

What Is Learned During Automatization? The Role of Attention in Constructing an Instance

Gordon D. Logan and Joseph L. Etherton

Seven experiments were conducted to examine the role of attention in automatization. Ss searched 2-word displays for members of a target category in divided-attention, focused-attention, and dual-task conditions. The main issue was whether attention conditions would affect what Ss learned about co-occurrences of the words in the displays. The attention hypothesis, derived from the instance theory of automaticity, predicts learning of co-occurrences in divided-attention and dual-task conditions in which Ss attend to both words but not in focused-attention conditions in which Ss only attend to 1 word. The data supported the attention hypothesis and therefore the instance theory.

This article concerns what is learned during automatization. This is an important question in the automaticity literature, especially from the perspective of memory-based theories, such as the instance theory of automaticity (Logan, 1988, 1990, 1992). Memory-based theories assume that automatic performance is based on retrieval of representations of past solutions from memory. What “gets into” those representations during learning and what is “taken out” of them during automatic performance are central questions in memory-based theories. The purpose of this article is to evaluate the answers offered by the instance theory of automaticity. The answers are important because they are derived from two of the three main assumptions of the theory: obligatory encoding and instance representation. The *obligatory encoding* assumption says that attention determines what gets into the representation. The *instance representation* assumption says that the representations in memory are instances—separate representations of co-occurrences. Attention determines what is in an instance; attention determines which co-occurrences are remembered. The experiments tested this hypothesis.

What Is an Instance?

The instance representation assumption consists of two related assumptions. The first is that each event is represented

separately. The second is that each event is represented as a constellation of co-occurrences of the features, properties, objects, and so forth, that comprise it. Previous investigations of instance theory focused on the idea that instances were separate, deriving and testing predictions about the power law of learning and extending it beyond mean reaction times to entire distributions of reaction times (Logan, 1988, 1992). This article focuses on the idea that instances represent co-occurrences, identifying instances by the co-occurrences they represent. The role of attention in constructing an instance is inferred from its effect on learning about co-occurrences.

Instance theories of all kinds assume that instances represent co-occurrences, usually co-occurrences of features or properties in specific examples (Barsalou, 1990; Hintzman, 1976, 1986; Jacoby & Brooks, 1984; Medin & Schaffer, 1978). In the instance theory of automaticity, instances represent co-occurrences because they represent processing episodes. Processing episodes consist of the goal the subject was trying to attain, the stimuli encountered in pursuit of that goal, the interpretation given to the stimuli with respect to the goal, and the response made to the stimulus (Logan, 1988, p. 495). Each episode represents a specific combination of goals, stimuli, interpretations, and responses that occurred together on a specific occasion. The present experiments asked how attention determines which co-occurrences are represented in an episode (also see Logan, 1990).

Evidence that co-occurrence is represented is not necessarily evidence for instance representations. All instance theories assume that instances represent co-occurrence. Existing strength theories typically do not, but in principle, strength theories could be constructed that represent co-occurrence (Barsalou, 1990). Thus, if we were to decide “instance representation” given evidence that co-occurrence was represented, our hit rate would be 1.0, but our false-alarm rate could be greater than zero. In the present experiments, we sought converging evidence for instance representation by testing the power function predictions, which depend on the idea that instances are separate, not on the idea that instances represent co-occurrences (see the *Evidence of Attention and Automaticity* section).

Obligatory Encoding and the Attention Hypothesis

The obligatory encoding assumption states that encoding into memory is an inevitable consequence of attending. What

Gordon D. Logan and Joseph L. Etherton, Department of Psychology, University of Illinois at Urbana-Champaign.

This research was supported by National Science Foundation Grants BNS 88-11026 and BNS 91-09856. The article was written while Gordon D. Logan was a visiting professor at the University of Amsterdam. We are grateful to the developmental psychology group for their hospitality and support. We are grateful to Jane Zbrodoff and the Wednesday night lab group for constructive criticism at all stages of the project. We are grateful to Julie Delheimer for help in testing subjects, analyzing the data, and managing intercontinental communication. We are grateful to Stan Taylor and Russ Poldrack for help with the analyses and comments on this article. We also thank Colin MacLeod, Tom Carr, and an anonymous reviewer for helpful comments on this article.

Correspondence concerning this article should be addressed to Gordon D. Logan, Department of Psychology, University of Illinois, 603 East Daniel Street, Champaign, Illinois 61820. Electronic mail may be sent to glogan@s.psych.uiuc.edu.

is attended is encoded; what is attended is represented in memory. Previous investigations of instance theory focused on the quantitative consequences of this assumption, explaining how the accumulation of instances through separate acts of attention accounts for the power law of learning (Logan, 1988, 1992). This article concerns the qualitative consequences of the assumption, focusing on what people attend to and how that determines what co-occurrences they learn. The experiments test an *attention hypothesis*, derived from the obligatory encoding assumption. The attention hypothesis predicts that people will learn the co-occurrences they attend to. Attention is sufficient for learning co-occurrences; it may even be necessary.

The attention hypothesis is important because it is testable. Attention can be manipulated experimentally, using well-understood techniques developed from a half-century of research (Broadbent, 1958; Kahneman, 1973). The effects of the manipulations can be measured in immediate performance, independent of learning, so the effects of attention on learning can be defined without circularity. The attention hypothesis can be tested by manipulating attention and co-occurrence at the same time. The attention hypothesis predicts that subjects will learn about co-occurrence when they attend to it.

Attention can be manipulated in many ways. Attention is selective in many different senses, and there are manipulations corresponding to each sense (e.g., Posner & Boies, 1971; Treisman, 1969). Logan (1990) manipulated *analyzer selection*, holding stimuli constant and varying the task (analysis) subjects performed on them. The present experiments manipulated *input selection*, presenting and withholding cues that controlled which parts of the display subjects attended to. The attention hypothesis predicts that subjects will learn co-occurrences of attended parts but not co-occurrences of unattended parts.

Testing the Attention Hypothesis

Consistent Context

In the present experiments, subjects searched through two-word displays for members of a target category (e.g., metals). Co-occurrence was manipulated by pairing particular words consistently throughout the training phase of the experiment (e.g., every time *Steel* appeared, it was paired with *Canada*) or varying the pairing from presentation to presentation (e.g., *Steel* might appear with *Canada* on one presentation and with *Broccoli* on the next). If subjects are sensitive to co-occurrence, they should learn which words went together on each trial, whether pairing is consistent or varied. However, consistent pairing provides subjects with the opportunity to take advantage of their knowledge of co-occurrence because they see the same pairs again. By contrast, varied pairing denies subjects the opportunity to take advantage of their knowledge of co-occurrence because the pairings are not repeated.

The effect of consistency of pairing was assessed in two ways: First, consistent-pairing subjects were compared with varied-pairing subjects. If subjects are sensitive to co-occurrence, then consistent-pairing subjects should perform better than varied-

pairing subjects. Second, consistent-pairing subjects were given a transfer test consisting of new pairings of old words. If subjects are sensitive to co-occurrence, then changing the pairing after consistent training should disrupt performance considerably. The strategies developed to exploit knowledge of consistent pairing will no longer support performance, and new strategies will have to be developed to cope with the task.

Attention

Attention was manipulated between experiments. All subjects saw the same displays, but their attention was directed to them in different ways by varying instructions. There were three attention conditions: divided attention, focused attention, and dual task. In the *divided-attention* condition (Experiments 1 and 2), subjects searched for members of a single target category but were not told which of the two words would be the target. Consequently, they had to pay attention to both words, or divide attention between words. In the *focused-attention* condition, subjects again searched for members of a single target category, but this time the position of the target was cued by coloring it (Experiments 3 and 4) or pointing to it with an arrowhead (Experiment 5). In the *dual-task* condition (Experiments 6 and 7), subjects again saw pairs of words, but this time they had two search tasks. They reported presence or absence of members of one category with one hand and reported presence or absence of members of another category with the other hand.

These manipulations of attention have (at least) two effects that can be measured independently of each other: effects on immediate performance, measured by reaction time and error rate on the current trial, and effects on learning, measured by performance on subsequent trials. Although the major focus of this article is on the effects of attention on learning, the immediate performance measures are important because they provide converging evidence of the operation of attention.

Traditionally, immediate performance measures are assumed to reflect the amount of "work" attention has to do. A common consensus and a lot of data suggest that there is more work to do in divided attention than in focused attention and more work to do in dual-task conditions than in divided-attention conditions (see, e.g., Broadbent, 1958; Kahneman, 1973). Thus, immediate performance should be worse in divided attention than in focused attention and worse in dual-task conditions than in divided-attention conditions. The data confirm these predictions.

The main focus of this article is on the effects of attention on subjects' sensitivity to co-occurrence. The attention hypothesis makes different predictions for the different attention conditions: It predicts that subjects will be sensitive to co-occurrence in the divided-attention experiments because the task requires them to attend to both words. It predicts that subjects will not be sensitive to co-occurrence in the focused-attention experiments because the cue allows them to ignore the context word. It predicts that subjects will be sensitive to co-occurrence in the dual-task experiments because the tasks require them to attend to both words. If subjects attend to both words successfully, they should learn their co-occurrence.

Automaticity

The predictions focus on the effects of a single act of attention on subsequent performance. In principle, they can be tested with as few as two presentations or as many as two thousand. Consequently, they should apply to the explicit- and implicit-memory literature as well as the automaticity literature. We focused on automaticity because we wanted to test the instance theory (Logan, 1988, 1990, 1992). The predictions were tested at two different levels of practice. In some experiments we used a single training session in which each stimulus was presented 16 times. In other experiments we used four training sessions to present each stimulus 64 times. In theory, sensitivity to context should appear at all practice levels, and the practice manipulation allows a test of that hypothesis. The instance theory (Logan, 1988, 1990, 1992) says that automaticity is a continuum, not a dichotomy, so the same qualitative effects should be observed at different degrees of automaticity (also see J. Cohen, Dunbar, & McClelland, 1990; Logan, 1985; MacLeod & Dunbar, 1988; Schneider, 1985; Schneider & Shiffrin, 1985). The effects may differ quantitatively—they may shrink or grow with practice—but they should be the same qualitatively despite the changes in scale. Varying context should have effects at both practice levels, though the effects may be stronger after four sessions than after one.

Automaticity was produced by practicing subjects under consistent stimulus-to-response mapping conditions (Shiffrin & Schneider, 1977): Subjects searched for members of the same category throughout the experiment, and the same set of exemplars was presented repeatedly. It is important to note that consistent mapping is not the same as consistent pairing. Consistent and varied pairing both involve consistent mapping. No target became a distractor and no distractor became a target in either condition. Pairing conditions differ in stimulus-to-stimulus mapping, not stimulus-to-response mapping. Consequently, automaticity should occur in both pairing conditions. The issue was whether there would be more automaticity (more learning) with consistent pairing than with varied pairing.

Automaticity was assessed in three ways. First, each experiment tested for a power function speedup during the training phase. In virtually all speeded tasks, reaction time decreases as a power function of practice, and the power function speedup is interpreted as a hallmark of automaticity (Logan, 1988, 1992; Newell & Rosenbloom, 1981). Second, the focused- and divided-attention experiments were compared to see if the difference between them decreased with practice. The processing load is greater in divided attention than in focused attention, and in search tasks, a reduction in processing load effects with practice is characteristic of automaticity (Schneider & Shiffrin, 1977). Third, the dual-task and divided-attention conditions were compared to see if the difference between them diminished with practice. In many tasks, a reduction in dual-task interference with practice is characteristic of automaticity (Bahrick & Shelly, 1958; Logan, 1979).

To avoid redundancy, we do not report tests of the criteria for automaticity as we discuss each experiment. The power function fits and the effects of practice on differences between attention conditions are presented in a section of their own

before the General Discussion section, after the experiments are discussed individually (see the *Evidence of Attention and Automaticity* section).

Experiment 1: Single-Session Divided Attention

The first experiment involved divided attention. Subjects searched for members of a target category in displays of two words. The position of the target was not cued. Subjects were trained for 16 blocks of trials and then transferred to a 17th block, all in one session. There were two groups of subjects: a consistent-pairing group and a varied-pairing group. For the consistent-pairing group, words were paired in the same way throughout the training blocks. Pairing was changed only in the transfer block. For the varied-pairing group, pairing was changed every block throughout training and transfer. A given word would be presented with one word on one presentation and another word on the next. The attention hypothesis predicts that subjects will be sensitive to the consistency of pairing because the task requires them to attend to both words. The two groups should start out performing at the same level because there is no difference between the conditions until the second block, when pairing either changes or stays the same. After that, consistent-pairing subjects should perform better than varied-pairing subjects, reflecting their knowledge of consistent co-occurrence. At transfer, when the pairing changes for both groups, the consistent-pairing group should have more difficulty than the varied-pairing group because the knowledge of co-occurrence they acquired and exploited during training can no longer support performance on the task.

Method

Subjects. Two groups of 32 subjects were recruited from the introductory psychology subject pool at the University of Illinois. They received course credit for participating in the experiment.

Apparatus and stimuli. The stimuli were 64 words drawn from four categories from the Battig and Montague (1969) norms. There were 16 words in each category. The categories were *metals*, *countries*, *vegetables*, and *articles of furniture*. The words are presented in the Appendix. The categories were matched with respect to frequency of mention in the Battig and Montague norms, prototypicality in the Uyeda and Mandler (1980) norms, word frequency in the Kucera and Francis (1967) norms, and word length in number of letters. These measures are presented in Table 1. There were no significant differ-

Table 1
Means and Standard Deviations of Measures for Each Word Category From the Experiments

Measure	Metals		Countries		Vegetables		Furniture	
	M	SD	M	SD	M	SD	M	SD
Word frequency ^a	18.9	15.2	51.4	60.9	8.8	11.7	47.8	64.5
Frequency of mention ^b	160	110	145	99	161	91	153	144
Prototypicality ^c	2.28	0.97	2.27	0.36	2.52	0.53	2.43	0.83
Word length	5.81	1.80	6.56	1.55	6.63	1.93	5.63	1.89

^aFrom Kucera and Francis (1967). ^bFrom Battig and Montague (1969). ^cFrom Uyeda and Mandler (1980).

ences between categories in any of these measures, except for word frequency, in which the difference between the highest and lowest frequency categories was statistically significant. This difference was not a problem: Word frequency is not an important variable in category verification tasks (Balota & Chumbley, 1984), and we counterbalanced assignment of categories to experimental conditions.

The words were displayed on Amdek model 722 color monitors driven by IBM PC XT and AT computers. There were four computers, each facing a different wall of a large room so that subjects could be tested in groups on their own computers without distracting each other.

Two words were displayed in each trial, one above the other. The words were presented in the center of the screen but left justified such that their initial letters were aligned. Their initial letters appeared in column 33 of row 12 and row 13 on the standard 80 × 24 IBM text screen. The words were written in lowercase except for the initial letter. The initial letter of country names had to be capitalized, so we capitalized the initial letter of all of the words. Viewed at a distance of 60 cm, single words subtended 0.48° of visual angle in height and a minimum of 0.76° and a maximum of 2.29° in length. The two words together subtended about 1.14° of visual angle vertically.

Each word pair was preceded by a fixation and warning display. It consisted of two lines of seven dashes centered on the screen. One line of dashes appeared on the screen line above the line that would contain the top word (i.e., row 11, columns 32–38), and the other line of dashes appeared one screen line below the line that would contain the bottom word (i.e., row 14, columns 32–38). Viewed at a distance of 60 cm, the fixation and warning display subtended 1.62° of visual angle horizontally and 1.72° vertically.

Each trial began with the fixation and warning display exposed for 500 ms. Then that display was extinguished and immediately replaced by the word pair for that trial, which was exposed for 1,000 ms. Then the screen went blank for 2,000 ms until the next trial began. Subjects responded by pressing the *z* and slash keys on the keyboards of the computers. These were the leftmost and rightmost keys on the bottom row of the AT keyboard and nearly so for the XT keyboard. Subjects' responses could be registered at any time during a 3,000-ms window from the onset of the word pair to the onset of the fixation and warning display for the next trial.

Procedure. The experiment was organized around blocks of 32 trials, in which the 64 words were paired and each pair was presented once. In the consistent-pairing group, words were paired randomly at the beginning of the experiment, and the pairing remained the same throughout training. The order in which the pairs were presented was randomized for each block, but the pairing remained the same, changing only at transfer. In the varied-pairing group, pairing changed with each block. The order of trials within blocks was randomized for each block. Both groups received 16 training blocks and 1 transfer block, so that each of the 64 words was presented 17 times. Thus, the frequency with which individual words were presented could not contribute to differences between groups in training or transfer.

There were two basic trial types, target-present trials and target-absent trials, and 16 of each type were presented in each block. On *target-present* trials, one word was selected from the target category, and one word was selected from one of two distractor categories. On *target-absent* trials, one word was selected from a fourth, nontarget category, and the other word was selected from one of two distractor categories. We used each of the four categories (metals, vegetables, countries, and furniture) equally often as targets, nontargets, and each of the two distractor categories. The categories were assigned to these roles with a balanced Latin square.

Members of the nontarget category were never presented together with members of the target category, so the presence or absence of a nontarget predicted presence or absence of a target perfectly. Consequently, subjects could have treated the task as a forced-choice

discrimination between the target and nontarget category rather than target detection. We risked the ambiguity of interpretation this might produce because it allowed us to present distractors from categories that were independent of presence or absence of the target category. The two distractor categories occurred equally often on target-present and target-absent trials, so distractor category membership could not predict the presence or absence of a member of the target category. This was important because we wanted to see whether subjects would learn to associate specific distractors with specific targets and nontargets. In principle, consistent pairing allows subjects to predict the presence or absence of a specific member of the target category from the presence or absence of a specific member of one of the two distractor categories. Our measure of learning (improved performance) required that subjects somehow exploit that predictive relation to reduce reaction time and increase accuracy. We balanced distractor categories as we did so that no other predictive relation would compromise our pairing manipulation.

Targets appeared equally often in the top and bottom positions in the display, as did nontargets and members of each of the two distractor categories. However, specific words were presented consistently in one position or the other. For example, if *Canada* were on top and *Steel* were on the bottom in the first block, they remained in those positions for the entire experiment, in both training and transfer. Position was held constant in both varied- and consistent-pairing conditions (e.g., *Canada* and *Steel* always appeared in the same positions whether or not they appeared with each other). We investigated the effects of changing position in a subsequent article and found they were small, relative to the effects of changing pairing (see Etherton, 1992).

Half of the subjects indicated target presence with their right hands and target absence with their left hands, and half did the opposite. Assignment to response-mapping conditions was orthogonal to the counterbalancing of categories.

