Journal of Experimental Psychology:
Learning, Memory, and Cognition
1995, Vol. 21, No. 5,1339-1348

0278-7393/95/$3.00

Attentional Limits in Memory Retrieval

L. Mark Carrier and Harold Pashler
University of California, San Diego

The hypothesis that episodic memory retrieval can occur in parallel with other cognitive processes was tested in 2 experiments. Participants memorized words and then performed speeded cued recall (Experiment 1) or speeded yes-no recognition (Experiment 2) in a dual-task situation. The psychological refractory period design was used: The participant was presented with a single test item at various stimulus onset asynchronies (SOAs; 50-1,200 ms) after a tone was presented in an auditory-manual 2-alternative choice reaction task. Reducing the SOA increased the memory task reaction times. This slowing was additive with the effect of variables slowing retrieval in the memory task. The results indicate that memory retrieval is delayed by central processes in the choice task, arguing that the central bottleneck responsible for dual-task interference encompasses memory retrieval as well as response selection.

A key question in the study of human cognition is which cognitive processes can operate in parallel with other ongoing cognitive processes, which operate serially, and which interact in more complex ways with other cognitive processes. The experiments presented in this article ask whether memory retrieval can occur in parallel with other cognitive processes. Previous work on this question has come from two separate lines of research. The first involves studies designed specifically to address this question. The second involves studies of processing limitations in the performance of simple tasks not involving memory retrieval. Interestingly, these two lines of research have led to opposite conclusions.

Dual-Task Studies of Memory Retrieval

Many investigators have required participants to perform a memory retrieval task while concurrently performing an unrelated cognitive task. The logic is simple: If memory retrieval cannot operate in parallel with other cognitive processes, then performance on one or both of the tasks will be worse in dual-task conditions than in single-task control conditions. Dual-task interference has been found in studies of free recall (Johnston, Greenberg, Fisher, & Martin, 1970; Johnston, Wagstaff, & Griffith, 1972; Macht & Buschke, 1983; Martin, 1970; Martin, Marston, & Kelly, 1973; Moscovitch, 1992; Park, Smith, Dudley, & Lafronza, 1989), serial recall (Johnston et al., 1970; Martin & Kelly, 1974; Trumbo & Milone, 1971), cued recall (Craik & McDowd, 1987; Park et al., 1989), and recognition (Craik & McDowd, 1987; Jacoby, 1991; Jacoby, Wolosyhn, & Kelley, 1989). These results might therefore be interpreted as indicating that memory retrieval cannot occur in parallel with other cognitive tasks.

However, this interpretation rests on the questionable assumption that there are no other confounding sources of interference between the two tasks in these studies. In fact, most of the studies required that participants make overt responses in one task while simultaneously making responses in the other task. Therefore, participants had to select and produce responses in one task while trying to select and produce responses in the other task. For this reason, it is necessary to rule out the confounding effects of interference between response-related processes to determine whether memory retrieval occurs in parallel with other cognitive processes.

Baddeley, Lewis, Eldridge, and Thomson (1984) compared the interference arising from a secondary task with the interference arising from a comparable motor control task, which contained the response requirements but not the cognitive requirements of the secondary task. If memory retrieval operates in parallel with other cognitive processes, then the secondary task should cause no more interference with retrieval than the control task. Across several experiments, the secondary task was either a card-sorting task or a digit-load task. In the control version of the card-sorting task, participants placed playing cards into four arbitrary stacks. In the regular version of the task, participants sorted cards into stacks on the basis of suit. In the control digit-load task, participants heard the numbers *one* through *six* and repeated them back at a prescribed rate. In the regular version, participants heard six random digits, then repeated them back. The results showed little or no decrease in free recall, cued recall, and recognition accuracy when performance in the control condition was compared with performance in the regular condition. The researchers therefore concluded that memory retrieval occurs in parallel with other cognitive processes. Primarily on the basis of these results, Moscovitch and Umilta (1990,1991) also

L. Mark Carrier and Harold Pashler, Department of Psychology, University of California, San Diego.

L. Mark Carrier was supported in part by National Institute of Mental Health Training Grant 5T32 MH14268. This research was supported by the Office of Naval Research under Contract N00014-88-K-0281. We thank Clark Fagot and Richard Schweickert for comments and advice and Renee Hwang and Krista McFarland for assistance in conducting the experiments.

Correspondence concerning this article should be addressed to L. Mark Carrier, who is now at Psychology Department 1051, Florida State University, Tallahassee, Florida 32306-1051. Electronic mail may be sent via Internet to mcarrier@psy.fsu.edu.

argued for the idea that memory retrieval operates in parallel with other mental processes.

