Hull, 1920; Posner & Keele, 1968). Similarly, it has been proposed that implicit learning is memory based, whereas explicit learning is rule based (Matthews et al., 1989; Stanley, Matthews, Buss, & Kotler-Cope, 1989).

We suggest that nonattentional learning operates primarily by simple associations, whereas attentional learning has a mechanism for encoding the position of an event within the sequence. This is supported by the inability of subjects to exhibit sequential knowledge with completely ambiguous associations under distraction in Experiment 4, even though sequences with a single unique event were learned under distraction in Experiment 2. Cohen et al. (1990) manipulated sequence structure within a single experiment and also found that attention was necessary to learn sequences completely composed of ambiguous associations. Earlier research concurs that keeping track of position within a sequence requires attention (Keele & Summers, 1976), and more recent work suggests that subjects are sensitive to the positional dependencies in artificial grammar only when they are free from distraction (Dienes et al., 1991). It has previously been argued that attentional learning is able to represent positional information by parsing the sequence into subcomponents and representing these components hierarchically (Cohen et al., 1990). Keele and Jennings (1992) simulated such a hierarchical learning system with a connectionist model developed by Jordan (in press). These simulations have provided computational support for these views. Further research is needed to show that such a mechanism actually operates within the brain.

It is not only important to ask about the nature of the learning mechanism that is supported by attention but also about the nature of the attentional mechanism itself. Our research did not address important questions concerning the specific role of attention. The inconsistency of the results concerning the effects of distraction on ambiguous sequence learning highlights the need for a more thorough understanding of the nature of the processes that are disabled by distraction. Attentional learning may benefit from attentional resources that modulate a wide variety of computations or from attentional resources that are selectively required by the learning mechanism (see Allport, 1989, for a review of selective vs. general theories of attention). For example, Dienes et al. (1991) suggested that distraction interferes with an articulatory loop (Baddeley, 1986) that holds positional information. A similar view of working memory loading, although not necessarily articulatory, could account for our results. Such questions can be properly addressed only by future experiments that systematically vary the distraction task. Knowledge of the specific effects of various distraction tasks will allow more specific hypotheses concerning the nature of the attentional mechanism that supports sequence learning.

**On the Separability of Attentional and Nonattentional Learning**

The demonstration that sequential knowledge can be learned by independent attentional and nonattentional mechanisms invites questions concerning the level at which these forms of learning are separable. At one extreme, it might be hypothesized that completely distinct neural and cognitive systems underlie these two forms of learning. A common alternative to hypotheses of multiple memory
systems invokes the notion of compatibility between the conditions of encoding and retrieval (e.g., Hintzman, 1990; Roediger, Weldon, & Challis, 1989). By this logic, the same system supports both attentional and nonattentional learning, but attention enhances sequence learning effects only if it is available during both original learning and later performance. For example, it might be supposed that attention provides additional retrieval cues that are blocked by distraction, but the basic storage medium is the same with and without attention. Similar interactions between the presence or absence of attention at encoding and retrieval have been previously observed (Baddeley, Lewis, Eldridge, & Thompson, 1984; Jacoby, 1991). Such a hypothesis may be able to account for the results of our first three experiments, but not for the interactions between attention and sequence structure. It does not explain why adding distraction following learning abolished learning of the ambiguous sequences of Experiment 4 but not the hybrid sequences of earlier experiments.

A connectionist model of sequence learning recently developed by Cleermans and McClelland (1991) was able to account for the interaction between attention and sequential structure (Cohen et al., 1990) using a single learning system. Distraction was simulated by the introduction of random noise into network connections. This simulated distraction disproportionately retarded the network’s ability to learn ambiguous sequences. Further simulations will be needed to assess this network’s ability to model the present experiments. One aspect of our results that may be particularly challenging to simulate is the effects of awareness on attentional learning. At present, we see no mechanism in the model for awareness and hence no mechanism for simulating Experiment 1.

Our experiments suggest that attentional and nonattentional learning operate independently, are parallel, do not share information, and differ in the kinds of sequences they can represent. It may be that these forms of learning are carried out by functionally distinct neural systems that do not share a common knowledge base. Alternatively, a common system may exist that is purely associative when attention is unavailable but interacts with some other mechanism to form hierarchical representations when attention is available. Such a view is inspired by Wicklegren’s (1979) theory of chunking. Wicklegren proposed that simple associative learning can proceed through horizontal associations between low-level memory nodes. Chunking is a product of vertical associations formed between these same low-level nodes and high-level nodes. Wicklegren speculated that chunking was dependent on limbic arousal. We doubt that the limbic system is important in the present case because Nissen and Bullemer’s (1987) patients with Korsakoff’s syndrome exhibited attentional learning of ambiguous sequences. However, if we posit that attention rather than limbic arousal is the binding force behind vertical associative memories, then such a theory may account for our results.

At present, we are not committed to either a view of completely independent attentional and nonattentional learning systems, or to a view with some common components but others differing and preventing cross-transfer of information. However, we can argue that attention enables some additional process or mechanism that changes the nature of learning and leads to representational differences.

References


Received January 23, 1992
Revision received April 21, 1992
Accepted May 19, 1992