Subjects were given written instructions that described the task, told them the name of the target category, and told them which keys to press to indicate target presence and target absence. Subjects were told to rest the index fingers of their right and left hands lightly on the response keys. Subjects were not told the number or the nature of the other categories, and they were not told about the consistency of pairing or the lack of it. After subjects read the instructions, the experimenter summarized them and answered questions. The subjects then began the experiment. They were allowed brief rests every 128 trials (four blocks). The last rest was just before the transfer trials. Immediately on completing the transfer trials, subjects were given a surprise recall task in which they were asked to recall all of the words that had appeared in training and transfer: targets, nontargets, and distractors. They were asked to write down the words they recalled in pairs if they could, but they were told it was important to recall as many words as they could, whether or not they were in pairs.

Results

Training. Mean reaction time and percentage correct were calculated for target-present and target-absent displays in each practice block for each subject in the consistent- and varied-pairing groups. The reaction time data are plotted as a function of practice in Figure 1.

Consistent- and varied-pairing subjects started out at about the same level of performance; mean reaction time on the first presentation was 836 ms for consistent pairing and 852 ms for varied pairing. Both groups improved with practice. Reaction time decreased as a negatively accelerated function of practice, characteristic of the power function speedup (fits are reported

in Tables 12 and 13). The consistent-pairing group gained more from practice, however. On the 16th presentation, mean reaction time was 637 ms for the consistent-pairing group and 686 ms for the varied-pairing group. These results suggest the sensitivity to co-occurrence that was predicted by the attention hypothesis.

These conclusions were supported by a 16 (presentations) \times 2 (target present vs. target absent) \times 2 (group: consistent pairing vs. varied pairing) analysis of variance (ANOVA) on the mean reaction times. The ANOVA revealed a significant main effect of presentations, $F(15, 930) = 67.58, p < .01, MS_e = 3,940.90$, a significant main effect of target presence, $F(1, 62) = 26.90, p < .01, MS_e = 7,855.43$, and a significant interaction between them, $F(15, 930) = 8.84, p < .01, MS_e = 1,172.91$. The interaction reflected a larger practice effect on

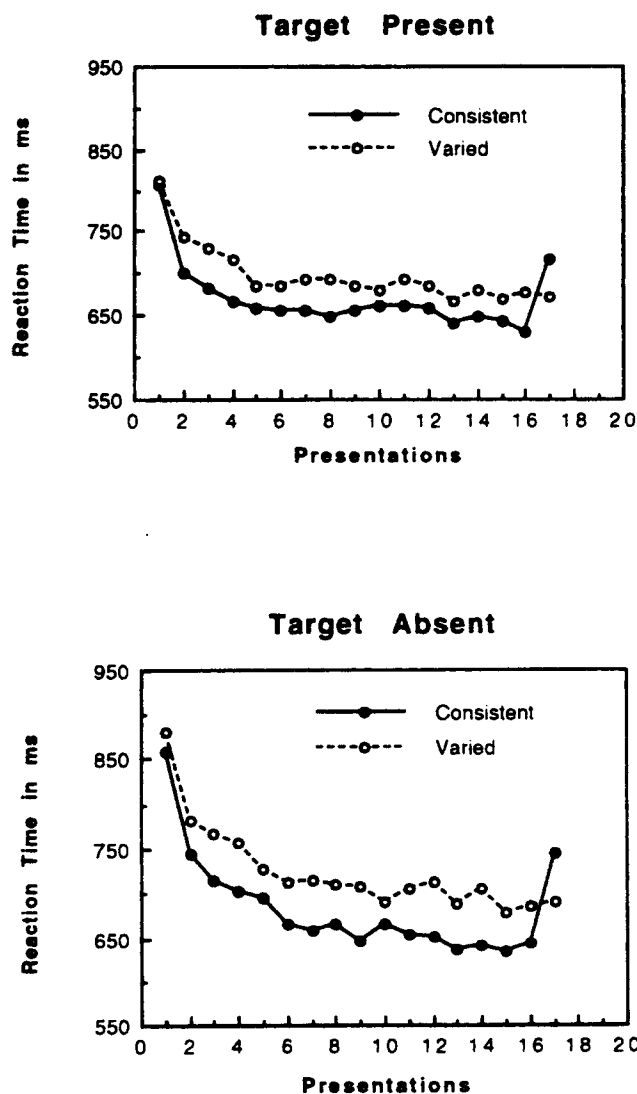


Figure 1. Mean reaction time as a function of the number of presentations for target present (top of figure) and target absent (bottom of figure) responses in Experiment 1 (divided attention). Solid lines = consistent pairing; broken lines = varied pairing.

Table 2
Percentage Correct in Each Condition of Experiment 1
(Single-Session Divided Attention) as a Function of
Number of Presentations

No. of presentations	Consistent pairing		Varied pairing	
	Present	Absent	Present	Absent
1	86	90	87	91
2	96	97	93	98
3	95	98	94	98
4	95	97	94	98
5	94	97	93	96
6	97	98	93	97
7	97	98	94	98
8	96	96	94	98
9	97	98	95	98
10	98	97	94	99
11	96	98	95	99
12	98	98	94	99
13	97	97	95	98
14	97	98	95	99
15	94	98	93	99
16	96	98	93	98
17	93	95	94	98

target-absent trials. The main effect of consistency of pairing approached significance, $F(1, 62) = 3.69, p < .06, MS_e = 229,735.54$. The critical interaction between pairing and practice was not significant, $F(15, 930) < 1, MS_e = 3,940.90$, but a planned comparison showed that the 49-ms difference on the 16th presentation was significantly larger than the 16-ms difference on the 1st presentation, $F(1, 930) = 8.57, p < .01, MS_e = 3,940.90$.

The accuracy data are presented in Table 2. Accuracy was generally high, averaging 95%. It correlated negatively with reaction time, corroborating the reaction time results. A 16 (presentations) \times 2 (target present vs. target absent) \times 2 (group: consistent pairing vs. varied pairing) ANOVA was performed on the accuracy data. It revealed a significant main effect of presentations, $F(15, 930) = 23.78, p < .01, MS_e = 23.23$; a significant main effect of target presence, $F(1, 62) = 23.29, p < .01, MS_e = 156.04$; and a significant interaction between them, $F(15, 930) = 1.84, p < .05, MS_e = 19.93$, corroborating the effects in the reaction time ANOVA. The only other significant effect was the interaction between target presence and consistency of pairing, $F(1, 62) = 5.80, p < .05, MS_e = 156.04$, reflecting a larger difference between target present and target absent with varied pairing.

Transfer. Mean reaction times from the transfer blocks (Block 17) also appear in Figure 1. Reaction time increased sharply from training to transfer for consistent-pairing subjects, averaging 637 ms before pairing changed and 736 ms after. The change from training to transfer hardly affected the varied-pairing subjects, who had seen 15 similar changes already. Reaction time averaged 686 ms before the final change and 689 ms after it. Consistent-pairing subjects lost more than the 49-ms advantage that they had gained over the varied-pairing subjects during training: They were 47 ms slower than varied-pairing subjects at transfer. These results also suggest the sensitivity to co-occurrence that was predicted by the attention hypothesis.

These conclusions were confirmed in a 2 (training block vs. transfer block) \times 2 (target present vs. target absent) \times 2 (group: consistent pairing vs. varied pairing) ANOVA on the mean reaction times. There was a significant main effect of transfer, $F(1, 62) = 49.32, p < .01, MS_e = 3,429.09$; a significant main effect of target presence, $F(1, 62) = 10.67, p < .01, MS_e = 2,629.21$; and a significant interaction between transfer and consistency of pairing, $F(1, 62) = 41.97, p < .01, MS_e = 3,429.09$. Nonorthogonal comparisons revealed a significant effect of transfer for the consistent-pairing group, $F(1, 62) = 91.46, p < .01, MS_e = 3,429.09$, but not for the varied-pairing group ($F < 1.0$).

The accuracy data from the transfer blocks are presented in Table 2. As in training, the accuracy data confirm the conclusions drawn from the reaction times. Accuracy decreased from training to transfer in the consistent-pairing subjects (from 96.8% to 94.3%), whereas it increased slightly from training to transfer in the varied-pairing subjects (from 95.6% to 95.9%).

These conclusions were confirmed in a 2 (training block vs. transfer block) \times 2 (target present vs. target absent) \times 2 (group: consistent pairing vs. varied pairing) ANOVA on the accuracy data. The main effect of transfer was not significant, $F(1, 62) = 2.40, p > .10, MS_e = 31.01$, but the main effect of target presence was significant, $F(1, 62) = 17.12, p < .01, MS_e = 35.78$. The interaction between transfer and consistency of pairing just missed significance, $F(1, 62) = 3.99, p < .06, MS_e = 31.01$. Nonorthogonal comparisons revealed a significant effect of transfer for the consistent-pairing group, $F(1, 62) = 6.29, p < .01, MS_e = 31.01$, but not for the varied-pairing group ($F < 1.0$).

Recall. The number of words recalled in target, nontarget, and distractor categories was calculated for each subject, and the numbers were converted to percentages. We collapsed the two distractor categories, as there was no theoretical distinction between them. Subjects in the consistent-pairing group recalled 62.7% of their 16 targets, 40.0% of their 16 nontargets, and 41.8% of their 32 distractors. Subjects in the varied-pairing group showed a similar pattern, with slightly lower performance. They recalled 57.2% of their targets, 31.6% of their nontargets, and 30.3% of their distractors. A 3 (target vs. nontarget vs. distractor) \times 2 (group: consistent pairing vs. varied pairing) ANOVA on the number of words recalled per category (maximum words = 16) revealed a significant main effect of target, $F(2, 124) = 84.10, p < .01, MS_e = 3.75$, and a significant main effect of consistency of pairing, $F(1, 62) = 5.85, p < .05, MS_e = 15.05$, but no significant interaction between them ($F = 1.0$). A planned comparison on the data averaged across pairing conditions revealed a significant difference between targets and the average of nontargets and distractors, $F(1, 124) = 168.19, p < .01, MS_e = 3.75$, suggesting more attention was paid to targets than to nontargets and distractors. A second planned comparison on the same data revealed no significant difference between nontargets and distractors ($F < 1.0$), which suggests equivalent attention to them. The difference between targets and the average of nontargets and distractors was smaller in the consistent-pairing group ($M = 21.8\%$) than in the varied-pairing group ($M = 26.3\%$), suggesting more attention to

nontargets and distractors in the consistent-pairing group (i.e., sensitivity to co-occurrence), but a contrast evaluating the effect was not significant, $F(1, 124) = 2.94, MS_e = 3.75$.

Discussion

The results suggest that subjects were sensitive to the co-occurrence of the words in the displays. Consistent-pairing subjects improved more with practice than varied-pairing subjects did. Consistent-pairing subjects suffered more than varied-pairing subjects when pairing changed at transfer. These results confirm the attention hypothesis, which predicts sensitivity to context under divided attention conditions.

It is interesting that consistent-pairing subjects were worse than varied-pairing subjects at transfer. This suggests that the two groups used different strategies. Consistent-pairing subjects somehow relied on the regular co-occurrence of words to support their performance, and that strategy would not work when the pairing changed. They would have to find a new basis for performance. Their search for it is reflected in the exceptionally long reaction times at transfer.

We performed an additional analysis on the transfer data to gain some insight into the associations that consistent-pairing subjects acquired during training. Distractors could have been associated with the targets and nontargets they were paired with, forming *stimulus-stimulus* associations. For example, if *Canada* were a distractor consistently paired with *Steel*, subjects would associate *Canada* with *Steel*. Alternatively, distractors could have been associated with the response that was appropriate to the display, forming *stimulus-response* associations.¹ In the previous example, *Canada* would be associated with a yes response because it appeared in a target-present display.

These alternatives make different predictions about the effects of changing pairing on transfer: If subjects learned stimulus-stimulus associations, performance should be poor regardless of the pairing. If subjects learned stimulus-response associations, performance with new pairings would depend on the responses that were associated with each member of the pairs during training. If the responses required for the new displays were compatible with the associated responses, there should be no disruption in performance. Responses are compatible if a distractor that was previously paired with a target is now paired with a different target or if a distractor that was previously paired with a nontarget is now paired with a different nontarget (for examples, see Table 3). If the responses required for the new display were incompatible with the associated responses, performance should be disrupted substantially. Responses are incompatible if a distractor that was previously paired with a nontarget is now paired with a

¹ In stimulus-response associations, responses refer to response categories, such as target present and target absent, not physical responses, such as key presses. Data from tasks similar to category search suggest that automaticity depends on associations between stimuli and response categories, independent of physical responses (Fisk & Schneider, 1984; Logan, 1990; Shiffrin & Schneider, 1977, Experiment 3).

target or if a distractor that was previously paired with a target is now paired with a nontarget (for examples, see Table 3).

To test these predictions, we partitioned the transfer data according to target presence and compatibility with training displays (see Table 3). The original program did not guarantee an equal number of compatible and incompatible trials at transfer because pairings were scrambled randomly. Nevertheless, each subject contributed some data to each cell of the design, so the data could be analyzed with conventional statistics (i.e., ANOVA). The means across subjects along with the means from the last training block are presented in Table 3 as a function of target presence and compatibility. The data suggest that subjects learned both stimulus-stimulus and stimulus-response associations.

Stimulus-stimulus associations were assessed by comparing transfer performance on compatible displays with performance on the last training block, in which displays were also compatible. On average, reaction time to compatible displays was 698 ms, compared with 637 ms in the last training block, which is consistent with the idea that subjects learned stimulus-stimulus associations. Planned comparisons showed that the difference was significant for target-present displays, $F(1, 62) = 34.40, p < .01, MS_e = 2,538.10$, and for target-absent displays, $F(1, 62) = 15.08, p < .01, MS_e = 2,538.10$.

Stimulus-response associations were assessed by comparing transfer performance on compatible displays with transfer performance on incompatible displays. On average, mean reaction time was 768 ms for incompatible displays, compared with 698 ms for compatible displays, which is consistent with the idea that subjects learned stimulus-response associations. Planned comparisons showed that the difference was significant for target-present displays, $F(1, 62) = 7.79, p < .05, MS_e = 2,538.10$, and for target-absent displays, $F(1, 62) = 67.00, p < .01, MS_e = 2,538.10$.

Stimulus-stimulus associations are important because they represent co-occurrence, which is characteristic of instance representation (Barsalou, 1990; Hintzman, 1976, 1986; Jacoby & Brooks, 1984; Medin & Schaffer, 1978). Thus, evidence of stimulus-stimulus associations is evidence for the instance theory. It suggests that automatic performance is based on retrieved instances.

Stimulus-response associations are harder to interpret. On the one hand, strength theories, such as Schneider's (1985), view learning as the strengthening of stimulus-response associations (also see J. Cohen et al., 1990), so evidence of stimulus-response associations may be interpreted as evidence for strength theories. On the other hand, the instance theory also assumes that subjects learn stimulus-response associations (Logan, 1988, 1990). Subjects learn what they attend to, and they certainly attend to the response they make and the stimulus that led them to make it. Thus, the evidence of stimulus-response associations does not uniquely support strength theory.

The evidence for stimulus-response associations is clearer than the evidence for stimulus-stimulus associations. Stimulus-response associations are defined operationally as a difference between two conditions randomly intermixed in the same experimental session. Subjects could not anticipate the condition for the next trial, so they could not adopt different

Table 3

Examples of Stimuli, Mean Reaction Times (RTs; in Milliseconds), and Percentage Correct From Consistent-Pairing Conditions of Experiment 1 (Single-Session Divided Attention) in the Last Training Block and in the Transfer Block, Partitioned Into Compatible and Incompatible Displays

Training		Transfer	
Steel (Y)	629	Steel (Y)	703
Canada (Y)	93	Carrot (Y)	94
Iron (Y)	629	Iron (Y)	738
Carrot (Y)	93	Beets (N)	91
Chair (N)	645	Chair (N)	693
Beets (N)	98	France (N)	98
Desk (N)	645	Desk (N)	797
France (N)	98	Canada (Y)	94

Note. Y and N in parentheses refer to the response (yes or no) made to each word during training. Steel-Carrot and Chair-France represent compatible displays; Iron-Beets and Desk-Canada represent incompatible displays.

strategies in advance for the different conditions. By contrast, stimulus-stimulus associations are defined operationally as a difference between two conditions from different sessions. It is possible that subjects used a different strategy on the 16th block before pairings changed than they did on the 17th block after pairings changed. The difference in reaction time may reflect different strategies rather than stimulus-stimulus associations. The slower reaction times on the transfer block could reflect a speed-accuracy tradeoff, but the error data presented in Table 3 do not support that interpretation. Subjects were more accurate on Block 16 than on Block 17.

We would like to interpret these results as evidence about what is learned during automatization. As we will see later, our preference is supported by the power function speedup apparent in both groups (see Tables 12 and 13) and a reduction in load effects and dual-task interference with practice (see Figures 8 and 9). Although these results satisfy three major criteria for automaticity, some researchers may be skeptical because of the low levels of practice involved. One session—16 stimulus presentations—is not much practice by common standards in the automaticity literature (e.g., Strayer & Kramer, 1990; Venturino, 1991). We believe that such skepticism would be misguided. In theory, large amounts of practice are not necessary to produce automaticity (Logan, 1988), and empirical tests of the necessity of extended practice have shown that the same effects can be produced at low levels of practice, provided that subjects have learned the requisite associations (Logan & Klapp, 1991; see also Lassaline & Logan, 1993). Nevertheless, we decided to replicate the experiment with a larger amount of practice to satisfy our own (unprincipled) uneasiness about the small amount of practice.

Experiment 2: Multisession Divided Attention

Experiment 2 was a replication of Experiment 1 but with more practice. Subjects completed four sessions of training and one session of transfer. Only consistent-pairing subjects were tested. In Experiment 2, we tested the same hypotheses as in Experiment 1, focusing on the effects of changing pairing

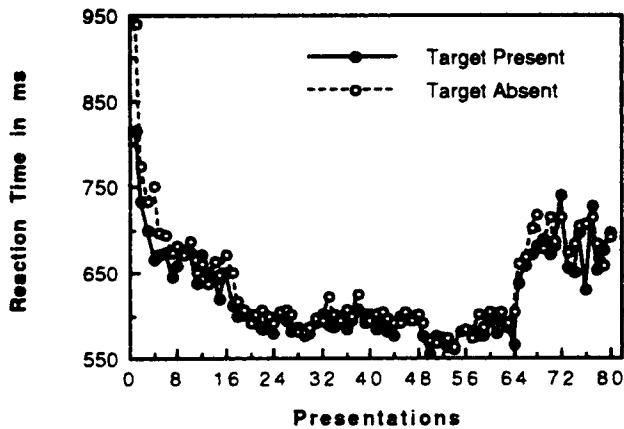


Figure 2. Mean reaction time as a function of the number of presentations for target present (filled circles) and target absent (open circles) responses in Experiment 2 (divided attention).

at transfer. The attention hypothesis predicts sensitivity to co-occurrence because both words are attended.