The Psychological Refractory Period and the Response Selection Bottleneck

Dual-task interference between two relatively simple cognitive tasks (not involving memory retrieval) has been studied using the psychological refractory period (PRP) design. In each trial in a PRP experiment, the participant performs two discrete, speeded tasks in close temporal proximity. For example, the first task might be classifying a tone as high or low in pitch by making a manual response, and the second task might be saying aloud the name of a visually presented letter. The duration of the interval between the onset of the stimulus to the first task and the onset of the stimulus to the second task (stimulus onset asynchrony, or SOA) is varied, and the effects of this manipulation on Task 1 and Task 2 reaction times (RTs) are observed. Typically, reducing the SOA from several hundred milliseconds toward zero increases Task 2 RTs by several hundred milliseconds (Welford, 1952).

This result suggests the existence of a processing bottleneck (Figure 1), that is, a crucial mental process that can only be carried out in one task at a time. When the two stimuli are presented close in time, the crucial process in the second task must await the completion of the same process in the first task. Furthermore, when the SOA is reduced beyond some point, the crucial stage of processing in Task 2 cannot begin any earlier. Therefore, the bottleneck hypothesis predicts that the slope of the function relating Task 2 RT (RT2) to SOA will approach -1 at the smallest SOAs. This often occurs.

Research has been carried out to determine which stage of cognitive processing is responsible for the bottleneck. The choice RT tasks used in PRP experiments can be broken down into three distinct processing stages: perception, response selection (i.e., the determination of an appropriate response for a given stimulus), and response execution (Sternberg, 1969). It is possible to determine which of these stages is included in the bottleneck by manipulating factors that vary the duration of particular stages of processing in Task 2 (cf. Keele, 1973; Schweickert, 1978). Two straightforward predictions can be made in this situation (see Figure 1). First, if one varies the duration of a stage that is at or after the bottleneck, then the effect on RT2 should be additive with the effect of SOA. Second, if one varies the duration of a stage that precedes the bottleneck, then the effect on RT2 should be smaller at the short SOAs than at the long SOAs, that is, there should be an underadditive interaction with the effect of SOA.[1]

Results of such manipulations have favored the response selection bottleneck hypothesis, which was proposed by Welford (1952, 1980). This hypothesis predicts that manipulating the duration of perceptual processes in Task 2 that precede response selection (in Task 2) will lead to an underadditive interaction with SOA on RT2. This has in fact been observed when the intensity of visual stimuli in a second task has been varied (Pashler, 1984; Pashler & Johnston, 1989). However, manipulations presumed to affect response selection in the second task should produce additive interactions with the effect of SOA. This has been found with variables such as decision outcome in a visual search task (Pashler, 1984), intertrial repetition of a stimulus (Pashler & Johnston, 1989), stimulus-response compatibility (McCann & Johnston, 1992), and the Stroop effect (Fagot & Pashler, 1992).

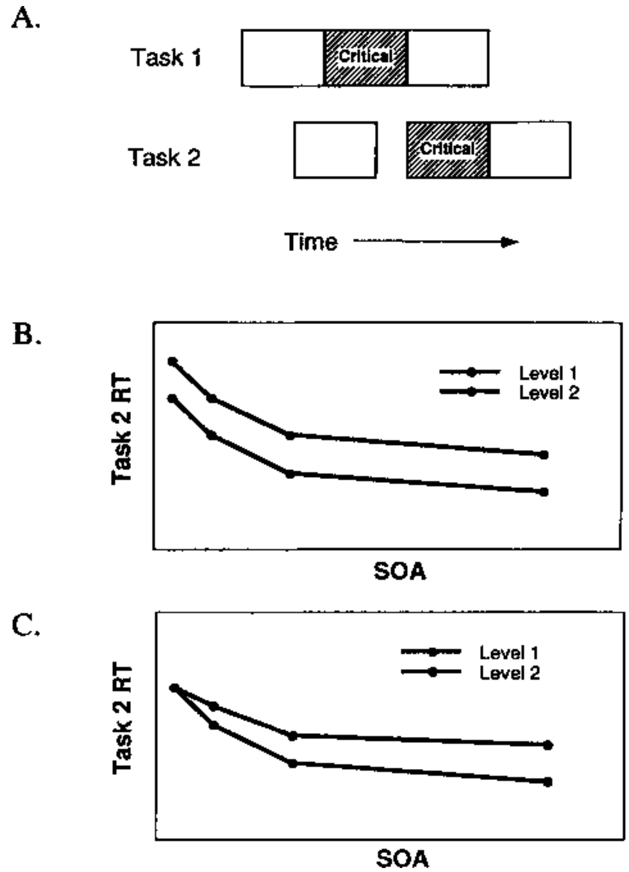

Figure 1. Panel A depicts a bottleneck model of information processing in a psychological refractory period (PRP) experiment. Panels B and C depict different predictions made within a PRP design. (b) When manipulating a stage of processing in the second task that occurs at or after the bottleneck, the manipulation interacts additively with the effect of stimulus onset asynchrony (SOA). (c) When manipulating a stage of processing that occurs before the bottleneck, an underadditive interaction is expected. RT = reaction time.

Studies of choice RT tasks thus favor the hypothesis that response selection in one task cannot occur in parallel with

[1] One does not have to assume that processing stages operate strictly in serial in order for this method to be useful. One could, for example, envision a cascade model in which each processing stage is continuously receiving information from the preceding stage and outputting information to the following stage (McClelland, 1979). If both tasks required the same process, and that process could not operate on a representation in Task 2 while operating on a representation in Task 1, then a processing bottleneck might arise. If the rate of processing of a Task 2 stage preceding the bottleneck stage were to be manipulated, then the output of that stage would reach asymptotic strength regardless of the level of the factor, and thus the factor would not affect RT2. However, manipulating a processing stage that comes at or after the bottleneck stage would have a measurable effect on RT2 (Pashler, 1994b).

response selection in a concurrently performed task. Response selection by definition involves the looking up of information (the appropriate response) associated with the stimulus (e.g., Sternberg, 1969). Therefore, response selection might be seen as the retrieval of a response code or response program. This raises the possibility that retrieval and response selection rely on the same cognitive mechanism and are subject to the same processing bottleneck. This will be termed the *response-selection / retrieval bottleneck hypothesis.* Alternatively, the bottleneck may not involve memory retrieval generally, but only retrieval of specifications of action. This hypothesis will be termed *the parallel retrieval hypothesis.*

The Present Approach

Baddeley et al. (1984) concluded that memory retrieval can occur in parallel with other cognitive processes. However, the evidence from the PRP studies suggests that memory retrieval cannot occur in parallel with response selection. It is interesting that the experimental design used by Baddeley et al. may have precluded the detection of interference between memory retrieval and response selection. In those experiments, the cognitive difficulty of the secondary tasks was varied, but the response demands of the tasks were kept constant. Conceptually, this tested whether concurrent cognitive processes other than the ones involved in responding interfere with memory retrieval. Our experiments were specifically designed to determine whether memory retrieval is delayed by a concurrently performed response selection.

The experiments reported here used the PRP design, with a choice reaction task as the first task and a cued memory retrieval task as the second task. We presumed that these tasks can be theoretically decomposed into constituent stages of information processing (see Figure 2). The choice reaction task is commonly divided into three stages: perception, response selection, and response execution. The memory retrieval task may be comprised of these basic stages plus an additional memory retrieval stage inserted between perception and response selection.

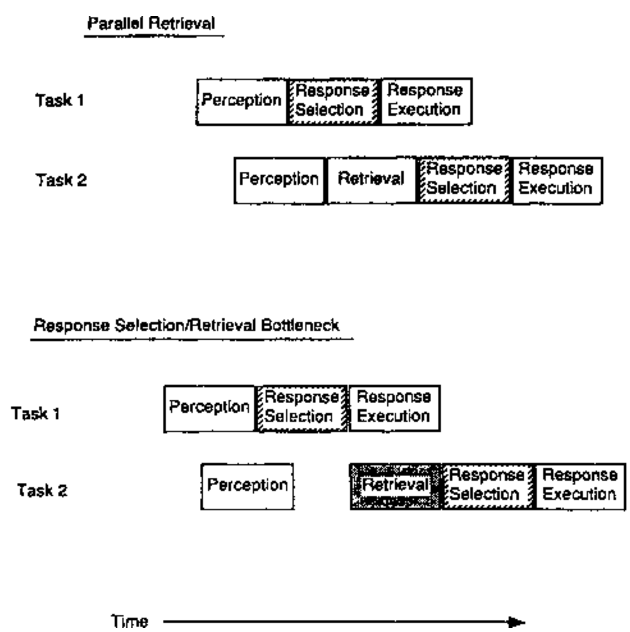

Figure 2. Decomposition of the tasks in Experiments 1 and 2 into stages of processing.