Method

The apparatus and stimuli were the same as in Experiment 1. The procedure was the same with the following exceptions: During training, there were 16 blocks of 32 consistent-pairing trials for each session (we dropped the 17th transfer block). Subjects saw the same pairings for four sessions or 64 presentations. The transfer session was like a varied-pairing session in Experiment 1. Pairings were changed at the beginning of the session and then changed again at the end of each 32nd trial block (i.e., after each pair had been presented once). Each subject completed his or her sessions in 5 consecutive days. No recall test was administered. There were 8 subjects, who were recruited from the general population at the University of Illinois and paid for their services.

Results

Training. Mean reaction times for target-present and target-absent trials are plotted as a function of number of presentations in Figure 2. There were 16 presentations per session. Reaction times decreased rapidly in the first 16 presentations, as they did in Experiment 1. There were some gains after the first session, but they were smaller in contrast.

These conclusions were confirmed in a 64 (presentation) \times 2 (target present vs. target absent) ANOVA on the mean reaction times. The main effect of presentations was significant, $F(63, 441) = 16.69, p < .01, MS_e = 2,638.42$. The main effect of target presence approached significance, $F(1, 7) = 5.00, p < .07, MS_e = 9,834.27$. The interaction between target presence and presentations was significant, $F(63, 441) = 1.82, p < .05, MS_e = 1,001.37$.

The accuracy data are presented in Table 4. Accuracy improved over presentations, consistent with the reaction time data. A 64 (presentations) \times 2 (target present vs. target absent) ANOVA on the accuracy scores revealed only a significant main effect of presentations, $F(63, 441) = 2.25, p < .05, MS_e = 16.26$.

Transfer. Mean reaction times from the transfer session are also plotted in Figure 2 (see Presentations 65–80). As in Experiment 1, reaction time increased from the last training block to the first transfer block (66 ms in this experiment), suggesting the same sensitivity to co-occurrence. Reaction times remained elevated throughout the transfer block. If anything, they increased over the first half of the session. These data indicate that the cost of changing pairing was not overcome quickly.

These conclusions were confirmed in a 2 (session: training vs. transfer) \times 16 (presentation within session) \times 2 (target present vs. target absent) ANOVA on the mean reaction times. The main effect of session was the only significant effect, $F(1, 7) = 13.58, p < .01, MS_e = 100,697.76$. The main effect of target presence approached significance, $F(1, 7) = 3.88, p < .10, MS_e = 5,291.22$.

The accuracy data are presented in Table 4 as well. The accuracy data were generally consistent with the reaction time data. A 2 (session: training vs. transfer) \times 16 (presentation within session) \times 2 (target present vs. target absent) ANOVA on the accuracy scores revealed only a significant main effect of session, $F(1, 7) = 16.92, p < .01, MS_e = 22.14$.

Discussion

The results replicated the results of Experiment 1 with more extensive practice. Reaction time decreased with practice with consistent pairing and then increased substantially when pairing changed in Session 5. On average, the difference was 103 ms, very close to the 99-ms cost observed after one session in Experiment 1. These results confirm the attention hypothesis.

We analyzed the transfer data for evidence of stimulus-stimulus and stimulus-response associations, following our analysis in Experiment 1. Response-compatible pairs were separated from response-incompatible pairs, and mean reaction times and accuracy scores were calculated for each kind of pair over the entire transfer session. The means across subjects are presented in Table 5.

Stimulus-stimulus associations were evidenced by the difference between response-compatible pairs and performance on the last training session. On average, reaction time was 654 ms for response-compatible trials versus 580 ms for Session 4. This suggests subjects did learn stimulus-stimulus associations. The effect was not compromised by a speed-accuracy tradeoff (see Table 5). Planned comparisons showed that the reaction time

Table 4
Percentage Correct for Target-Present and Target-Absent Responses From Experiment 2 (Multisession Divided Attention) as a Function of Practice Session

Practice session	Target	
	Present	Absent
1	95	97
2	96	98
3	96	98
4	96	98
5	95	97

difference was significant for target-present displays, $F(1, 14) = 69.09$, $p < .01$, $MS_e = 338.81$, and for target-absent displays, $F(1, 14) = 57.23$, $p < .01$, $MS_e = 338.81$.

Stimulus-response associations were evidenced by the difference between response-compatible and response-incompatible pairs at transfer. On average, reaction time was 654 ms for response-compatible trials and 715 ms for response-incompatible pairs, suggesting that subjects learned stimulus-response associations as well as stimulus-stimulus associations. Planned comparisons showed that the difference was significant for target-present displays, $F(1, 14) = 24.71$, $p < .01$, $MS_e = 338.81$, and for target-absent displays, $F(1, 14) = 69.32$, $p < .01$, $MS_e = 338.81$.

These results replicate Experiment 1. Even the magnitudes of the effects are very similar. The evidence of stimulus-stimulus associations provides further evidence for the instance theory, which says that automatic performance is based on retrieval of representations of co-occurring properties.

Experiment 3: Single-Session Focused Attention

The divided-attention experiments showed that subjects were sensitive to co-occurrence when they attended to both words, but attention may have nothing to do with their sensitivity. Subjects may be sensitive to co-occurrence under all conditions of attention, as if an instance were a snapshot that preserved all the details available to the senses. The focused-attention experiments were designed to test the necessity and sufficiency of attention. Attention was directed to targets by cues, so subjects did not have to attend to the distractors. The attention hypothesis predicts that subjects will not be sensitive to co-occurrence under these conditions.

Experiment 3 was almost an exact replication of Experiment 1. The main difference between this experiment and Experiment 1 is that the position of the target was cued by coloring it green. Each display contained one green word and one white word. The green word was either a target or a nontarget. The white word was a member of one of the two distractor categories. Subjects were told to decide whether the green word was a member of the target category. They were told that if the green word was not a target, the other (white) word would not be a target either, so they could respond correctly after examining just the green word.

We cannot discriminate the strong version of the attention hypothesis from the weak one in this experiment. The strong version says that attention is necessary for learning, so there should be no learning of unattended material. The weak version says attention is sufficient but not necessary, so there may be some learning of unattended material. The problem is that focused-attention procedures do not guarantee that subjects will not attend to the "unattended" material (see Hollender, 1986, and commentary). Focused-attention procedures guarantee attention to targets, but they do not prevent attention to distractors. Consequently, evidence of sensitivity to co-occurrence could be interpreted under either version of the attention hypothesis. The weak version would interpret it as learning without attention, whereas the strong version would interpret it as evidence that subjects paid some attention to the distractors. Thus, the key prediction of the attention

Table 5

Examples of Stimuli, Mean Reaction Times (RTs; in Milliseconds), and Percentage Correct From Consistent-Pairing Conditions of Experiment 2 (Multisession Divided Attention) in the Last Training Session and in the Transfer Session, Partitioned Into Compatible and Incompatible Displays

Training		Transfer	
Steel (Y)	576	Steel (Y)	653
Canada (Y)	96	Carrot (Y)	97
Iron (Y)	576	Iron (Y)	699
Carrot (Y)	96	Beets (N)	91
Chair (N)	584	Chair (N)	654
Beets (N)	98	France (N)	98
Desk (N)	584	Desk (N)	731
France (N)	98	Canada (Y)	95

Note. Y and N in parentheses refer to the response (yes or no) made to each word during training. Steel-Carrot and Chair-France represent compatible displays; Iron-Beets and Desk-Canada represent incompatible displays.

hypothesis is not that there will be no learning of distractors under focused-attention conditions, but rather that there will be less learning of distractors under focused-attention conditions than under divided-attention conditions. The effects here should be smaller than those in Experiment 1.

Method

Subjects. Two groups of 32 subjects were recruited from the introductory psychology subject pool at the University of Illinois. They received course credit for participating in the experiment, which took about 1 hour.

Apparatus and stimuli. The apparatus and stimuli were the same as in Experiment 1. The only difference was that the target or the nontarget word in each display was colored green (IBM 10).

Procedure. The procedure was the same as in Experiment 1. The only difference was that subjects were told that the target word would be colored green and that if the green word was not a target, then no target would be present in the display.

Results

Training. Mean reaction time and percentage correct were calculated for target-present and target-absent displays in each practice block for each subject in the consistent- and varied-pairing groups. The reaction time data are plotted as a function of practice in Figure 3.

Consistent-pairing subjects started out a little slower than varied-pairing subjects; mean reaction time on the first presentation was 735 ms for consistent pairing and 718 ms for varied pairing. Both groups improved with practice. Reaction time decreased sharply from the first block to the second and gradually after that. The consistent-pairing group gained more from practice than the varied-pairing group. On the 16th presentation, the consistent-pairing group was faster than the varied-pairing group, averaging 643 ms compared with 668 ms. These results suggest a sensitivity to co-occurrence under focused-attention conditions. However, the effect was much smaller than the one observed in Experiment 1 under divided-attention conditions (cf. Figures 1 and 3). Part of the effect

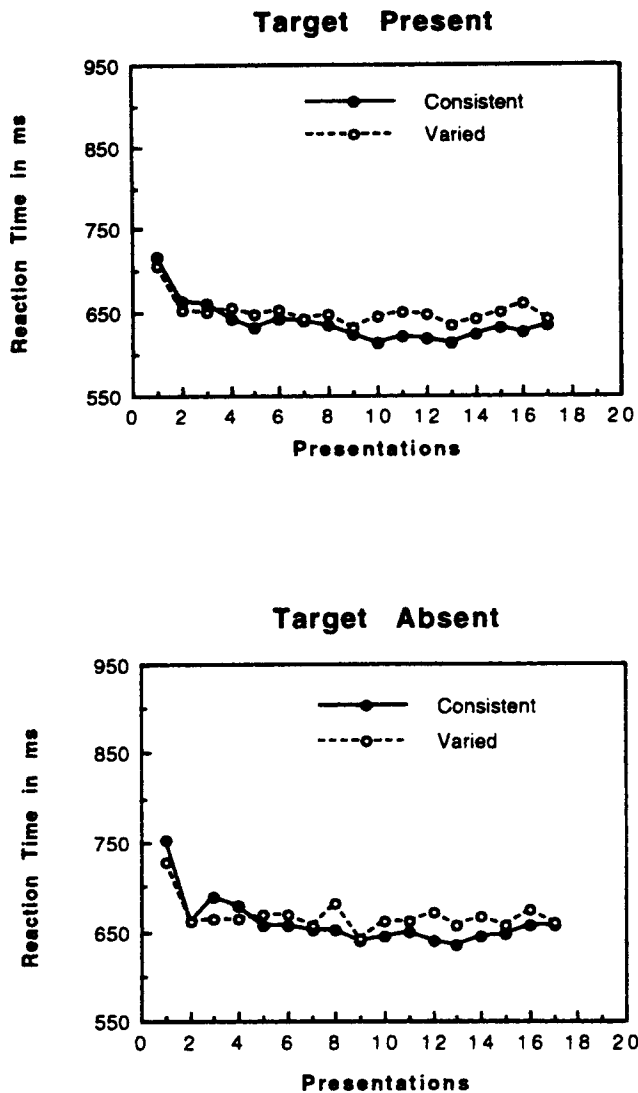


Figure 3. Mean reaction time as a function of the number of presentations for target present (top of figure) and target absent (bottom of figure) responses in Experiment 3 (focused attention). Solid lines = consistent pairing; broken lines = varied pairing.

may reflect a speed-accuracy tradeoff. Consistent- and varied-pairing subjects were equally accurate in the first block (94.0% vs. 93.5%, respectively), but consistent-pairing subjects were less accurate than varied-pairing subjects in the last half of training. On Block 16, consistent-pairing subjects averaged 95.4% correct, whereas varied-pairing subjects averaged 98% correct (see Table 6).

These conclusions were supported by a 16 (presentations) \times 2 (target present vs. target absent) \times 2 (group: consistent pairing vs. varied pairing) ANOVA on the mean reaction times. The ANOVA revealed a significant main effect of presentations, $F(15, 930) = 23.87, p < .01, MS_e = 2,421.36$, and a significant main effect of target presence, $F(1, 62) = 35.09, p < .01, MS_e = 5,003.22$. The main effect of consistency of pairing was not significant, $F(1, 62) < 1.0, MS_e = 98,476.38$, but consistency interacted significantly with presentations,

$F(15, 930) = 2.95, p < .01, MS_e = 2,421.36$. A planned comparison showed that the difference on the 16th presentation was significantly larger than the one on the 1st presentation, $F(1, 930) = 12.06, p < .01, MS_e = 2,421.36$.

The accuracy data are presented in Table 6. Accuracy was generally high, averaging 96.8%. For the most part, it correlated negatively with reaction time, corroborating the reaction time results. A 16 (presentations) \times 2 (target present vs. target absent) \times 2 (group: consistent pairing vs. varied pairing) ANOVA performed on the accuracy data revealed a significant main effect of presentations, $F(15, 930) = 5.28, p < .01, MS_e = 17.39$, and a significant main effect of target presence, $F(1, 62) = 8.41, p < .01, MS_e = 59.91$. The interaction between consistency and presentations was significant, $F(15, 930) = 1.92, p < .05, MS_e = 17.39$, indicating that the difference between consistent- and varied-pairing subjects increased over practice. Consistent-pairing subjects got progressively worse than varied-pairing subjects, suggesting that the opposite effect in the reaction time data was due, in part, to a speed-accuracy tradeoff.

Transfer. Mean reaction times from the transfer block (Block 17) also appear in Figure 3. Reaction time was not affected much by changing pairing, slowing only 4 ms in consistent-pairing subjects and speeding up by 17 ms in varied-pairing subjects. These effects are small compared with the effects in divided-attention conditions (99 ms in Experiment 1 and 103 ms in Experiment 2). Subjects did not show convincing evidence of sensitivity to co-occurrence.

The only significant effect in a 2 (training block vs. transfer block) \times 2 (target present vs. target absent) \times 2 (group: consistent pairing vs. varied pairing) ANOVA on the mean reaction times was the main effect of target presence, $F(1, 62) = 13.46, p < .01, MS_e = 1,720.39$.

The accuracy data from the transfer block are presented in Table 6. Accuracy decreased slightly from Block 16 to Block 17 for both groups. Consistent-pairing subjects went from 95.4% to 94.6% correct; varied-pairing subjects went from 98.1% to

Table 6

Percentage Correct in Each Condition of Experiment 3 (Single-Session Focused Attention) as a Function of Number of Presentations

No. of presentations	Consistent pairing		Varied pairing	
	Present	Absent	Present	Absent
1	94	94	94	93
2	95	98	97	98
3	96	97	97	98
4	96	98	97	97
5	96	97	97	98
6	97	98	96	98
7	96	98	95	97
8	96	97	97	98
9	96	97	99	97
10	95	98	97	97
11	96	98	98	98
12	97	98	97	97
13	96	96	98	98
14	95	96	96	98
15	97	98	96	97
16	94	97	98	98
17	94	96	97	96

96.3% correct. If anything, the effect was larger for varied-pairing subjects. These data provide no evidence of sensitivity to co-occurrence.

These conclusions were supported in a 2 (training block vs. transfer block) \times 2 (target present vs. target absent) \times 2 (group: consistent pairing vs. varied pairing) ANOVA on the accuracy scores. The main effect of transfer was significant, $F(1, 62) = 4.96, p < .05, MS_e = 22.50$, as was the main effect of consistency, $F(1, 62) = 6.53, p < .05, MS_e = 47.91$, and the interaction between target presence and consistency, $F(1, 62) = 4.23, p < .05, MS_e = 25.78$. The critical interaction between transfer and consistency was not significant, $F(1, 62) < 1.0$.

Recall. The number of words recalled in target, nontarget, and distractor categories was calculated for each subject, and the numbers were converted to percentages. The two distractor categories were averaged together in this calculation. Subjects in the consistent-pairing group recalled 57.1% of their targets, 41.1% of their nontargets, and 24.9% of their distractors. Subjects in the varied-pairing group recalled 58.4% of their targets, 43.4% of their nontargets, and 24.4% of their distractors. The results for targets and nontargets, which received attention, are similar to the results from the divided-attention experiment (Experiment 1), in which subjects recalled 60.0% of their targets and 35.8% of their nontargets. The results for distractors are much worse than the divided-attention experiment, in which 36.1% of distractors were recalled. Distractors did not receive as much attention in focused attention as in divided attention, so they were not recalled as well.

These conclusions were supported by a 3 (target vs. nontarget vs. distractor) \times 2 (group: consistent pairing vs. varied pairing) ANOVA on the number of words recalled. The only significant effect was the main effect of target, $F(2, 124) = 94.32, p < .01, MS_e = 5.02$. Orthogonal comparisons revealed a significant difference between targets and the average of nontargets and distractors, $F(1, 124) = 128.28, p < .01, MS_e = 5.02$, which suggests more attention was paid to targets than to nontargets and distractors, and a significant difference between nontargets and distractors, $F(1, 124) = 50.33, p < .01, MS_e = 5.02$, which suggests more attention was paid to nontargets than to distractors.

We compared the present results with those of Experiment 1 in a 3 (target vs. nontarget vs. distractor) \times 2 (group: Experiment 1 vs. Experiment 3) \times 2 (group: consistent pairing vs. varied pairing) ANOVA. The main effect of target was significant, $F(2, 248) = 155.66, p < .01, MS_e = 4.34$, as was the interaction between target and experiment, $F(2, 248) = 20.93, p < .01, MS_e = 4.34$. Planned comparisons showed that the interaction was mostly due to distractors. One contrast showed that the difference between distractor recall and the mean of target and nontarget recall was different in the two experiments, $F(1, 248) = 34.07, p < .01, MS_e = 4.34$. Separate comparisons between each target condition yielded no significant difference in target recall, $F(1, 248) < 1$, significantly better recall of nontargets in Experiment 3, $F(1, 248) = 10.68, p < .01, MS_e = 4.34$, and significantly worse recall of distractors in Experiment 3, $F(1, 248) = 34.29, p < .01, MS_e = 4.34$. The first contrast and the last one are (redundantly)

critical: They indicate that subjects were less sensitive to distractors in focused attention (Experiment 3) than in divided attention (Experiment 1).