We varied the duration of memory retrieval in the second task to determine whether memory retrieval in Task 2 occurs in parallel with processes in Task 1. If retrieval occurs in parallel with other cognitive processes, then one of two things must happen. If response selection in Task 1 and response selection in the retrieval task do not temporally overlap (i.e., at long SOAs), then there should be no slowing of the retrieval task. Alternatively, if response selection in Task 1 and response selection in Task 2 do abut each other, then there should be an underadditive interaction of the effects of the retrieval manipulation and of SOA. In contrast, the response-selection/ retrieval bottleneck hypothesis predicts that when Task 1 response selection and Task 2 memory retrieval do not abut each other (i.e., at long SOAs), there should be no slowing of the retrieval task. When Task 1 response selection and Task 2 memory retrieval do abut each other at the shorter SOAs, there should be slowing of the retrieval task. Furthermore, there should be an additive interaction of SOA and the retrieval manipulation.

Experiment 1: Cued Recall

We tested whether the retrieval component of cued recall occurs in parallel with other cognitive processes. The first task was a two-alternative choice reaction task with an auditory stimulus and a manual response. The second task was the cued recall of a paired associate. The participants had previously studied the set of paired associates. The duration of retrieval in cued recall was manipulated by having each target pair studied and tested twice during the course of the experiment. We presumed that it should take longer to respond the first time we tested the pair than the second time. Henceforth, we refer to the first study and test of a pair as the first-presentation condition and to the second study and test as the second-presentation condition.

Method

Participants. The participants were 30 undergraduates at the University of California, San Diego, who were fulfilling part of a course requirement.

Materials and apparatus. All stimuli were presented using IBM personal computers. Vocal RTs were measured using a Gerbrands Model G1341T Voice-Activated Relay (Arlington, MA) connected to the personal computer.

The paired associates were word pairs. Seventy-eight nouns were randomly selected from Kučera and Francis's (1967) database, with the restrictions that the nouns had to be from four to seven letters in length (inclusive) and had to have frequencies of occurrence greater than 100 per million. Sixty of the nouns were randomly paired to form the target pairs, and 18 were used to form the pairs for the distractors (see below). The same pairings were used for all participants. The word pairs used in the practice were 38 nouns chosen in the same manner as the experimental words except the frequency of occurrence was restricted to ≥ 50 and < 100 occurrences per million. Thirty of the practice words were randomly paired to form the target pairs (for practice); the remaining words were randomly paired to form the distractor pairs. Words were presented visually, in the center of the computer screen (white letters on a black background). The height of the words subtended a visual angle of 1.34° at a viewing distance of 60cm.

Design and procedure. Two within-subjects variables (SOAs from the tone to the word of 50, 250, and 1,200 ms and the presentation condition of first presentation vs. second presentation) and one between-subjects counterbalancing variable (assignment of word pairs to SOAs) comprised the design. There were three different assignments of word pairs to SOAs such that across subjects all pairs contributed equally often to each of the SOAs.

For each participant, there were 30 first-presentation trials and 30 two-presentation trials. Ten of the first-presentation trials were at the 50-ms SOA, 10 were at the 250-ms SOA, and 10 were at the 1,200-ms SOA. The same held for the second-presentation trials. Word pairs were assigned to the same SOA for the first and second presentations. The assignment of SOAs to trials was designed to minimize any systematic influence of the position of the trial in the overall experiment on retrieval RTs. Six random SOAs were assigned in random order to the first block of trials. For Blocks 2-9, three of the SOAs from the previous block (from first-presentation trials) and three new randomly selected SOAs comprised each block's set of trials. On the final block, three of the SOAs were the three first-presentation SOAs used in the previous block, and the other three SOAs were the remaining first-presentation SOAs used in the first block of trials. For each participant, word pairs were assigned one at a time to the first unassigned first-presentation trial at the appropriate SOA for the word pair set and then to the first unassigned second-presentation trial at the appropriate SOA contained in the remaining blocks. Thus, for a word pair first presented in Block i, the second presentation of the word pair could appear anywhere in Blocks $i + 1$ to 10.

In the experimental session, participants received written instructions regarding their tasks, then practiced all types of conditions that appeared in the experiment. Next, the experimenter answered any questions and then administered the experiment. The instructions to the participant emphasized the importance of responding as quickly and as accurately as possible to each of the two speeded tasks. The instructions placed no special emphasis on one of the tasks over the other and did not say which response should be made first.

Five blocks of six trials each comprised the practice. The main experiment consisted of 10 blocks of six PRP trials each. At the beginning of each block, the participants studied seven word pairs for testing. Six of the word pairs were tested in immediate PRP trials; the remaining word pair served only as a distractor during study (to prevent the participants from knowing which pair would appear in the last PRP trial in a block). The study phase of each block went as follows: First, all seven word pairs appeared together on the screen (in random order) for 70 s. Then, the participant was given an informal test over each pair. In the informal test, the stimulus member of a pair would appear on the screen by itself for 5 s. The participant was told to try to remember the response member of the pair during this time period. Then, both the stimulus and the response member of the pair would appear on the screen for 5 s. The presentation order of informal tests of pairs was randomized. Finally, all of the word pairs together were presented on the screen again for 60 s.

Following a brief warning message to prepare for the speeded trials, the PRP trials began. In all blocks except the first and last, the set of trials included three first-presentation trials and three second-presentation trials. Only first-presentation trials comprised the first block and only second-presentation trials comprised the last block.

The timing of an individual PRP trial was as follows: At the beginning of the trial, a fixation point (a plus sign, 0.95° x 0.95° of visual angle, based on a 60-cm viewing distance) appeared in the center of the screen for 1,500 ms. Fifteen-hundred milliseconds after the offset of the fixation point, the first stimulus occurred (500 Hz for low-pitched tones; 1,000 Hz for high-pitched tones) and lasted 200 ms. The participant responded by using his or her right middle and index fingers to press the "HI" key on the computer keyboard (the semicolon key) or the "LO" key on the computer keyboard (the period key), respectively. At varying intervals after the onset of the tone (the SOA), the stimulus word was presented on the screen. The participant responded by saying the appropriate response into a microphone attached to the voice trigger. Manual RTs were measured from the onset of the tone to the keypress response. Vocal RTs were measured from the onset of the word to the first detection of a vocal response by the voice trigger. Two seconds after the second (verbal) response was made, the next trial began.

Results

Data from trials in which no cued recall response was made or in which either RT was less than 200 ms or greater than 10,000 ms were not included in the analyses. These criteria excluded data from 45 out of 1,800 trials (3%). Both correct and incorrect trials were included in the RT analyses.

First task (tones). Figure 3 shows the mean RTs in the choice reaction task. A 2 (presentation condition) x 3 (SOA) analysis of variance (ANOVA) was performed on the RT data. There was no reliable main effect of presentation condition, $F(1, 29) = 0.16$, but there was a reliable main effect of SOA, $F(2, 58) = 9.02$, $p < .001$, $MSE = 9{,}008$, consistent with the small amount of apparent slowing as SOA was reduced. The interaction between presentation condition and SOA was not reliable, $F(2, 58) = 1.19$, $p = .31$, $MSE = 6{,}440$. Accuracy in this task was very high. The overall mean accuracy was 99%. An ANOVA performed on the accuracy data indicated that neither the main effect of number of tests, $F(1, 29) = 2.07$, $p = .16$, $MSE = .002$, the main effect of SOA, $F(2, 58) = 0.27$, nor the interaction between the two, $F(2, 58) = .38$, were reliable.

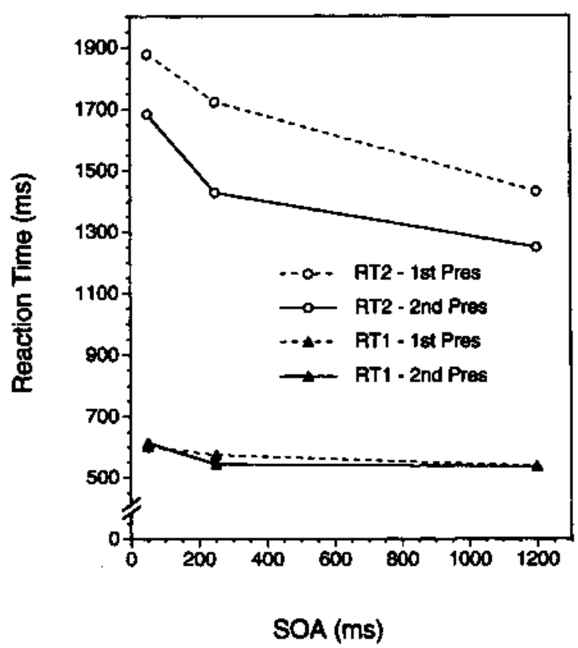

Figure 3. Mean reaction times (RTs) from Experiment 1. Pres = presentation; SOA = stimulus onset asynchrony.