Discussion

The results suggest that subjects had very little sensitivity to the co-occurrence of the pairs of words under focused-attention conditions. Consistent-pairing subjects improved more than varied-pairing subjects did, but not much more, and the small improvement in reaction time was compromised by a small reduction in accuracy, suggestive of a speed-accuracy tradeoff. Consistent-pairing subjects were barely affected by the change in pairing at transfer, both absolutely and in comparison to varied-pairing subjects. These results disconfirm the snapshot hypothesis, which says that sensitivity to co-occurrence is independent of attention. They confirm the attention hypothesis, which says that subjects should not be sensitive to co-occurrence under focused-attention conditions, in which attention is directed toward only one of the words in each pair.

The results are remarkable in contrast with Experiment 1. Focused attention produced faster responses, poorer recall of distractors, a smaller benefit from consistent pairing during training, and a much smaller cost of changing pairing at transfer. Faster responses and poorer recall of distractors suggest that the focused-attention manipulation was successful. The smaller benefit from consistent pairing and the smaller cost of changing pairing indicate that focused attention reduced sensitivity to co-occurrence, relative to divided attention. Manipulating attention made a difference, just as the attention hypothesis predicted.

The recall data are important because they allow us to resolve an ambiguity about whether the attention effects occurred at encoding or retrieval. The attention hypothesis assumes that the effects occur at encoding: Attention determines what is learned. However, the effects could occur at retrieval time. Subjects could learn co-occurrences whether attention is focused or divided but retrieve them only if attention is divided. It is difficult to rule out this interpretation in the reaction time data. The attention conditions were the same in training and transfer, so it is difficult to tell whether attention affected encoding or retrieval.

The recall data are important because the attention conditions during recall were different from what they were in training, and the attention conditions during recall were the same for subjects trained in focused and divided attention. Thus, differences in recall can be attributed to encoding rather than retrieval. The data were clear: Fewer distractors were recalled in focused attention than in divided attention. This suggests that the attention effects occurred at encoding, resolving the ambiguity in the reaction time data.

We would like to interpret these results as evidence about what is learned in automatization, but we recognize that some readers may be skeptical because of the low levels of practice (one session, 544 trials, and 17 presentations of each stimulus), so we replicated the experiment with a larger amount of practice.

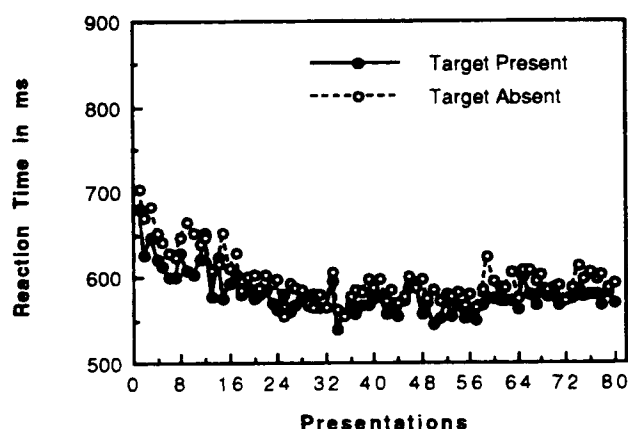


Figure 4. Mean reaction time as a function of the number of presentations for target present (filled circles) and target absent (open circles) responses in Experiment 4 (focused attention).

Experiment 4: Multisession Focused Attention

Method

Experiment 4 was a replication of Experiment 2, except that the target was cued by coloring it green. The attention hypothesis predicts no sensitivity to co-occurrence under focused-attention conditions, even with extended practice. There were 8 subjects recruited from the general population at the University of Illinois, who were paid for their services.

Results

Training. Mean reaction times for target-present and target-absent trials are plotted as a function of number of presentations in Figure 4. There were 16 presentations per session. Reaction times decreased gradually over the four training sessions. The reduction was greater in the first session than in the subsequent ones, but the learning curve was not as sharply inflected as it was in the single-session experiment (Experiment 2).

These conclusions were confirmed in a 64 (presentations) \times 2 (target present vs. target absent) ANOVA on the mean reaction times. The main effect of presentations was significant, $F(63, 441) = 4.30, p < .01, MS_e = 3,138.15$, as was the main effect of target presence, $F(1, 7) = 6.12, p < .05, MS_e = 16,001.04$. The interaction was not significant, $F(63, 441) = 1.05, MS_e = 1,019.54$.

The accuracy data are presented in Table 7. In general, they were consistent with the reaction time data. A 64 (presentations) \times 2 (target present vs. target absent) ANOVA revealed no significant effects.

Transfer. Mean reaction times from the transfer session are also plotted in Figure 4 (see Presentations 65–80). Reaction time increased 35 ms from the last training block to the first transfer block. That increase was not large compared with previous increases from the last block of one session to the first block of the next. The increases were 15, 36, and –8 ms from Sessions 1 to 2, 2 to 3, and 3 to 4, respectively. The increase was also not large compared with the 66-ms increase from the last

training block to the first transfer block in multisession divided attention (Experiment 2). Averaged over presentations within sessions, the increase in reaction time from the last training session to the transfer session was negligible (13 ms). Reaction times did not change much throughout the transfer session. These results suggest that subjects were insensitive to co-occurrence even after 64 presentations.

These conclusions were confirmed in a 2 (session: training vs. transfer) \times 16 (presentations within sessions) \times 2 (target present vs. target absent) ANOVA on the mean reaction times. The only significant effect was target presence, $F(1, 7) = 11.00, p < .05, MS_e = 4,524.13$.

The accuracy data are presented in Table 7. They were generally consistent with the reaction time data. A 2 (session: training vs. transfer) \times 16 (presentations within sessions) \times 2 (target present vs. target absent) ANOVA on the accuracy data revealed only a significant interaction between session and presentation, $F(15, 105) = 2.02, p < .05, MS_e = 20.06$, reflecting a tendency for accuracy to decrease slightly over presentations in Session 4 and to increase slightly over presentations in Session 5. There was no abrupt transition between training and transfer. Accuracy was 95.5% on the last training block and 94.3% on the first transfer block.

Discussion

The results replicate the results of Experiment 3 with more extensive practice. Reaction time improved with practice with consistent pairing but was not affected much when pairing changed on Session 5. On average, the cost was 13 ms, close to the 3-ms cost observed after one session in the same focused-attention conditions in Experiment 3, but very different from the 99-ms cost observed after one session in divided-attention conditions in Experiment 1 or the 103-ms cost observed after four sessions in divided-attention conditions in Experiment 2. These results confirm the attention hypothesis, which predicts no sensitivity to co-occurrence under focused-attention conditions.

Experiment 5: Focused Attention With Arrowhead Cues

Experiment 5 was a single-session focused-attention experiment, testing the attention hypothesis once more. This time, we used an arrowhead (i.e., $>$) to cue the position of the target or nontarget. The arrowhead cue was important because it could be presented independently of the target display, allow-

Table 7
Percentage Correct for Target Present and Target Absent Responses From Experiment 4 (Multisession Focused Attention) as a Function of Practice Session

Practice session	Target	
	Present	Absent
1	94	94
2	95	96
3	95	96
4	95	96
5	95	95

ing us greater temporal control over attention. If the cue were presented sufficiently early, subjects would be able to use it to focus exclusively on the target word. If the cue were presented sufficiently late, subjects may be able to find the target without it. By varying the delay between the cue and the target display, we could change smoothly from focused-attention conditions to divided-attention conditions. According to the attention hypothesis, subjects' sensitivity to co-occurrence should vary with cue delay, ranging from what we saw in Experiments 3 and 4 when the cue was early to what we saw in Experiments 1 and 2 when the cue was delayed.

In Experiment 5 we used two different delays in a between-subjects design: The cue appeared either 500 ms or 0 ms before the target display. The 500-ms cue should provide ample time to move attention to the position of the target. Thus, the 500-ms cue provides a pure focused-attention condition, in which the attention hypothesis predicts no sensitivity to co-occurrence. There should be no difference between consistent- and varied-pairing subjects during training and no cost from changing pairing at transfer. By contrast, the 0-ms cue should not allow subjects to focus attention on the position of the target before the display appears, as with the color cue in Experiments 3 and 4. Subjects may spend some time dividing attention between the words before they find the cue and locate the target with respect to it (Logan, Withey, & Cowan, 1977), or they may inadvertently select the distractor word before they locate the cue. To the extent that they attend to the distractors, they should show sensitivity to co-occurrence. Consistent-pairing subjects should do better than varied-pairing subjects during training and have more difficulty than they do from the change in pairing at transfer.

Method

Subjects. We recruited four groups of 32 subjects from the introductory psychology subject pool at the University of Illinois. Subjects received course credit for participating in the experiment.

Apparatus and stimuli. These were the same as in Experiments 1 and 2 (i.e., two white words were displayed on each trial). The only difference was that an arrowhead cue (i.e., >) was presented one character space to the left of the initial letter of the target or nontarget word. In the 500-ms delay condition, the cue appeared with the fixation and warning display and remained on when the target display appeared. In the 0-ms delay condition, the cue appeared (and disappeared) with the target display.

Procedure. There were four groups of subjects, defined by the factorial combination of consistent versus varied pairing and 0- versus 500-ms delay. Otherwise, the procedure was the same as the one used in Experiments 1 and 3, including the surprise recall test at the end of the transfer block.

Results

Training. Mean reaction time and percentage correct were calculated for target-present and target-absent displays in each practice block for each subject in each combination of consistency and cue delay conditions. The reaction time data are plotted as a function of practice in Figure 5.

All subjects showed evidence of learning: Reaction time decreased with practice. The learning curves were sharply

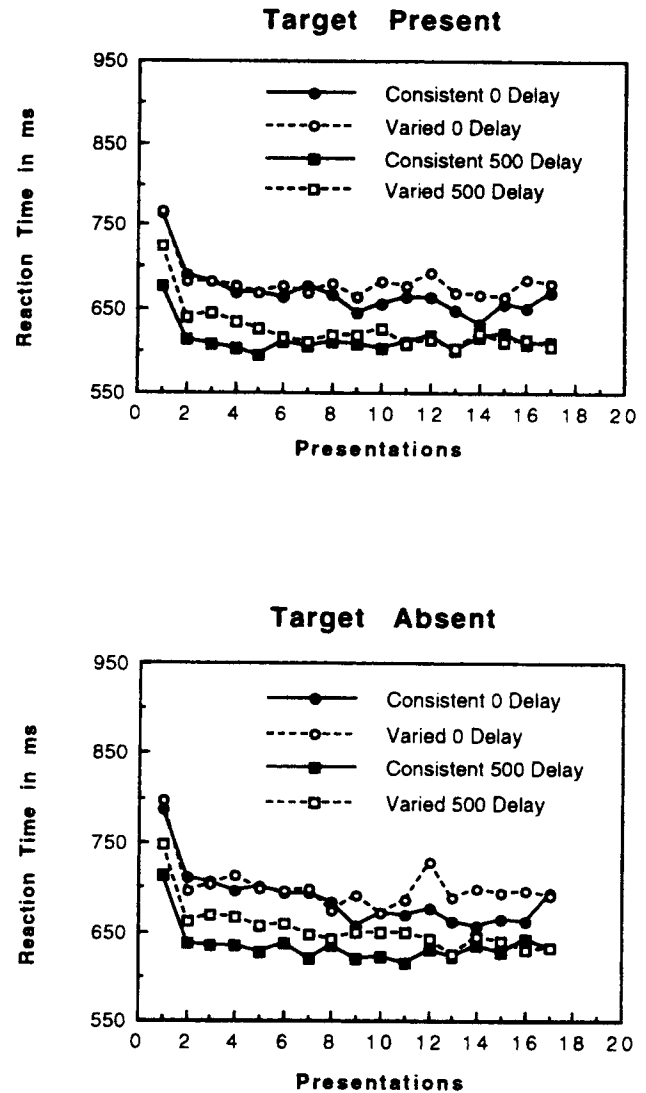


Figure 5. Mean reaction time as a function of the number of presentations for target present (top of figure) and target absent (bottom of figure) responses in Experiment 5 (focused attention with arrow head cues). Solid lines = consistent pairing; broken lines = varied pairing; circles = 0-ms delay; squares = 500-ms delay.

inflected, like those in Experiment 3, indicating that most of the performance improvement came between the first and second blocks. Cue delay had a strong effect on performance. Subjects were 53 ms faster when the cue came 500 ms before the target display than when the cue appeared simultaneously with the target.

Subjects in the 0-ms delay condition showed some sensitivity to co-occurrence. Consistent-pairing subjects started out performing at the same level as varied-pairing subjects (776 and 781 ms, respectively), but after the 8th block, consistent-pairing subjects were faster. On the last training block, they were 34 ms faster than varied-pairing subjects (656 and 690 ms, respectively).

Subjects in the 500-ms delay condition showed a pattern that is difficult to interpret. On the one hand, there was no

Table 8
*Percentage Correct in Each Condition of Experiment 5
 (Single-Session Focused Attention With Arrow Cues) as a
 Function of Number of Presentations*

No. of presentations	0-ms cue delay				500-ms cue delay			
	C		V		C		V	
	P	A	P	A	P	A	P	A
1	91	93	91	94	92	95	92	96
2	95	97	94	96	95	98	96	99
3	94	97	94	96	95	97	97	98
4	94	97	94	96	96	98	96	98
5	95	97	94	96	95	97	95	98
6	95	96	92	96	94	96	95	98
7	94	95	94	94	95	98	94	97
8	93	97	93	92	96	97	96	98
9	94	96	91	92	95	98	97	98
10	93	97	92	93	95	98	96	98
11	93	97	94	94	94	96	96	97
12	93	96	94	93	95	97	96	98
13	93	96	92	95	94	98	96	97
14	91	97	92	95	94	96	96	96
15	92	97	93	94	93	98	95	96
16	94	96	91	94	93	97	95	97
17	91	94	91	92	93	94	96	97

Note. C = consistent pairing; V = varied pairing; P = target present; A = target absent.

difference between consistent- and varied-pairing subjects in the last practice block (mean reaction times were 625 and 621 ms, respectively), suggesting the insensitivity to co-occurrence predicted by the attention hypothesis. On the other hand, the consistent-pairing subjects showed a 41-ms advantage over varied-pairing subjects in the first practice block (mean reaction times were 695 and 736 ms, respectively). This difference is hard to interpret because there were no differences in procedure in the first block. Consistency of pairing was defined in terms of what happened in the second (and subsequent) block, not the first.

These conclusions were confirmed in three ANOVAs on the mean reaction times. The whole data set was subjected to a 2 (group: cue delay) \times 16 (presentations) \times 2 (target present vs. target absent) \times 2 (group: consistent pairing vs. varied pairing) ANOVA. However, the difference in initial performance between consistent- and varied-pairing subjects in the 500-ms delay condition made us suspect the validity of that analysis, so we analyzed each delay condition separately in a 16 (presentations) \times 2 (target present vs. target absent) \times 2 (group: consistent pairing vs. varied pairing) ANOVA and focused our conclusions primarily on those analyses.

The ANOVA on the 0-ms delay group revealed a significant main effect of target presence, $F(1, 62) = 26.40, p < .01, MS_e = 8,709.03$; a significant main effect of presentations, $F(15, 930) = 30.39, p < .01, MS_e = 2,908.16$; and a significant interaction between presentations and consistency of pairing, $F(15, 930) = 2.52, p < .01, MS_e = 2,908.16$. No other effects were significant.

The ANOVA on the 500-ms delay group revealed a significant main effect of target presence, $F(1, 62) = 42.29, p < .01, MS_e = 6,817.03$, and a significant main effect of presentations, $F(15, 930) = 20.23, p < .01, MS_e = 3,424.30$. The interaction

between presentations and consistency of pairing approached significance, $F(15, 930) = 1.68, p < .06, MS_e = 3,424.20$. No other effects were significant.

The ANOVA including both delays revealed significant effects of delay, $F(1, 124) = 25.24, p < .01, MS_e = 118,489.25$; presentations, $F(15, 1860) = 48.77, p < .01, MS_e = 3,166.23$; and target presence, $F(1, 124) = 66.54, p < .01, MS_e = 7,763.03$. The interaction between delay, presentations, and consistency of pairing was significant, $F(15, 1860) = 3.37, p < .01, MS_e = 3,166.23$. This interaction is consistent with the idea that the difference between consistent and varied pairing became larger over practice in the 0-ms delay group than in the 500-ms delay group. However, it may also reflect the fact that the difference between consistent and varied pairing converged with practice in the 500-ms delay group but diverged with practice in the 0-ms delay group.

The accuracy data are presented in Table 8. The accuracy data were consistent with the reaction times. They were analyzed in a 2 (group: cue delay) \times 16 (presentations) \times 2 (target present vs. target absent) \times 2 (group: consistent pairing vs. varied pairing) ANOVA. The only significant effects were the main effect of cue delay, $F(1, 124) = 6.92, p < .01, MS_e = 509.10$; the main effect of presentations, $F(15, 1860) = 4.13, p < .01, MS_e = 37.31$; and the main effect of target presence, $F(1, 124) = 21.65, p < .01, MS_e = 256.27$.

Transfer. Mean reaction times for the transfer block also appear in Figure 5. None of the groups was affected much by the change in pairing at transfer. However, delay had a small effect in the expected direction: Consistent-pairing subjects showed a 24-ms cost of changing pairing in the 0-ms delay condition but a -3-ms cost in the 500-ms delay condition. Subjects in the 0-ms delay condition showed some sensitivity to co-occurrence, whereas subjects in the 500-ms delay condition did not.

These conclusions were confirmed in three ANOVAs on the mean reaction times. Each delay condition was analyzed separately in a 2 (training vs. transfer) \times 2 (target present vs. target absent) \times 2 (group: consistent pairing vs. varied pairing) ANOVA, and then the two conditions were compared in a 2 (group: cue delay) \times 2 (training vs. transfer) \times 2 (target present vs. target absent) \times 2 (group: consistent pairing vs. varied pairing) ANOVA. Again, we have more confidence in the separate analyses than in the analysis that included both delays because of the unexplained initial differences between consistent and varied pairing in the 500-ms delay condition.

In the ANOVA on the 0-ms delay group, only the main effect of target presence was significant, $F(1, 62) = 4.63, p < .05, MS_e = 2,882.77$. The crucial interaction between transfer and consistency approached significance, albeit from afar, $F(1, 62) = 3.03, p < .09, MS_e = 3,929.33$. A planned comparison showed that the 24-ms transfer cost in the consistent-pairing condition was significant, $F(1, 62) = 4.26, p < .05, MS_e = 3,929.33$.

In the ANOVA on the 500-ms delay group, only the main effect of target presence was significant, $F(1, 62) = 18.88, p < .01, MS_e = 1,887.78$. The interaction between transfer and consistency was not significant, $F(1, 62) < 1.0, MS_e = 2,352.36$, nor was a planned comparison testing the -3-ms transfer cost in the consistent-pairing group, $F(1, 62) < 1.0$.

The ANOVA on both delay groups revealed only a significant main effect of delay, $F(1, 124) = 18.17, p < .01, MS_e = 22,835.11$, and a significant main effect of target presence, $F(1, 124) = 19.41, p < .01, MS_e = 2,385.27$. The interaction between delay, consistency of pairing, and training versus transfer was not significant, $F(1, 124) = 2.02, p < .16, MS_e = 3,140.85$. A planned contrast comparing the 24-ms transfer cost in the consistent-pairing 0-ms delay group and the -3-ms cost in the varied-pairing and the 500-ms delay group approached significance, $F(1, 124) = 3.52, p < .10, MS_e = 3,140.85$.

The accuracy data from the transfer block also appear in Table 8. Again, the accuracy data were consistent with the reaction times. They were analyzed in a 2 (group: cue delay) \times 2 (training vs. transfer) \times 2 (target present vs. target absent) \times 2 (group: consistent pairing vs. varied pairing) ANOVA. The only significant effects were the main effect of cue delay, $F(1, 124) = 4.33, p < .05, MS_e = 154.94$, and the main effect of target presence, $F(1, 124) = 8.21, p < .01, MS_e = 69.87$. The main effect of training versus transfer approached significance, $F(15, 1860) = 3.49, p < .07, MS_e = 45.15$.

Recall. The number of words recalled in target, nontarget, and distractor categories was calculated for each subject, and the numbers were converted to percentages. The means across subjects in each cue delay and consistency condition are presented in Table 9. On average, subjects recalled 59.1% of the targets, 41.2% of the nontargets, and 23.4% of the distractors. The pattern was more like the one observed in the previous single-session focused-attention experiment (Experiment 3: 54.8%, 42.3%, and 24.7%, respectively) than the one in the single-session divided-attention experiment (Experiment 1: 61.0%, 35.8%, and 36.1%, respectively): Nontargets, which received attention, were recalled better than distractors, which received little or no attention.

Note that the recall data provide converging evidence that the effects of cuing attention occurred at encoding rather than at retrieval. Subjects were less likely to recall distractors in focused attention because they were less likely to encode them in training.

Consistent-pairing subjects recalled distractors better than varied-pairing subjects (26.3% vs. 20.8%, respectively) even though their recall of targets and nontargets was slightly worse (averaging 47.6% vs. 52.7%, respectively). This suggests that consistent-pairing subjects were more sensitive to the distractors than varied-pairing subjects.

Overall, recall was slightly better in the 0-ms delay condition than in the 500-ms delay condition (42.9% vs. 39.7%, respectively). There was a tendency for consistent-pairing subjects to recall more distractors when cue delay was 0 ms than when it was 500 ms (22.8% vs. 19.9%, respectively), which suggests that more attention was paid to distractors when cue delay was 0 ms. However, consistent-pairing subjects also recalled targets and nontargets better when cue delay was 0 ms than when it was 500 ms (49.7% vs. 45.5%, respectively), which does not support the suggestion that subjects paid more attention to distractors when cue delay was 0 ms.

These conclusions were supported by a 2 (group: cue delay) \times 2 (group: consistent pairing vs. varied pairing) \times 3 (target vs. nontarget vs. distractor) ANOVA on the number of words

Table 9

Percentage Correct in Recall in Each Condition of Experiment 5 (Single-Session Focused Attention With Arrow Cues) as a Function of Category

Category	0-ms cue delay		500-ms cue delay	
	C	V	C	V
Target	59.7	63.7	55.9	57.2
Nontarget	39.7	47.1	35.2	43.0
Distractor	27.0	20.1	25.6	21.1

Note. C = consistent pairing; V = varied pairing.

recalled. The only significant effects were the main effect of target, $F(2, 248) = 298.14, p < .01, MS_e = 3.50$, and the interaction between target and consistency, $F(2, 248) = 10.53, p < .01, MS_e = 3.50$. The main effect of cue delay was not significant, $F(1, 124) = 2.39, p < .20, MS_e = 10.69$, and cue delay did not interact with any of the other factors. Planned comparisons showed that targets were recalled better than the average of nontargets and distractors, $F(1, 248) = 448.43, p < .01, MS_e = 3.50$, and nontargets were recalled better than distractors, $F(1, 248) = 147.85, p < .01, MS_e = 3.50$. Another planned comparison showed that distractors were recalled relatively better (and targets and nontargets were recalled relatively worse) in the consistent-pairing conditions than in the varied-pairing conditions, $F(1, 248) = 18.17, p < .01, MS_e = 3.50$.

Discussion

The results suggest that attention modulates sensitivity to co-occurrence. There was virtually no sensitivity to co-occurrence in the 500-ms delay condition, when attention was sharply focused on the target. Consistent pairing showed no advantage over varied pairing at the end of training and no cost from changing pairing at transfer. The training results were compromised somewhat by unexplained initial differences between consistent- and varied-pairing groups, but the transfer results are clear. In the 0-ms delay condition, in which subjects may have attended to distractors before focusing on the target, there was some sensitivity to co-occurrence. Consistent pairing showed an advantage over varied pairing at the end of practice and a cost from changing pairing at transfer.

The effects in the 0-ms delay condition were large relative to the single-session divided-attention effects (Experiment 1). The advantage at the end of training was 69% of the (49-ms) advantage at the end of training in the divided-attention experiment. The cost at transfer was 24% of the (99-ms) cost in the divided-attention experiment. If we delayed the cue even more (e.g., presenting it 500 ms after the display), these effects should grow stronger, reaching asymptote at the level of divided-attention effects.

These results provide strong support for the attention hypothesis. The results with the 500-ms delay show no evidence of sensitivity to co-occurrence, and the results with the 0-ms delay show some evidence of sensitivity to co-occurrence. The attention hypothesis accounts for both sets of results by arguing that cue delay modulates attention to distractors and therefore, sensitivity to co-occurrence. The argument that cue

delay modulates attention is supported by a lot of data (e.g., Eriksen & Colegate, 1971; Eriksen & Hoffman, 1972; Logan et al., 1977) and by the 53-ms difference between the 0- and 500-ms delay conditions.

Experiment 6: Single-Session Dual Task

In the next two experiments we extended the attention hypothesis to dual-task situations. The attention hypothesis predicts sensitivity to co-occurrence because dual-task conditions require subjects to pay attention to both words. The attention hypothesis predicts that subjects will learn what they attend to in dual-task conditions. If they have sufficient time to attend to all of the information, they should learn about co-occurrences of stimuli. Subjects in Experiments 6 and 7 should show the same sensitivity to co-occurrence as subjects in Experiments 1 and 2.

In Experiment 6 we examined a single session of learning in dual-task conditions. We selected conditions that we thought would produce severe interference but at the same time provide opportunities to learn about co-occurrence. Stimuli were exposed simultaneously and remained on until subjects responded to both tasks. The simultaneous onsets should maximize interference (Duncan, 1980; Pashler, 1989). The long exposure durations should allow subjects to attend successfully to both words, which is a necessary condition for learning co-occurrence. Sensitivity to co-occurrence was assessed as it was in the previous single-session experiments, by training subjects for 16 blocks of consistent or varied pairing and then transferring them to a single block of varied pairing.

In Experiment 6 we used exactly the same displays as in the divided-attention experiments (Experiments 1 and 2). It differed primarily in the response requirements: Subjects had two category search tasks, one assigned to each hand. One of the tasks required subjects to discriminate between targets and nontargets, as in the previous experiments. The other task required subjects to discriminate between the two distractor categories, one of which was treated as a target and the other of which was treated as a nontarget.

Another important difference was in the duration of the target display. We exposed it for 1,000 ms in the previous experiments. In Experiment 6, we exposed the target display until subjects registered both responses. We wanted to be sure that subjects would have enough time to encode both words before the display disappeared. Rather than choosing an arbitrary duration for the display, we let the subjects control the duration.

Finally, we ran more subjects in this experiment than in the previous ones (48 per condition rather than 32). We expected reaction times to be longer and more variable, so we would have less power to detect the same differences. We hoped to regain some of that power by adding subjects.

Method

Subjects. Two groups of 48 introductory psychology students from the University of Illinois served as subjects. They received course credit for their participation.

Apparatus and stimuli. The apparatus and stimuli were the same as in the previous experiments. Subjects responded differently, however.

For one task, they pressed the period key and the slash key with the index and middle fingers, respectively, of their right hands. For the other task, subjects pressed the *z* key and the *x* key with the middle and index fingers, respectively, of their left hands. Target presence versus absence was mapped onto response keys in four ways: NYYN, YNNY, NYNY, and YNYN, in which Y and N referred to target present and target absent, respectively, and the ordering from left to right referred to the *z*, *x*, period, and slash keys, respectively. We assigned 12 subjects to each mapping condition.

The timing of the displays was different from the previous experiments. The target displays were exposed until subjects' second response was registered. Then the screen went blank for a 2,000-ms intertrial interval.

Procedure. The procedure was very similar to the previous experiments. The displays were constructed in the same manner following the same counterbalancing scheme. As in Experiment 1, the four categories were divided into two pairs—a target and a nontarget pair and a pair of distractor categories—and the two pairs were assigned to displays orthogonally. Half of the targets occurred with one distractor category, and half occurred with the other; half of the nontargets occurred with one distractor category, and half occurred with the other. Thus, subjects could not predict which distractor category would be present once they knew whether a target or a nontarget was present. Similarly, they could not predict whether a target or a nontarget would be present once they knew which distractor category was present.

The dual-task procedure was implemented by requiring subjects to discriminate between the two distractor categories as well as discriminating between target and nontarget categories. They reported the different discriminations by pressing different keys with different hands.

Results

Training. Mean reaction time and percentage correct were computed for target-present and target-absent displays in each task in each practice block in each consistency condition. The reaction time data are plotted as a function of practice in Figure 6.

Reaction times were substantially longer in these dual-task conditions than in the divided-attention conditions of Experiment 1; subjects seem to have suffered dual-task interference. Nevertheless, performance improved dramatically for both groups of subjects, dropping 1,357 ms on average from the 1st block to the 16th. Consistent-pairing subjects improved more than varied-pairing subjects. Their performance was almost equivalent in the 1st block—2,544 ms for consistent-pairing subjects versus 2,483 ms for varied-pairing subjects. But consistent-pairing subjects were 149-ms faster than varied-pairing subjects in the 16th block ($M = 1,082$ and $1,231$ ms, respectively). This suggests some sensitivity to co-occurrence, even under dual-task conditions.

Apparently, subjects coped with dual-task conditions by grouping their responses, a strategy that sometimes appears in dual-task experiments (Kahneman, 1973; Pashler & Johnston, 1989). Response grouping involved emitting the responses to the two tasks close together in time, as if the two tasks required a single two-handed response. Reaction times to the two tasks were almost the same (1,380 vs. 1,342 ms for the consistent-pairing group; 1,430 vs. 1,440 ms for the varied-pairing group), suggesting a very short interresponse interval, which is characteristic of grouping (Kahneman, 1973).

We calculated the correlation between the reaction times

for Task 1 and Task 2 on a trial-by-trial basis to assess grouping further. If the responses were grouped, reaction times should be correlated. When one is fast, the other should be fast; when one is slow, the other should be slow. If the responses were not grouped but instead were executed independently, reaction times should not be correlated (Pashler & Johnston, 1989). We calculated correlations within training blocks so that the large reduction in reaction times to both tasks over practice did not contribute to the correlation. We disregarded target presence versus target absence in the two tasks to increase the number of observations going into the correlation (maximum = 32 per block). There were 32 correlations in all, 16 for each group. In the consistent-pairing group, the smallest correlation was .890, observed on the 1st block, the largest was .964, observed on the 6th block, and the average was .937. The correlation for the 16th block was .950. In the varied-pairing group, the smallest correlation was .906, again observed in the 1st block, the largest was .969, observed in the 7th block, and the average was .949. The correlation for the 16th block was .959. These are high correlations, particularly for correlations computed on raw reaction times from individual subjects. Correlations of this magnitude are often seen with means, in which the process of averaging removes much of the error variance (see, e.g., the power function fits to the mean reaction times in Tables 12 and 13). It is remarkable that they are so large within subjects. The tendency to group responses was very strong.

The mean reaction times were analyzed in a 16 (presentations) \times 2 (target present vs. target absent) \times 2 (Task 1 vs. Task 2) \times 2 (group: consistent pairing vs. varied pairing) \times 4 (group: mapping condition) ANOVA. The two most important effects were significant: the main effect of presentations, $F(15, 1320) = 248.67$, $p < .01$, $MS_e = 1,090,510.95$, and the interaction between presentations and consistency of pairing, $F(15, 1320) = 3.49$, $p < .01$, $MS_e = 1,090,510.95$. The main effect of target presence was significant, $F(1, 88) = 3.98$, $p < .05$, $MS_e = 260,954.41$, reflecting faster responses when the target was present. The main effect of mapping condition was significant, $F(3, 88) = 4.50$, $p < .01$, $MS_e = 5,589,217.45$, reflecting faster responses with symmetrical mappings (NYYN and YNNY) than with asymmetrical ones (NYNY and YNYN). Many of the interactions were significant, but they did not compromise the main results. Task 1 versus Task 2 interacted with consistency of pairing, $F(1, 88) = 6.30$, $p < .05$, $MS_e = 133,804.39$. Presentations interacted with target presence, $F(15, 1320) = 5.65$, $p < .01$, $MS_e = 42,270.37$, and mapping, $F(45, 1320) = 2.47$, $p < .01$, $MS_e = 190,510.95$.

There were significant three-way interactions between presentations, consistency, and mapping, $F(45, 1320) = 1.98$, $p < .01$, $MS_e = 190,510.95$; presentations, target presence, and mapping, $F(45, 1320) = 1.59$, $p < .01$, $MS_e = 42,270.37$; presentations, Task 1 versus Task 2, and mapping, $F(15, 1320) = 3.82$, $p < .01$, $MS_e = 14,639.84$; and presentations, target presence, and Task 1 versus Task 2, $F(15, 1320) = 8.22$, $p < .01$, $MS_e = 42,270.37$. Finally, the four-way interaction between presentations, target presence, consistency, and mapping was also significant, $F(45, 1320) = 2.21$, $p < .01$, $MS_e = 42,270.37$. We have no interpretation for these interactions.

The accuracy data are presented in Table 10. We analyzed the accuracy data in a 16 (presentations) \times 2 (target present

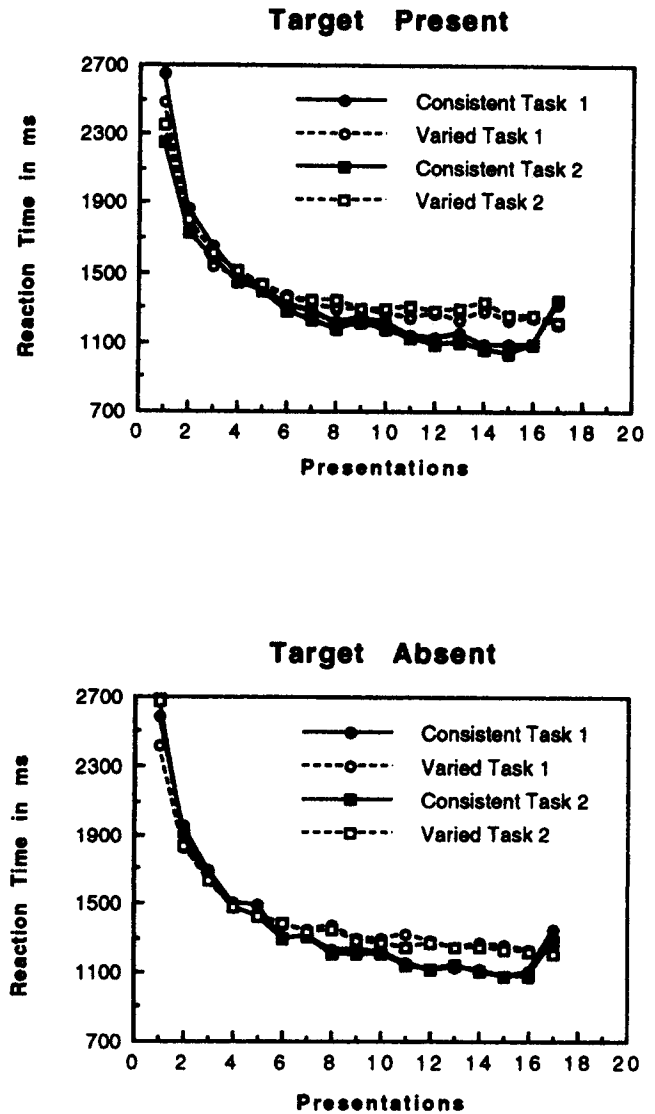


Figure 6. Mean reaction time as a function of the number of presentations for target present (top of figure) and target absent (bottom of figure) responses in Experiment 6 (dual task). Solid lines = consistent pairing; broken lines = varied pairing; circles = Task 1; squares = Task 2.

vs. target absent) \times 2 (Task 1 vs. Task 2) \times 2 (group: consistent pairing vs. varied pairing) \times 4 (group: mapping condition) ANOVA. There was a significant main effect of presentations, $F(15, 1320) = 4.13$, $p < .01$, $MS_e = 149.58$, and target presence versus target absence, $F(1, 88) = 35.86$, $p < .01$, $MS_e = 292.26$. The interaction between task and consistency of pairing was significant, $F(1, 88) = 6.06$, $p < .05$, $MS_e = 323.64$.

Transfer. The mean reaction times for the transfer block are also plotted in Figure 6. Consistent-pairing subjects were strongly affected by the change in pairing at transfer, slowing their reaction times by 241 ms. Varied-pairing subjects were not affected at all. They were 20 ms faster in transfer than in training. These effects may not appear very large in Figure 6, mostly because the reaction time scale was compressed substan-

Table 10
*Percentage Correct in Each Condition of Experiment 6
 (Single-Session Dual Task) as a Function
 of Number of Presentations*

No. of presentations	Task 1				Task 2			
	C		V		C		V	
	P	A	P	A	P	A	P	A
1	91	93	87	95	92	92	89	92
2	95	97	92	96	93	95	93	97
3	96	97	92	96	94	96	94	97
4	96	96	93	97	93	96	95	97
5	95	98	94	97	94	96	94	97
6	96	99	94	96	94	96	95	97
7	96	98	94	95	95	98	95	96
8	96	97	94	98	95	97	95	97
9	96	98	95	98	94	97	96	98
10	96	99	94	96	95	98	94	96
11	97	98	94	96	95	99	94	96
12	95	98	94	97	95	97	94	97
13	96	97	91	94	93	97	95	96
14	96	98	91	94	95	97	95	95
15	96	99	93	96	95	97	95	95
16	96	96	94	96	93	98	95	96
17	94	94	93	93	94	97	93	94

Note. C = consistent pairing; V = varied pairing; P = target present; A = target absent.

tially. There was much more learning in this experiment than in the previous ones, and though a lot was lost at transfer, performance did not revert to anything like initial, unpracticed performance.