Second task (cued recall). Figure 3 also shows the mean RTs in the cued recall task. An ANOVA performed on the RT data indicated that the main effect of presentation condition, $F(1, 29) = 19.9, p < .001$, $MSE = 112{,}831$, and the main effect of SOA, $F(2, 58) = 30.9$, $p < .001$, $MSE = 95{,}048$, were reliable. The interaction between the two was not reliable, however, $F(2, 58) = 0.87$. Accuracy in this task was very high. The overall mean accuracy was 98%. An ANOVA performed on the accuracy data indicated that neither main effects, $F(1, 29) = 0.70$, presentation condition; $F(2, 58) = 0.74$, SOA; nor the interaction, $F(2,58) = 0.89$, were reliable.

Discussion

Visual inspection of the results reveals possible additivity of the retrieval manipulation and SOA (Figure 3). Indeed, there was no reliable interaction between these two variables (see above). Furthermore, RT2s were approximately 400-ms slower at the shortest SOA than at the longest. This slowing would be due to the postponement of response selection in Task 2 according to the parallel hypothesis. The postponement would have resulted in a 400-ms gap in processing time in Task 2 at the shortest SOA during which no processing would be occurring in Task 2. If so, then the 200-300 ms effect of the number of presentations (which affects the retrieval stage) on RT2 should virtually have been eliminated at the shortest SOA. However, a post hoc analysis of the effect of presentations at the 50-ms SOA showed a reliable difference, $f(29) = 2.31$, $p < .05$. This additivity is consistent with the response selection-retrieval bottleneck hypothesis but is not consistent with the hypothesis that memory retrieval occurs in parallel with processing in Task 1.

An analysis of the pattern of slowing of the responses in Task 2 (R2s) reveals more evidence supporting the response selection-retrieval bottleneck hypothesis. According to the parallel retrieval hypothesis, R2 slowing with reduced SOA (Figure 3) comes about from response selection interference. However, even at the shortest SOAs, response selection interference is probably not involved in the slowing (see Figure 4). The data from the 1,200-ms SOA condition (first presentation) suggests that it would have taken 537 ms to perform the tone task alone and 1,431 ms to perform the cued recall task alone. It takes approximately 500 ms to read aloud a word (Balota, Black, & Cheney, 1992; Monsell, Doyle, & Haggard, 1989), and reading aloud a word probably contains the same response processing as the cued recall task.[2] Therefore, response selection and response execution are contained in the last 500 ms of the 1,431 ms of processing in the cued recall task. When the two tasks overlapped at the 50-ms SOA, Task 1 should have overlapped with only the first [537 - 50 =] 487 ms of Task 2 processing, leaving the remaining [1,431 - 487 =] 944 ms of Task 2 processing interference free. Assuming the response processing in the cued recall task is contained within the last 500 ms of processing in cued recall, cued-recall response processing should essentially be free from overlap with response processing in Task 1. Yet, in the data there were large amounts of R2 slowing at this shortest SOA, suggesting there was interference between response selection in Task 1 and memory retrieval in Task 2. The same was true for the second-presentation condition. The response selection-retrieval bottleneck hypothesis fits the data well. If most of the latency in the cued recall task is consumed by the memory retrieval, the first task could overlap with just the initial processing in the cued recall task and cause the observed slowing of the cued recall RTs.

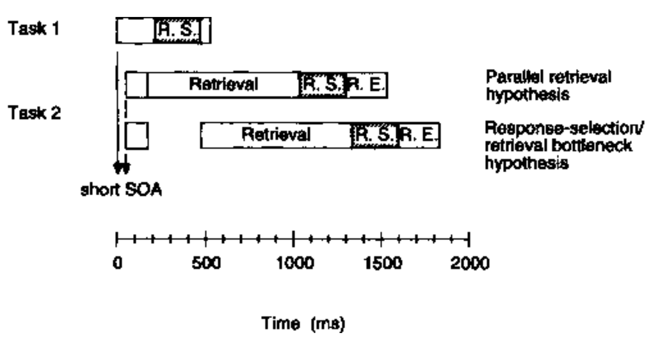

Figure 4. The response-selection/retrieval bottleneck hypothesis predicts responses in Task 2 (R2) slowing when Task 1 response selection overlaps with Task 2 retrieval. However, the parallel retrieval hypothesis predicts R2 slowing only when Task 1 response selection overlaps with Task 2 response selection. R. S. = response selection; R. E. = response execution; SOA = stimulus onset asynchrony.