These conclusions were confirmed in a 2 (training block vs. transfer block) \times 2 (target present vs. target absent) \times 2 (Task 1 vs. Task 2) \times 2 (group: consistent pairing vs. varied pairing) \times 4 (group: mapping condition) ANOVA on the mean reaction times. The main effect of training versus transfer was significant, $F(1, 88) = 22.08, p < .01, MS_e = 106,272.09$, as was the crucial interaction between transfer and consistency of pairing, $F(1, 88) = 30.84, p < .01, MS_e = 106,272.09$. In addition, the main effect of mapping condition was significant, $F(3, 88) = 3.56, p < .05, MS_e = 724,531.90$, as was the interaction between transfer, target presence, Task 1 versus Task 2, consistency, and mapping, $F(3, 88) = 3.66, p < .05, MS_e = 33,510.05$.

The accuracy data from the transfer block also appear in Table 10. They were analyzed in a 2 (training block vs. transfer block) \times 2 (target present vs. target absent) \times 2 (Task 1 vs. Task 2) \times 2 (group: consistent pairing vs. varied pairing) \times 4 (group: mapping condition) ANOVA. The only significant effects were the main effect of training versus transfer, $F(1, 88) = 10.59, p < .01, MS_e = 33.00$, and the main effect of target presence, $F(1, 88) = 7.85, p < .01, MS_e = 48.39$.

Discussion

The results suggest that subjects were sensitive to co-occurrence in dual-task conditions. Consistent-pairing subjects improved more than varied-pairing subjects during training, and they suffered more than varied-pairing subjects at transfer. These results are consistent with the attention hypothesis. The

sensitivity to co-occurrence was stronger than that observed in single-session divided attention (Experiment 1).

The effects of consistent pairing are interesting because they show that subjects learned associations between the words in the pairs. These results may be related to findings in the sequence learning literature: Subjects have difficulty learning arbitrary associations between responses in a sequence under dual-task conditions (Nissen & Bullemer, 1987), but the difficulty is restricted primarily to nonunique (i.e., inconsistent) associations. Unique associations can be learned as well in dual-task and single-task conditions (A. Cohen, Ivry, & Keele, 1990). In Experiment 6, the (consistent-pairing) associations that subjects learned under dual-task conditions were also unique.

The contrast between consistent pairing and varied pairing may account for some of the difficulty subjects had with learning in dual-task conditions. The contrast suggests that subjects learn arbitrary co-occurrences of events in the two tasks and that performance is better when the co-occurrences are repeated than when they change from trial to trial. Dual-task experiments in the literature probably involved varied pairing rather than consistent pairing. Experimenters often manipulate events in the two tasks independently, so arbitrary co-occurrences are unlikely to be repeated and subjects do not have the opportunity to show that they learned them. Independent manipulation may be good experimental design, but it underestimates what subjects learn in dual-task conditions.

As in the previous experiments, we thought it was important to replicate these results with a greater degree of practice to increase generalizability and to convince people who think extensive practice is necessary to produce automaticity. In Experiment 7 we replicated Experiment 6 with four training sessions and one transfer session.

Experiment 7: Multisession Dual Task

Subjects were trained under consistent-pairing conditions for four sessions and then transferred to varied pairing for the fifth session. No varied-pairing subjects were run. The main question was whether there would be cost when pairing changed on the fifth session and whether that cost would persist throughout the session.

Method

Subjects. In this experiment we used 12 paid volunteer subjects from the University of Illinois. This was different from the previous multisession experiments, which used only 8 subjects. The extra subjects were run to offset a reduction in power we anticipated because dual-task reaction times should be longer and more variable than single-task (i.e., divided and focused attention) reaction times.

Apparatus and stimuli. The apparatus and stimuli were the same as in Experiment 6.

Procedure. The procedure was the same as the consistent-pairing procedure in Experiment 6 for the first four sessions, except that there were only 16 training blocks each session. There was no 17th transfer block. The procedure on the fifth session was the same as the varied-pairing procedure from Experiment 6, except that there were only 16 training blocks.

Results

Training. Mean reaction times for the two tasks are presented as a function of practice in Figure 7. Reaction time decreased over sessions in both tasks. Mean reaction time was 1,241 ms on Session 1, 970 ms on Session 2, 822 ms on Session 3, and 742 ms on Session 4. The regular decrease was interrupted by a kind of "scaloping": Reaction time was faster in the last block of Session *N* than it was in the first block of Session *N* + 1. The differences were 275, 120, and 24 ms for the transitions between Sessions 1 and 2, 2 and 3, and 3 and 4, respectively. Despite the scaloping, reaction times for both tasks were well fit by power functions (see Tables 12 and 13).

As in the single-session experiment, there was evidence that subjects grouped their responses. The interresponse interval (Task 2 reaction time minus Task 1 reaction time) was short, averaging 6, -6, -2, and -11 ms for Sessions 1 to 4, respectively. The reaction times for the two tasks were correlated strongly. As in Experiment 6, we calculated correlations between reaction times for the two tasks in each practice block, disregarding target presence versus target absence. The correlations ranged from .847 (Block 16, Session 4) to .992 (Block 14, Session 1). The average correlations were .960, .937, .917, and .893 for Sessions 1 to 4, respectively. These correlations are high, especially for individual subject data. They indicate a strong tendency to group responses, to respond to the two tasks as if they were one.

The conclusions about the practice effects were confirmed in an ANOVA. Our ANOVA program did not have access to sufficient memory to run the entire 64 (presentations) \times 2 (target presence) \times 2 (Task 1 vs. Task 2) design, so we analyzed each task separately in 64 (presentations) \times 2 (target presence) ANOVAs. This was justified by the high correlations between the raw reaction times in each task and by a 32 (presentations) \times 2 (target presence) \times 2 (task) ANOVA, in which we collapsed adjacent presentations to fit within memory limitations. That ANOVA showed no main effect of tasks and no significant interactions between tasks and the other variables.

The main effect of presentations was significant in Task 1, $F(63, 693) = 17.66, p < .01, MS_e = 82,359.27$, and in Task 2, $F(63, 693) = 17.60, p < .01, MS_e = 92,248.93$. The main effect of target presence and the interaction between target presence and presentations were not significant in either task.

The accuracy data are presented in Table 11. They were analyzed similarly, in 64 (presentations) \times 2 (target presence) ANOVAs performed on each task separately. No effects were significant in either analysis.

Table 11
Percentage Correct in Each Condition of Experiment 7
(Multisession Dual Task) as a Function of Practice Session

Session	Task 1		Task 2	
	Present	Absent	Present	Absent
1	96	98	96	98
2	96	99	97	98
3	97	99	97	98
4	96	98	98	97
5	94	95	96	97

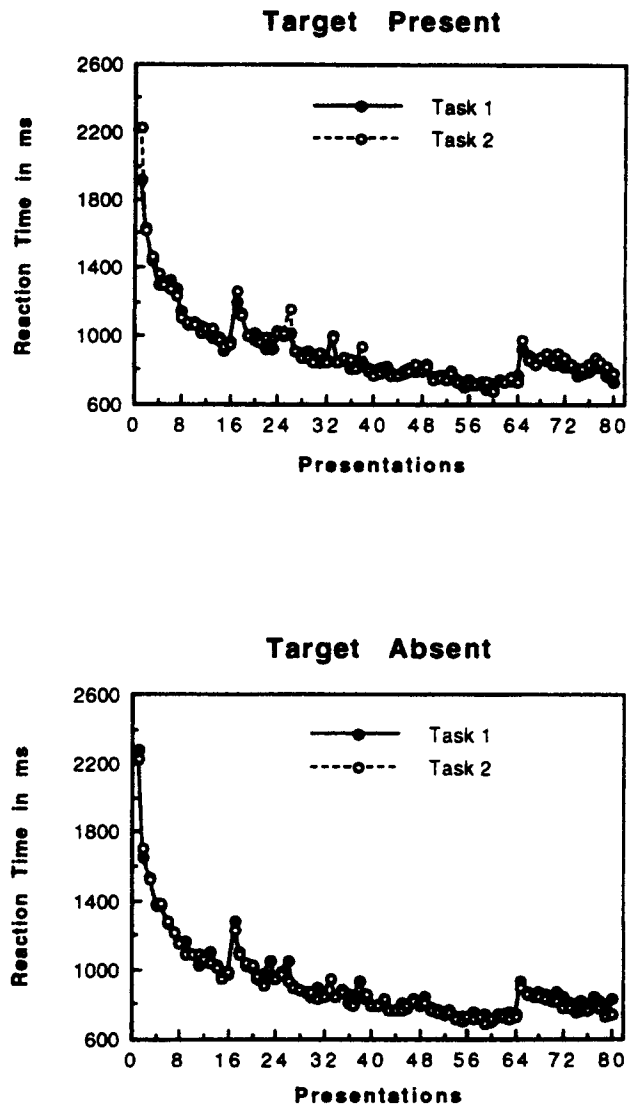


Figure 7. Mean reaction time as a function of the number of presentations for target present (top of figure) and target absent (bottom of figure) responses in Experiment 7 (dual task). Filled circles = Task 1; open circles = Task 2.

Transfer. The mean reaction times from the transfer session are also plotted in Figure 7. Reaction time increased sharply on the first transfer block and remained elevated thereafter, decreasing gradually over the session. The increase from the last block of training to the first block of transfer was 191 ms, close to the 241-ms cost at the end of one session in Experiment 6. The 191-ms increase was also large compared with the transition between Sessions 3 and 4 (24 ms); changing pairing had more of a cost than changing sessions. Moreover, changing pairing had a persistent cost: Reaction time increased 87 ms on average from Session 4 to Session 5. The average cost was smaller than the initial cost but still large relative to the cost in multisession divided attention (103 ms) and in multisession focused attention (13 ms). The cost

diminished over the session, however. In the last transfer block, reaction time was only 28 ms slower than the last training block.

These conclusions were confirmed in a 2 (session: training vs. transfer) \times 16 (presentations within sessions) \times 2 (target presence) \times 2 (Task 1 vs. Task 2) ANOVA on the mean reaction times. The main effect of sessions approached significance, $F(1, 11) = 4.12, p < .07, MS_e = 692,816.67$. Eight of the 12 subjects were slower on average in the transfer block than in the training block. A contrast comparing the last training block with the first transfer block was significant, $F(1, 165) = 49.58, p < .01$, tested against the $MS_e (17,660.56)$ from the interaction between session and presentations. All 12 subjects were slower on the first transfer block than on the last training block. We compared mean reaction time in each transfer block with the mean reaction time from the last training block (which, coincidentally, was the same as the mean reaction time for the last training session, 742 ms) by using Fisher's least significant difference test. All of the transfer blocks were significantly slower than the last training block ($p < .05$), except for Blocks 10, 11, 12, 15, and 16.

The main effect of presentations was significant, $F(15, 165) = 6.34, p < .01, MS_e = 16,636.08$, but the main effect of target presence was not ($F < 1$), and neither was the main effect of task, $F(1, 11) = 3.53, MS_e = 8,125.91$. The interaction between task and target presence was significant, $F(1, 11) = 6.15, p < .05, MS_e = 20,520.35$, as was the interaction between task, target presence, and session, $F(1, 11) = 6.53, p < .05, MS_e = 14,290.45$. No other effects were significant.

The accuracy data also appear in Table 11. They were analyzed in a 2 (session: training vs. transfer) \times 16 (presentations within sessions) \times 2 (target presence) \times 2 (Task 1 vs. Task 2) ANOVA. The main effect of session was not significant, $F(1, 11) = 2.22, MS_e = 515.51$. The only significant effect was the interaction between session and task, $F(1, 11) = 9.17, p < .05, MS_e = 32.15$.

Discussion

This experiment showed evidence of sensitivity to co-occurrence after extended practice in dual-task conditions. Sensitivity to co-occurrence is consistent with the attention hypothesis. Sensitivity to co-occurrence was larger initially and almost as large on average as the sensitivity observed in multisession divided attention (Experiment 2).

Evidence of Attention and Automaticity

The experiments were intended to investigate the effects of attention on what was learned during automatization. Were we successful in manipulating attention, and were we successful in producing automaticity? Two between-experiment contrasts test the success of our attention manipulation: Performance should be better in the focused-attention experiments than in the divided-attention experiments, and performance should be better in the divided-attention experiments than in the dual-task experiments. Our success in producing automaticity can be tested in three ways: a power function speedup in reaction time, a reduction in load effects, and a reduction in dual-task

interference. The power function speedup is a within-experiment (within-condition) effect; the reduction in load effects and dual-task interference are between-experiment comparisons. None of these effects were tested formally up to this point. This section reports and discusses the tests.

Power Function Speedup

The power function speedup was tested by fitting power functions to the mean reaction times in each condition of each experiment. The function fitted to the data was

$$RT = a + bN^{-c}, \quad (1)$$

where RT is reaction time; a is the asymptote, reflecting an irreducible limit on performance; b is the difference between initial and asymptotic performance, reflecting the amount of performance improvement that can result from learning; N is the amount of practice, measured in trials per item; and the exponent c is the learning rate.

The power function speed-up in mean reaction time is predicted by all major theories of automaticity and skill acquisition. The instance theory also predicts a power function reduction in the standard deviation of reaction times, and it predicts that the exponent of the power function for standard deviations should equal the exponent of the power function for mean reaction times (Logan, 1988). Some other theories can account for this effect (see Anderson, 1992; J. Cohen et al., 1990), but it does not follow as a logical consequence of their initial assumptions, as it does with instance theory (Logan, 1992). We fitted power functions (Equation 1) to standard deviations of reaction times as well to test the instance theory prediction that the exponents would be equal.

Finally, we fitted power functions to means and standard deviations simultaneously, constraining them to have the same exponents but allowing them to have different scaling parameters (a and b). These fits were a direct test of the instance theory prediction that means and standard deviations would decrease as power functions of practice with the same exponent. If means and standard deviations are linked in that way, they should be well fit by constrained power functions. The fit should not be (much) worse than the unconstrained fits.

Power functions were fitted to means, standard deviations, and means and standard deviations simultaneously from the average data in each condition of each experiment. Altogether, 32 data sets were fitted. Table 12 contains two measures of goodness of fit: the squared product-moment correlation (r^2) between observed and predicted values and the square root of the mean squared difference (or deviation; $rmsd$) between observed and predicted values.

Table 12 also contains composite measures of r^2 from the separate fits to means and standard deviations, to be compared with r^2 from the constrained fits. The composite r^2 was a single correlation between observed and predicted values calculated over both means and standard deviations, using power functions from the separate fits to generate the predicted values. We formed the composite measure because the r^2 from the constrained fits reflects in part the ability of the separate scaling parameters to capture the large difference between

Table 12
Measures of Goodness of Fit (r^2 and $rmsd$) for Power Functions (Equation 1) Fitted to Mean Reaction Times, Standard Deviations of Reaction Times, and Means and Standard Deviations Simultaneously From Each Condition of Each Experiment

Experiment	r^2				$rmsd$		
	RT	SD	Both	Comp.	RT	SD	Both
1: Single-session divided							
Consistent present	.968	.856	.999	.999	7.1	9.3	10.2
Consistent absent	.982	.816	.999	.999	7.4	6.7	7.5
Varied present	.955	.951	.999	.999	7.4	5.3	8.3
Varied absent	.966	.816	.999	.999	9.0	6.2	8.6
2: Multisession divided							
Present	.888	.517	.996	.996	15.2	18.7	17.4
Absent	.926	.744	.997	.997	16.2	14.5	15.5
3: Single-session focused							
Consistent present	.919	.619	.999	.999	7.2	7.7	7.6
Consistent absent	.854	.725	.998	.999	10.5	10.8	9.7
Varied present	.828	.731	.999	.999	6.6	6.5	6.4
Varied absent	.773	.650	.999	.999	8.5	8.7	8.7
4: Multisession focused							
Present	.705	.215	.994	.994	14.9	22.7	19.2
Absent	.707	.149	.993	.993	17.5	24.6	21.4
5: Single-session focused							
0-ms cue delay							
Consistent present	.897	.858	.999	.999	9.1	5.6	8.1
Consistent absent	.911	.943	.999	.999	9.4	3.3	7.3
Varied present	.885	.671	.999	.998	7.9	10.1	9.2
Varied absent	.771	.544	.998	.998	13.3	11.3	12.3
500-ms cue delay							
Consistent present	.876	.795	.999	.999	6.3	8.7	7.6
Consistent absent	.893	.803	.999	.999	7.0	7.9	7.7
Varied present	.941	.912	.999	.999	6.7	5.1	7.5
Varied absent	.897	.826	.999	.999	8.6	5.9	8.7
6: Single-session dual task							
Task 1							
Consistent present	.997	.986	.993	.997	22.8	29.8	43.2
Consistent absent	.995	.972	.996	.997	28.0	35.8	33.7
Varied present	.996	.957	.992	.995	20.0	43.7	44.0
Varied absent	.994	.954	.997	.997	22.6	28.0	28.3
Task 2							
Consistent present	.995	.948	.996	.996	22.9	35.0	30.5
Consistent absent	.997	.980	.994	.997	21.2	38.3	42.6
Varied present	.993	.933	.995	.996	22.6	37.5	35.6
Varied absent	.998	.993	.993	.999	14.4	20.4	43.4
7: Multisession dual task							
Task 1							
Present	.941	.824	.982	.983	55.6	43.8	51.0
Absent	.946	.740	.964	.967	61.1	82.2	75.7
Task 2							
Present	.920	.730	.956	.959	72.1	89.1	84.5
Absent	.965	.888	.987	.987	49.1	39.2	44.5

Note. RT = fits to mean reaction time. SD = fits to standard deviation of reaction time; Both = fits to means and standard deviations simultaneously, constraining them to have the same exponent; Comp. = the composite score from the separate fits.

means and standard deviations. (The separate a and b parameters can be adjusted arbitrarily to account for the average differences between means and standard deviations.) This may inflate r^2 from the constrained fits, relative to the average r^2 from the separate fits. The composite measure is an improvement, in that it also reflects in part the ability of the separate scaling parameters to capture the large difference between means and standard deviations. In addition, it reflects the ability of separate exponents to capture differences in the shapes of the learning curves for means and standard deviations

that are contrary to the predictions of instance theory. The difference between the composite measures and measures from the constrained fits reflects the importance of variation in the exponents.