Neither hypothesis predicted the 50-ms slowing of the response in Task 1 (R1) as the SOA was reduced (Figure 3). A tentative post hoc account of this slowing is that participants selectively used a *conjoint responding* strategy, in which R1 was delayed until after R2 was selected so that R1 and R2 were emitted together (Borger, 1963; Pashler & Johnston, 1989). At the short SOAs, participants knew that R2 would be ready soon and therefore engaged in conjoint responding. At the long SOA, participants knew that R2 would not be ready for a while, so they emitted R1 without delay. One speculative explanation for why participants engaged in conjoint responding in this experiment is that the instructions placed equal emphasis on the two tasks. In previous PRP studies, participants have usually been instructed to give special emphasis to the speed of R1, and no Task 1 RT (RT1) effects have been obtained. If participants had been given the standard instructions, RT1 might not have shown these effects. However, there is no reason to believe that the RT2 effects would have been much different.

In summary, these results support the response selection-retrieval bottleneck hypothesis and do not support the hypothesis that memory retrieval occurs in parallel with other cognitive processes.

Experiment 2: Recognition

In Experiment 2, we tested whether the process of recognition is subject to postponement from a concurrently performed choice reaction task. Task 1 was the choice reaction task used

[2] Some readers may find this assumption questionable. There is little direct evidence available on the duration of response selection and response execution in tasks similar to the cued recall task. However, if one relaxes this assumption by allowing that the critical duration may be up to 900 ms, the argument still holds.

in Experiment 1, whereas Task 2 was a speeded yes-no word recognition task. We conceptually divided the recognition task into four stages: perception, recognition, response selection, and response execution. We manipulated the duration of the recognition stage by varying the number of presentations of a word that had been made prior to the word being tested. A target word was presented either one time (the old1 condition) or five times (the old5 condition) before being tested. Presumably, five presentations of a word would lead to a shorter duration of the recognition stage than would one presentation of a word.

Method

Participants. The participants were 30 students enrolled in undergraduate courses at the University of California, San Diego. They participated as part of a course requirement.

Materials and apparatus. The words used in the recognition task were 240 nouns from Kučera and Francis's (1967) database. We randomly selected 240 nouns with lengths from four to eight letters inclusive and frequencies per million from 50 to 300 inclusive. The words used during practice were 20 words chosen from the same set. The word lists were prerecorded on audio tape and presented using a tape player. The PRP trials were run on IBM PCs.

Design. The main design was a 2 (old1 vs. old5) x 3 (SOA) factorial. In addition, there were the same number of (recognition) distractor PRP trials as there were old1 and old5 trials combined.

Procedure. Each participant received written instructions regarding the experiment. The instructions emphasized that the participants should make all responses as quickly and as accurately as possible, but did not say whether one response should be made before the other. Also, the participants were told to do as well as possible on the recognition task. Following the presentation of the instructions, each participant completed one practice list.

During the experiment, the participants attempted to memorize and were tested on six different word lists. The presentation of each word list was followed by the performance of a brief interpolated task, which was followed by the performance of a series of PRP trials.

Details of the presentation phase are as follows. The experimenter played the tape containing the word lists. Each word list was constructed to ensure that the average retention intervals for old1 and for old5 items would be equal. Each word list contained 20 different items. The 10 old1 items were presented once and the 10 old5 items were presented five times each. The word list was divided into five parts. In each of the first four parts, the complete set of 10 old5 words was presented in a random order. In the last part of the word list, the complete set of 10 old1 items was mixed randomly with the old5 items. The assignment of specific words to the old1 and old5 conditions was counterbalanced such that across all subjects each word contributed equally often to both conditions. At the end of the taped word list, the participant engaged in a 2 min 30 s arithmetic task. In this task, the participant solved as many written multiplication problems with two- and three-digit operands as possible in the time allotted.

At the end of the arithmetic task, the participant began 40 PRP trials based on the preceding word list. Twenty of the trials involved the speeded recognition of old items (items presented in the word list) and the remaining 20 involved foils (words never presented). Half of the old items were old1 items and the other half were old5 items. Each PRP trial began with the presentation of either a low-pitched (500 Hz) or a high-pitched (1,000 Hz) tone, randomly selected. The tone lasted 200 ms. The participant responded to the tone by pressing the "z" key on the computer keyboard for a low-pitched tone or the "a" key for a high-pitched tone. The test word appeared after an SOA of 50,150, or 1,100 ms from the onset of the tone. The order of SOAs within a block was random, with the constraint that across all blocks there were equal numbers of old1 and old5 items at each SOA. Furthermore, there were equal numbers of old and new items at each SOA. The word remained on the screen until the participant made a response. The participant indicated a yes response to the word by pressing the single-quote key and a no response by pressing the forward slash key. The assignment of words to SOAs was counterbalanced such that across all subjects each specific word contributed equally often to each SOA condition. In a given word list, the order of PRP trials was randomized. At the end of a block of PRP trials, the computer informed the participant of the proportion of correct recognition responses. A rest break was given before the next word list started.