Fits to mean reaction times. The measures in Table 12 suggest that power functions provided satisfactory fits to the mean reaction times in each experiment, confirming precedent and fulfilling the power function speedup criterion for automaticity. The average r^2 was .915, and the average $rmsd$ was 19.0 ms. Two of the data sets were fit by "degenerate" power

Table 13
Exponents of Power Functions (Equation 1) Fitted to Mean Reaction Times, Standard Deviations of Reaction Times, and Means and Standard Deviations Simultaneously From Each Condition of Each Experiment

Experiment	RT	SD	Both
1: Single-session divided			
Consistent present	1.388	7.976	1.909
Consistent absent	0.771	1.553	0.817
Varied present	0.785	2.309	1.101
Varied absent	0.664	2.122	0.729
2: Multisession divided			
Present	0.337	0.281	0.329
Absent	0.513	0.645	0.530
3: Single-session focused			
Consistent present	0.843	1.262	0.921
Consistent absent	1.184	10.000 ^a	1.692
Varied present	2.822	10.000 ^a	4.119
Varied absent	10.000 ^a	10.000 ^a	10.000 ^a
4: Multisession focused			
Present	0.231	0.315	0.257
Absent	0.235	0.563	0.269
5: Single-session focused			
0-ms cue delay			
Consistent present	1.141	3.743	1.374
Consistent absent	0.705	1.289	0.813
Varied present	3.352	10.000 ^a	9.999
Varied absent	10.000 ^a	10.000 ^a	10.000 ^a
500-ms cue delay			
Consistent present	4.683	10.000 ^a	10.000 ^a
Consistent absent	2.805	10.000 ^a	10.000 ^a
Varied present	1.420	10.000 ^a	2.077
Varied absent	1.252	10.000 ^a	1.960
6: Single-session dual task			
Task 1			
Consistent present	0.786	1.527	0.960
Consistent absent	0.616	0.808	0.654
Varied present	1.103	1.995	1.331
Varied absent	0.901	1.346	0.965
Task 2			
Consistent present	0.531	0.733	0.571
Consistent absent	0.773	1.333	0.923
Varied present	0.927	1.521	1.043
Varied absent	1.093	2.239	1.382
7: Multisession dual task			
Task 1			
Present	0.251	0.322	0.313
Absent	0.383	0.818	0.467
Task 2			
Present	0.392	0.876	0.503
Absent	0.357	0.388	0.376

Note. RT = fits to mean reaction time. SD = fits to standard deviation of reaction time; Both = fits to means and standard deviations simultaneously, constraining them to have the same exponent.

^aExponents of 10 reflect degenerate power functions.

functions with unacceptably large exponents. The varied-pairing target-absent data from Experiment 3 (single-session focused attention) and the varied-pairing target-absent data from the 0-ms delay condition of Experiment 5 (single-session focused attention with arrow cues) produced exponents of 10 (see Table 13), which represent the limit imposed by the fitting program. An exponent of 10 produces a power function in which all the learning happens on the first presentation. The

first point is higher than all of the others, and all of the rest of the points are not different from each other.

Fits to standard deviations of reaction time. The fits for standard deviations were not as good as the fits for means (see Table 12). The average r^2 was .782, and the average *rmsd* was 22.5 ms. Nine of the fits were degenerate, in that they produced unacceptably large exponents. All of them came from the single-session focused-attention experiments (Experiments 3 and 5; see Table 13). The fits in multisession focused attention (Experiment 4) produced low r^2 values, .215 and .149 for target present and target absent, respectively. The fits in the other 23 data sets were reasonable. They confirm precedent and confirm a prediction of the instance theory.

Constrained fits to means and standard deviations. The constrained fits were reasonable in most cases. There were four degenerate fits, one from Experiment 3 and three from Experiment 5. The average r^2 was .995, and the average *rmsd* was 24.0 ms, not much different from the average r^2 (.995) and *rmsd* (21.1 ms) from the composite measures of the separate fits. The goodness of the constrained fits confirm predictions of the instance theory (see also Logan, 1988; Strayer & Kramer, 1990).

Equal exponents for means and standard deviations. From the perspective of the theories at issue in this article, the most important parameters of the power functions are the exponents, c . The instance theory predicts that the exponents should be the same for means and standard deviations. The exponents from each fit are presented in Table 13. In 31 of 32 cases, the exponent for the standard deviation was larger than the exponent for the mean, contrary to the instance theory prediction.

Does this difference falsify the instance theory or call into question the applicability of the instance theory to these data? Perhaps not. We suspect that the observed difference may be an artifact of low levels of practice. The difference was much larger in the single-session experiments (the average exponent for the means was 0.862, compared with 1.191 for the standard deviations) than in the multisession experiments (0.337 vs. 0.526, respectively). The difference in the multisession experiments was small relative to the variation in exponents.

The exponents from the multisession experiments were generally smaller than the exponents from the single-session experiments, averaging 0.337 versus 0.862 for means and 0.526 versus 1.191 for standard deviations. These differences could reflect strategic differences between single-session and multisession subjects (e.g., multisession subjects may try harder), or they could be artifacts of estimating parameters with different numbers of data points. Small numbers of data points give more weight to the rapid changes seen in the first few presentations. Larger numbers of data points allow the more gradual changes seen on later presentations to influence parameter estimation. To test for measurement artifacts, we fitted power functions to the first session of the multisession experiments, using 16 data points per function as we did in the single-session fits. The exponents rose sharply. The average exponent for the means was 0.627 (vs. 0.337); the average for the standard deviations was 1.388 (vs. 0.526). More important, the exponents diverged more than they did when they were

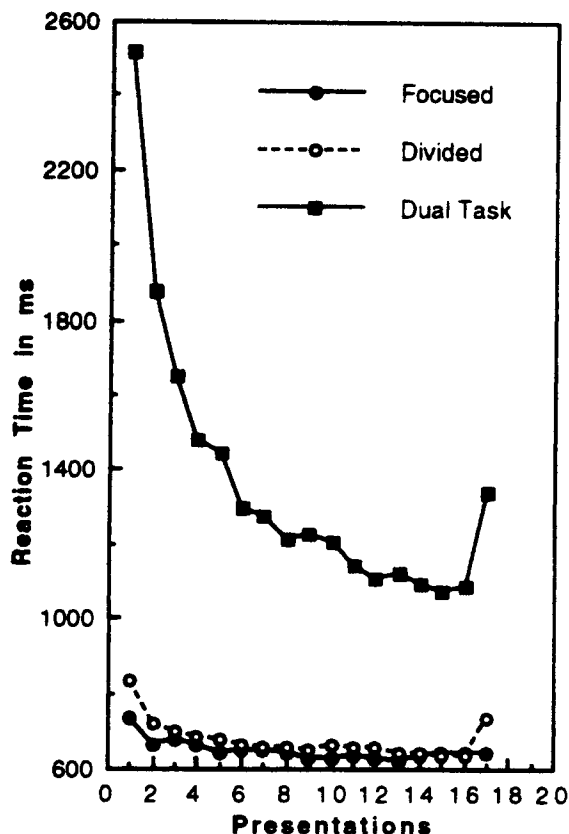


Figure 8. Mean reaction time as a function of the number of presentations for focused attention (Experiment 3; filled circles), divided attention (Experiment 1; open circles) and dual task (Experiment 6; filled squares). Data are from the consistent-pairing conditions of each experiment, averaged over target presence. Dual-task data are averaged over tasks as well.

based on all 64 data points; the difference was 0.761 in the fits to the first session versus 0.189 in the fits to all four sessions.²

We draw two conclusions from this analysis: First, the large divergence between exponents for means and standard deviations in the single-session data may have been inflated by a measurement artifact. Second, the smaller divergence in the multisession experiments may disappear if enough data were collected. These conclusions allow us some latitude in interpreting the observed differences in exponents. We may not have had enough data points to test the instance theory predictions adequately. Logan (1992) also found better fits with larger numbers of data points.

Load Effects

If our attention manipulation were successful, focused- and divided-attention conditions should differ in processing load. Performance should be better with focused attention than with divided attention. If we were successful in producing automaticity, the processing load effects should diminish with practice. These hypotheses were tested twice: once in the single-session data and once in the multisession data. Mean reaction times from the consistent-pairing conditions of the single-session

experiments (Experiment 1 vs. Experiment 3) are plotted in Figure 8. Mean reaction times from the multisession experiments (Experiment 2 vs. Experiment 4) are plotted in Figure 9. Mean reaction times from the 1st and 16th blocks of the single-session experiments and the first and fourth sessions of the multisession experiments are presented in Table 14.

In each case, divided attention was worse than focused attention early in practice, and the difference diminished. These effects demonstrate a load effect that indicates that our attention manipulation was successful and a reduction in load effects with practice that fulfills our second criterion for automaticity.

Performance on the recall test also suggests that our manipulation of attention was successful. Subjects in divided attention (Experiment 1) recalled as many distractors as they did nontargets (36.1% vs. 35.8%, respectively), which suggests they paid equal attention to them. By contrast, subjects in focused attention (Experiments 3 and 5) recalled fewer distractors than nontargets (23.8% vs. 41.6%, respectively, averaging across experiments), which suggests they paid less attention to them (nontargets were cued, but distractors were not).

Dual-Task Interference

If our attention manipulation was successful, dual-task conditions should be harder than divided-attention conditions.

² Two factors contribute to the distortion of the exponents in these experiments. The first is theoretical: The instance theory predictions apply best to memory-based performance, when the algorithm does not contribute to performance. The algorithm necessarily determines performance on the first trial because there are no prior memories to be retrieved. The algorithm also influences the early portion of the learning curve, where it is fast enough to win the race with memory. Its influence drops out at higher levels of practice when memory nearly always wins the race (Logan, 1988). Thus, distortions from the algorithm affect the initial part of the learning curve. We should expect more distortion in fits based on smaller amounts of practice because those fits will emphasize the contribution of the algorithm. Logan (1992) found improved fits to instance theory distributional predictions when the first few sessions were dropped from the fitting. The second factor contributing to the distortion in the exponents is statistical: Power functions fitted to small amounts of noisy data appear to overestimate the exponent. We performed a number of Monte Carlo simulations, in which we generated power functions with known parameters, added noise to them, and then fitted power functions to the noisy data and compared the fitted parameters with the known ones. We varied the number of points we fitted (8, 16, 32, 48, and 64) and the amount of noise we added. We sampled noise randomly from a normal distribution with a mean of 0 and a standard deviation of 0, 10, 15, 20, or 25. (The standard deviation of the noise corresponds to the standard error of the mean, which may be calculated by taking the square root of the mean square error from the ANOVA and dividing it by the square root of N .) When there was no noise ($SD = 0$) there was no bias. The estimated parameters were almost identical with the known parameters regardless of the number of points in the function. When noise was added, the picture changed. Estimated exponents were larger than the known ones when the fits were based on 8 and 16 points. Bias was negligible for fits based on 32 or more points.

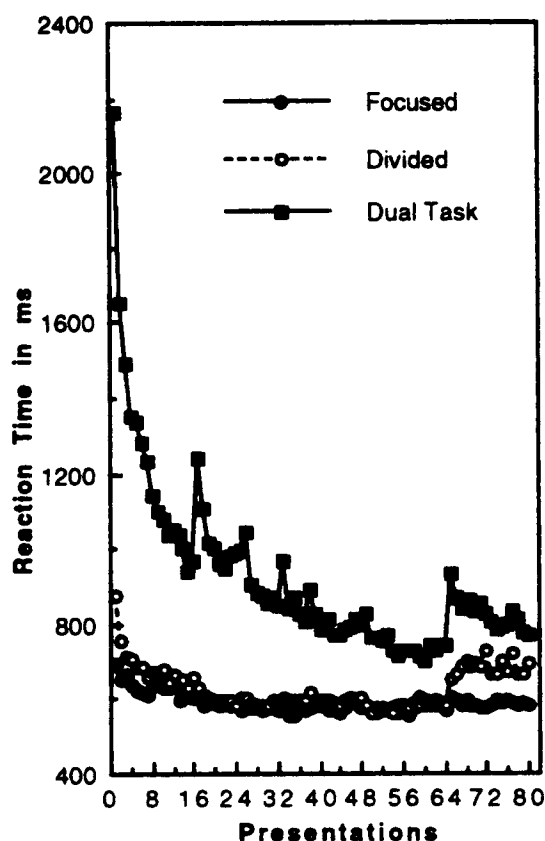


Figure 9. Mean reaction time as a function of the number of presentations for focused attention (Experiment 4; filled circles), divided attention (Experiment 2; open circles), and dual task (Experiment 7; filled squares).

If we succeeded in producing automaticity, this dual-task interference should diminish with practice. We tested these hypotheses twice: once comparing the single-session divided-attention experiment with the single-session dual-task experiment and once comparing the multisession divided-attention experiment with the multisession dual-task experiment. Both tests focused on consistent-pairing conditions. Mean reaction times from the single-session experiments (Experiment 6 vs. Experiment 1) are plotted in Figure 8; mean reaction times from the multisession experiments (Experiment 7 vs. Experiment 2) are plotted in Figure 9. The mean reaction times from the 1st and 16th blocks of the single-session experiments and the first and fourth sessions of the multisession experiments appear in Table 14.

In each case, dual-task performance was worse than divided-attention performance initially, and the difference became smaller with practice. These effects demonstrate dual-task interference, which means our attention manipulation was successful, and show a reduction in dual-task interference that fulfills our third criterion for automaticity.

Automaticity and Extended Practice

Qualitatively, the single-session experiments produced the same results as the multisession experiments. The three

criteria for automaticity were fulfilled both in single-session and in multisession experiments. Extended practice does not seem necessary to produce automaticity (see also Logan & Klapp, 1991). Moreover, the effects of consistency of pairing were the same in the single-session and in the multisession experiments. The same effects can be seen with small and large amounts of practice. This is now the fourth time we have demonstrated that point in our laboratory (see Lassaline & Logan, 1993; Logan, 1988; Logan & Klapp, 1991; see also Strayer & Kramer, 1990). Perhaps future investigators (and readers) will be more willing to accept experiments with low levels of practice as evidence for automaticity. The amount of practice is not the issue. Changes in performance with practice are the issue. Automaticity is evident if reaction time gets faster, load effects get smaller, and dual-task interference diminishes. Reaction time does not have to reach asymptote, and load effects and dual-task interference do not have to disappear entirely. All that is required is a detectable change in the direction of automaticity, and that may be observed after a small amount of practice.

Quantitatively, the distributional predictions of the instance theory were confirmed more readily in the multisession experiments than in the single-session experiments. Power function exponents for means were closer to the exponents for standard deviations in the multisession experiments. We have also seen this before in our laboratory (Logan, 1988, 1992). However, this is the first time we compared different levels of practice on the same task. The fits to the first session of the multisession experiments (see previous section) suggest that the problem stems from estimating parameters from the initial portion of the learning curve. Instance theory predictions are clearest at

Table 14

Benefits From Training and Costs From Transfer for Mean Reaction Times From Each Experiment

Experiment	First	Last	Transfer	Cost
1: Single-session divided				
Consistent pairing	833	637	732	95
Varied pairing	846	681	682	1
2: Multisession divided				
Consistent pairing	689	581	684	103
3: Single-session focused				
Consistent pairing	735	643	647	4
Varied pairing	718	668	651	-17
4: Multisession focused				
Consistent pairing	632	573	586	13
5: Single-session focused				
0-ms cue delay				
Consistent pairing	776	656	680	24
Varied pairing	781	690	686	-4
500-ms cue delay				
Consistent pairing	695	625	621	-4
Varied pairing	736	621	619	-2
6: Single-session task				
Consistent pairing	2,544	1,082	1,323	241
Varied pairing	2,483	1,231	1,211	-20
7: Multisession dual task				
Consistent pairing	1,241	742	829	87

Note. First, last, and transfer refer to Blocks 1, 16, and 17 in the single-session experiments and Sessions 1, 4, and 5 in the multisession experiments.

high levels of practice when the algorithm has dropped out and differences in retrieval time distributions no longer matter (see also Logan, 1992).

These findings lead us to conclude that the importance of extended practice in the designs of automaticity experiments depends on the nature of the hypotheses the researchers intend to test. Researchers interested in testing qualitative hypotheses should be able to test them adequately with relatively low levels of practice. Many times, qualitative hypotheses can discriminate between major theories. Thus, important research on automaticity can be done with relatively low levels of practice. However, researchers testing quantitative predictions of the instance theory and other formal theories of automatization should run experiments with much higher levels of practice to stabilize parameter estimation.

General Discussion

What is learned during automatization, and what role does attention play in learning? The instance theory says that instances are learned—separate representations of episodic co-occurrences—and that attention determines which co-occurrences go into an instance. We tested these hypotheses in Experiments 1–7 by manipulating co-occurrence (consistency of pairing) and attention at the same time. The results, summarized in Table 14, were clear. Subjects showed strong sensitivity to co-occurrence in the divided-attention experiments: At the end of training, consistent-pairing subjects were 44 ms faster than varied-pairing subjects. When the pairing changed at transfer, consistent-pairing subjects were slowed by 99 ms. Subjects showed a similar sensitivity to co-occurrence in the dual-task experiments: Consistent-pairing subjects were 149 ms faster than varied-pairing subjects at the end of training, and they were slowed by 164 ms when pairing changed at transfer. By contrast, subjects in the focused-attention experiments showed very little sensitivity to co-occurrence. Consistent-pairing subjects were only 19 ms faster than varied-pairing subjects at the end of training, and changing pairing at transfer cost them only 9 ms.

The results were consistent with the attention hypothesis, derived from the instance theory. Subjects were sensitive to co-occurrence when the task required them to pay attention to both words (i.e., the divided-attention and dual-task experiments) and insensitive to co-occurrence when the task allowed them to pay attention to one word and ignore the other (i.e., the focused-attention experiments). The results suggest that performance was based on instance retrieval and that attention determined what got into an instance. The remainder of this section discusses implications of these conclusions.

Co-Occurrence and Instance Representation

The experiments showed that the memory representations that support automatic performance contain information about co-occurrence. This is an important demonstration because representation of co-occurrence is characteristic of instance theories in general (Barsalou, 1990; Hintzman, 1976, 1986; Jacoby & Brooks, 1984; Medin & Schaffer, 1978) and assumed by the instance theory of automaticity (Logan, 1988, 1990,

1992). The analysis of transfer performance in Experiments 1 and 2 suggested that two kinds of associations underlie automatic performance: associations between single words and responses (i.e., targets and responses and distractors and responses) and associations between word pairs and responses. The associations between word pairs and responses provide the clearest evidence of representation of co-occurrence, but the associations between single words and responses can also be interpreted as representations of co-occurrence (i.e., of words and responses; see also Logan, 1990).

Is the evidence of co-occurrence evidence of instance representation? The instance representation assumption assumes that instances are stored and retrieved separately as well as assuming that instances represent co-occurrence. Evidence of co-occurrence is a necessary part of the evidence for instance representation, but it is not sufficient in itself. The power function fits address the assumption that instances are separate because that assumption is essential in the proofs of the power function prediction. The power function speed-up follows as a logical consequence of the assumption of separate instances (and a few others; see Logan, 1988, 1992). However, the power function fits are not a unique prediction of the instance theory; some strength theories can account for the simultaneous fits to means and standard deviations (Anderson, 1992; J. Cohen et al., 1990; but see Logan, 1992).

More important, the effects of consistency of pairing, on which evidence of co-occurrence is based, can be predicted by strength models that use standard connectionist learning mechanisms such as the Hebb rule or back propagation (Rumelhart, Hinton, & Williams, 1986). These learning algorithms are sensitive to consistent correlations among features in their input. Consistent pairing would lead to an advantage over varied pairing because both the target and the context would activate the response. Changing the pairing following consistent training would also disrupt performance because the network would have been trained to be maximally responsive to the consistently repeated pairs. New pairs would evoke a weaker response.

The varied-pairing conditions may provide a better test of instance versus strength representations because instance representations would preserve the details of individual pairings, whereas many strength representations would average them away (Barsalou, 1990). Different retrieval tasks could be designed to distinguish between alternatives (see, e.g., Hintzman, 1976). The effects could appear in automaticity experiments if a few pairs were repeated occasionally in a varied-pairing condition. *Steel* might appear twice with *Canada* and once with *Davenport* and *Broccoli*. Instance theory would predict benefit on the second presentation with *Canada*. Strength theories that average away details would not predict an advantage. The prior presentation of *Canada* would be merged with prior presentations of *Davenport* and *Broccoli* and lost to the retrieval process. Unfortunately, we could not test for these effects in our data.

Thus, the evidence that co-occurrence was represented does not force us to conclude that the representations were instances. The evidence is certainly consistent with instance representation, but it is also consistent (in principle) with

certain strength models. The evidence that co-occurrence was represented is important nevertheless because it is suggestive of instance representation and because it has implications for other theories of skill acquisition and automaticity. It is a constraint that they must account for. Moreover, the main focus of the article is on the role of attention in learning co-occurrences, and for that purpose, it is not absolutely essential that the co-occurrences are represented as instances.

Co-Occurrence in Other Theories of Skill Acquisition and Automaticity

The conclusion that automatic performance depends on representations of co-occurrence has implications for N. Cohen and Squire's (1980; Squire, 1992) procedural-declarative theory of learning and memory and Schneider's (1985; Schneider & Detweiler, 1987, 1988) connectionist-control theory of automatization. N. Cohen and Squire argued that declarative and procedural memory rely on different physiological substrates and use different kinds of representation. Declarative memory depends on hippocampal structures and uses "relational" representations to represent arbitrary co-occurrences, instances. Procedural memory relies on nonhippocampal systems and uses "nonrelational" representations. Declarative memory supports instancelike episodic memory phenomena, whereas procedural memory supports skill acquisition. The present evidence that automatic performance depends on representations of co-occurrence is not compatible with these assumptions. The present evidence suggests that automaticity is a declarative memory phenomenon (see also Logan, 1988, Experiment 5; Logan, 1990) or that procedural memory can represent arbitrary co-occurrences. The idea that automaticity may depend in part on declarative memory need not be inconsistent with the theory: Automaticity may depend on both kinds of memory. The idea that procedural representations capture arbitrary co-occurrences seems contrary to the basic assumptions of the theory.

The conclusion that automatic performance depends on representations of co-occurrence is also incompatible with Schneider's (1985; Schneider & Detweiler, 1987, 1988) connectionist-control theory of automatization. Schneider's model is very explicit about what is learned during automatization. It assumes that learning occurs only during controlled processing (Schneider & Detweiler, 1987, p. 66), and it assumes that controlled processing is strictly serial (Schneider, 1985, pp. 480–484; Schneider & Detweiler, 1988, p. 550). Thus, subjects may learn about occurrences of individual items that are presented to them, but they will not learn how they co-occur. Experiments 1 and 2 are especially relevant because they are a nearly exact implementation of the task that Schneider (1985) explicitly modeled. Schneider described in detail what the model learned when it searched for category targets in divided-attention conditions, and the model did not learn co-occurrence.

Schneider's model may not predict learning of co-occurrences in its current version (i.e., Schneider, 1985; Schneider & Detweiler, 1987, 1988), but it should be possible to change it to accommodate the present results. First, the assumptions about strict serial processing could be relaxed: If subjects paid some

attention to the distractors while they were focusing on the targets (or vice versa), they would learn about co-occurrence. Schneider's (Hebb) learning algorithm was designed to pick up regularities like the pairing manipulation in the present experiments. There are limits to this approach, however. The idea that controlled processing is serial is a key assumption in Schneider's theory (1985, pp. 480–484; Schneider & Detweiler, 1988, p. 550) and has been central to his thinking since the seminal papers (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). It cannot be changed much or relaxed very far without changing the essence of the theory. The data from the focused-attention experiments provide an empirical constraint on how far serial processing assumptions can be relaxed. In serial processing models, cues to focus attention change the order in which display items are processed but not the nature of processing. Thus, subjects in focused attention should display the same sensitivity to co-occurrence as subjects who process the display serially. The data suggest that sensitivity is low (see Table 14). Serial processing with the same "leakage" observed in the focused-attention experiments could not account for subjects' sensitivity to co-occurrence in the divided-attention experiments.

Second, it may be possible to account for sensitivity to co-occurrence within the same connectionist-control architecture by extending Schneider and Detweiler's (1987, 1988) theory of working memory, which accounts for short- and long-term memory phenomena as well as automaticity. That theory assumes that items are associated with their context, and that context acts as a retrieval cue. Schneider and Detweiler (1987) interpreted context as a "temporal tag" (p. 65) that could be used to relate different items that occurred at the same time—co-occurring items would have the same temporal tag. In the current model, temporal tags are short-lived and quickly forgotten (Schneider & Detweiler, 1987, p. 67). They may not last long enough to account for the results of the multisession experiments (2, 4, and 7), in which the first transfer trial occurred about 24 hr after the last training trial. Longer lasting temporal tags would be necessary to explain those effects. It is not clear how essential it is for the model to have short-lasting temporal tags. Longer lasting tags may destroy the model's ability to account for short- and long-term memory phenomena, which are currently explained with short-lived temporal tags. Fortunately, the answer is empirical: We can run the simulations and find out what happens.

Third, the assumptions about learning could be changed or augmented. Schneider's (1985) model learned according to a Hebb learning rule that associated single items with single responses. It may be possible to replace this learning rule or supplement it with a simple recurrent net that associates successive states of the network (Elman, 1990) so that *Canada* followed by *Steel* would build an association between *Canada* and *Steel*. It remains to be seen whether the model would work in a different way with a different learning algorithm. The model makes a number of interrelated assumptions that interact in determining how the model performs. Changing one assumption to accommodate a new result may leave the model's performance intact, or it may change it so much that

the model no longer accounts for the original results it was built to account for. This is an empirical question to be addressed by simulation.

The Attention Hypothesis

The results suggest that attention determines what gets into an instance. Attention determines what subjects learn about individual words in the display, as evidenced by the contrast between focused and divided attention in recall of targets and distractors (Experiment 1 vs. Experiments 3 and 5), and attention determines whether subjects learn co-occurrences of words, as evidenced by the contrast in the reaction time data between focused-attention conditions on the one hand and divided-attention and dual-task conditions on the other (Experiments 3–5 vs. Experiments 1, 2, 6, and 7). These results are consistent with the attention hypothesis and encourage further investigation.

The attention hypothesis has relevance beyond the automaticity literature. It concerns the effects of attention on a single presentation, predicting that subjects will learn what they attend to. That learning may be evident later on as automaticity, explicit memory, implicit memory, or implicit learning, depending on the retrieval task (and the perspective of the investigator). The attention hypothesis is about encoding rather than retrieval, so it should apply regardless of the retrieval task.

The attention hypothesis has important implications for the memory literature. It implies that the memorability of particular attributes or particular kinds of attributes is not absolute, but rather, depends on whether they are attended. In principle, all kinds of attributes can be attended, so all kinds of attributes can be preserved in instances. All kinds of attributes can be ignored, so no attribute will be retained in every instance (cf. Hasher & Zacks, 1979). Instances can be concrete or abstract, depending on what is attended (cf. Brown & Carr, 1993; Carr, Brown, & Charalambous, 1989).

The strength of the attention hypothesis lies in its testability. It makes predictions for immediate performance as well as memory. Immediate performance can often be understood with reference to the attention literature (Broadbent, 1958; Duncan, 1980; Kahneman, 1973; Pashler, 1989; Posner & Boies, 1971; Treisman, 1969). Memory predictions require a detailed task analysis to figure out what will be in an instance. The analysis needs to show (a) what traces are produced by running the initial algorithm and (b) whether any of those traces can support subsequent memory-based performance. Traces that fulfill both conditions will be retained in an instance. The nature of the traces that fulfill both conditions depends on the algorithm and the retrieval task.

The attention hypothesis is related to ideas like Kirsner and Dunn's (1985) perceptual record, Morris, Bransford, and Franks's (1977) transfer-appropriate processing (also see Roediger, 1990), and even Craik and Lockhart's (1972) levels of processing. Those ideas suggest that memory records the "trajectory" a stimulus follows through the information-processing system, and memory performance depends on the similarity of the trajectory taken at retrieval to the one taken at encoding. The greater the similarity, the better the perfor-

mance. The attention hypothesis is similar, except that the trajectory is defined in terms of the acts of attention necessary to perform the task. Attention is treated as an independent variable that can be manipulated to test specific predictions, not as a post hoc factor invoked to explain unanticipated results.

Toward a Propositional Theory of Attention

The attention hypothesis is an attempt to integrate literatures. Memory is not an important topic in modern attention theories, and attention is not an important topic in modern memory theories. Automaticity is a bridge between literatures, forcing theorists to deal with attention and memory at the same time. The attention hypothesis is important because it makes testable predictions about the relation between attention and memory and about the relation between immediate performance and subsequent performance on retrieval tasks. The attention hypothesis is very general, stating only that the things that are selected will be remembered. It does not explain how selection works or why selection results in memory. In the remainder of this section, we suggest how the attention hypothesis might be extended to deal with these issues.

We propose that attention constructs propositions. Instances are propositions (see also Logan, 1990, in press). Propositions are important because they are discrete representations, and their discreteness captures the idea that each instance is represented separately. Propositions are also important because they represent co-occurrence in a natural way. A proposition is a predicate with a truth value, and a predicate is a relation that takes one or more arguments. Multiargument relations express the co-occurrence of their arguments directly (e.g., "the ball is on the table" expresses the co-occurrence of *ball* and *table*). Different propositions can express co-occurrence indirectly through reference to the same arguments (e.g., "the ball is on the table" and "the cup is on the table" imply the co-occurrence of *ball* and *cup* because they are both "on the table"). Large episodic structures can be built by argument repetition (e.g., Kintsch & van Dijk, 1978).

Propositions are important to attention because they set computational goals for the attentional mechanisms to implement. They tell attention what to select. The predicated relation must be chosen from an indefinitely large number of possible relations, and the arguments to be related must be chosen from an indefinitely large number of possible arguments (Logan, in press). The choice of relation corresponds to analyzer selection in classical analyses of attention, and the choice of arguments corresponds to input selection (Posner & Boies, 1971; Treisman, 1969). Thus, propositions imply selectivity in the same sense as classical analyses.

Our idea that attention creates propositions stems from the definition of a proposition as a predicate with a truth value. The act of attending establishes the truth of a predicate, and in doing so, turns the predicate into a proposition. The idea that attention selects perceptual evidence is common (e.g., Broadbent, 1958; Kahneman, 1973; Posner & Boies, 1971; Treisman, 1969). We go a step further and assume that the evidence leads to a decision about what is likely to be true of the situation from which the evidence was sampled; it leads to a proposition.

Our view has implications for immediate performance: Propositions specify the computational goals that the attention system must satisfy, the goals shape the algorithms that the attention system implements, and the algorithms determine how performance unfolds in real time. Implications for memory are intertwined with implications for immediate performance: The algorithm determines what is attended when—which propositions get constructed—and that determines what is remembered on subsequent retrieval tests. We have only begun to explore the implications for immediate performance (Logan, in press).

References

- Anderson, J. R. (1992). Automaticity and the ACT* theory. *American Journal of Psychology*, 105, 165–180.
- Baehrick, H. P., & Shelly, C. (1958). Time-sharing as an index of automatization. *Journal of Experimental Psychology*, 48, 288–293.
- Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? *Journal of Experimental Psychology: Human Perception and Performance*, 10, 340–357.
- Barsalou, L. W. (1990). On the indistinguishability of exemplar memory and abstraction in category representation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 3, pp. 61–88). Hillsdale, NJ: Erlbaum.
- Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, 80, 1–46.
- Broadbent, D. E. (1958). *Perception and communication*. Elmsford, NY: Pergamon Press.
- Brown, J. S., & Carr, T. H. (1993). Limits on perceptual abstraction in reading: Asymmetric transfer between surface forms differing in typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 1277–1296.
- Carr, T. H., Brown, J. S., & Charalambous, A. (1989). Repetition and reading: Perceptual encoding mechanisms are very abstract but not very interactive. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 763–778.
- 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.
- Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, 97, 332–361.
- Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of “knowing how” and “knowing that.” *Science*, 210, 207–209.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–676.
- Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, 87, 272–300.
- Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14, 213–252.
- Eriksen, C. W., & Colegate, R. L. (1971). Selective attention and serial processing in briefly presented visual displays. *Perception and Psychophysics*, 10, 321–326.
- Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception and Psychophysics*, 12, 201–204.
- Etherton, J. L. (1992). *Encoding of position and color during the acquisition of automaticity*. Unpublished master's thesis, University of Illinois at Urbana-Champaign.
- Fisk, A. D., & Schneider, W. (1984). Consistent attending versus consistent responding in visual search: Task versus component consistency in automatic processing development. *Bulletin of the Psychonomic Society*, 22, 330–332.
- Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, 108, 356–388.
- Hintzman, D. (1976). Repetition and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 10 pp. 47–91). San Diego, CA: Academic Press.
- Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace model. *Psychological Review*, 93, 411–428.
- Hollender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking. *Brain and Behavioral Sciences*, 9, 1–23.
- Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 18, pp. 1–47). San Diego, CA: Academic Press.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.
- Kirsner, K., & Dunn, J. (1985). The perceptual record: A common factor in repetition priming and attribute retention. In M. I. Posner & O. S. Marin (Eds.), *Attention and performance XI* (pp. 547–565). Hillsdale, NJ: Erlbaum.
- Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Lassaline, M. L., & Logan, G. D. (1993). Memory-based automaticity in the discrimination of visual numerosity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 561–581.
- Logan, G. D. (1979). On the use of a concurrent memory load to measure attention and automaticity. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 189–207.
- Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology*, 39, 367–386.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95, 492–527.
- Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, 22, 1–35.
- Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 883–914.
- Logan, G. D. (in press). Linguistic and conceptual control of visual spatial attention. *Cognitive Psychology*.
- Logan, G. D., & Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 179–195.
- Logan, G. D., Withey, M. J., & Cowan, W. B. (1977). Cue search and comparison processes in visual search for letters. *Canadian Journal of Psychology*, 31, 113–121.
- MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 126–135.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207–238.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16, 519–533.
- Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition.

- tion and the law of practice. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 1-55). Hillsdale, NJ: Erlbaum.
- Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, 19, 1-32.
- Pashler, H. (1989). Dissociations and dependencies between speed and accuracy: Evidence for a two-component theory of divided attention in simple tasks. *Cognitive Psychology*, 21, 469-514.
- Pashler, H., & Johnston, J. C. (1989). Chronometric evidence for central postponement in temporally overlapping tasks. *Quarterly Journal of Experimental Psychology*, 41A, 19-45.
- Posner, M. I., & Boies, S. J. (1971). Components of attention. *Psychological Review*, 78, 391-408.
- Roediger, H. L., III. (1990). Implicit memory: Retention without remembering. *American Psychologist*, 45, 1043-1056.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318-364). Cambridge, MA: MIT Press.
- Schneider, W. (1985). Toward a model of attention and the development of automatic processing. In M. I. Posner & O. S. Marin (Eds.), *Attention and performance XI* (pp. 475-492). Hillsdale, NJ: Erlbaum.
- Schneider, W., & Detweiler, M. (1987). A connectionist/control architecture for working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 21, pp. 53-119). San Diego, CA: Academic Press.
- Schneider, W., & Detweiler, M. (1988). The role of practice in dual-task performance: Toward workload modeling in a connectionist/control architecture. *Human Factors*, 30, 539-566.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 84, 1-66.
- Schneider, W., & Shiffrin, R. M. (1985). Categorization (restructuring) and automatization: Two separable factors. *Psychological Review*, 92, 424-428.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127-190.
- Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99, 195-231.
- Strayer, D. L., & Kramer, A. F. (1990). An analysis of memory-based theories of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 291-304.
- Treisman, A. (1969). Strategies and models of selective attention. *Psychological Review*, 76, 282-299.
- Uyeda, K. M., & Mandler, G. (1980). Prototypicality norms for 28 semantic categories. *Behavior Research Methods and Instrumentation*, 12, 587-595.
- Venturino, M. (1991). Automatic processing, code dissimilarity, and the efficiency of successive memory searches. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 677-695.

Appendix

Words Used in the Experiments

Metals	Countries	Vegetables	Furniture
Iron	France	Carrot	Chair
Copper	America	Peas	Table
Steel	Russia	Corn	Bed
Gold	England	Bean	Sofa
Aluminum	Germany	Lettuce	Desk
Silver	Canada	Spinach	Lamp
Tin	Italy	Asparagus	Couch
Zinc	Spain	Broccoli	Dresser
Magnesium	Switzerland	Tomato	Footstool
Uranium	Norway	Cucumber	Buffet
Tungsten	Australia	Beets	Bench

Received March 29, 1993

Revision received October 25, 1993

Accepted October 29, 1993 ■