Results

Data from trials in which either RT was less than 200 ms or greater than 10,000 ms were not included in the analyses. These criteria excluded data from 12 out of 7,200 trials (0.2%). Both correct and incorrect trials were included in the RT analyses.

First task (tones). The mean RTs in the tone task are displayed in Figure 5. For purposes of analysis, the data were divided according to the type of item on the recognition test (new items, old items presented once, and old items presented five times). A 3 (SOA) x 3 (test item type) ANOVA was performed on the RT data. The results indicate that both the main effect of test item type, $F(2,58) = 6.82$, $p < .005$, $MSE = 3,957$, and the main effect of SOA, $F(2, 58) = 16.0$, $p < .001$, $MSE = 42,534$, were reliable. We consider the significance of these two effects in the *Discussion* section below. The interaction between these two effects was not reliable, $F(4, 116) = 1.64$, $p = .17$, $MSE = 3,584$. Accuracy in this task overall was high. The mean proportion of correct trials was .98 across all conditions. According to an ANOVA performed on the accuracy data, the main effect of test item type was not reliable, $F(2, 58) = 1.01$, $p = .37$, $MSE = .0009$, whereas the main effect of SOA was reliable, $F(2, 58) = 4.65$, $p < .05$, $MSE = .002$. The mean proportions correct across SOAs and

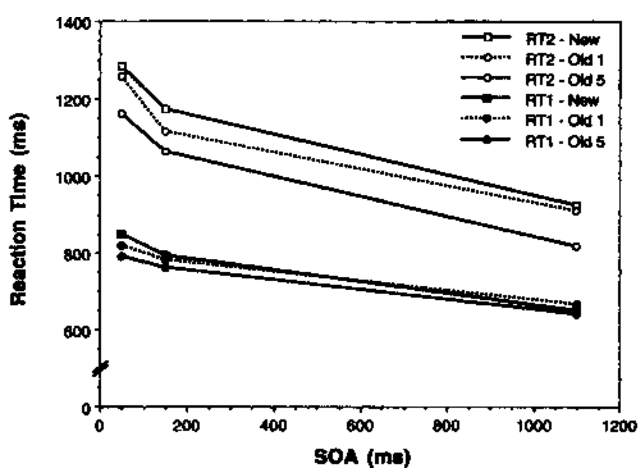

Figure 5. Mean reaction times (RTs) from Experiment 2. "Old1" refers to items that were presented one time; "Old5" refers to items that were presented five times. SOA = stimulus onset asynchrony.

collapsed across test item type were .97, .98, and .99 in the 50-, 150-, and 1,100-ms conditions, respectively. The interaction between the two effects was not reliable, $F(4,116) = 0.18$.

Second task (recognition). The mean RTs to perform the recognition task are shown graphically in Figure 5. An ANOVA performed on the RT data indicated that the main effects of test item type, $F(2, 58) = 19.6$, $p < .001$, $MSE = 15{,}962$, and SOA, $F(2, 58) = 134$, $p < .001$, $MSE = 21{,}392$, were both reliable. It apparently took longer to respond to a distractor item (the new targets) than it did to respond to the old items. Furthermore, as expected, there was massive slowing of the RTs as the SOA was reduced. The interaction between the two main effects was not reliable, $F(4, 116) = .81$. An ANOVA performed on just the data from the trials with old items indicated a reliable main effect of number of presentations, $F(1, 29) = 28.8$, $p < .001$, $MSE = 10{,}456$, and a reliable main effect of SOA, $F(2, 58) = 106.4$, $p < .001$, $MSE = 17{,}406$, but no reliable interaction between the two effects, $F(2, 58) = 1.32$, $p = .28$, $MSE = 6{,}596$.

The mean proportions correct in the recognition task collapsed across SOA were .89, .78, and .91 for the new items, old1 items, and old5 items, respectively. An ANOVA performed on the accuracy data indicated that the main effect of test item type was reliable, $F(2, 58) = 30.6$, $p < .001$, $MSE = 0.014$. However, the main effect of SOA was not reliable, $F(2, 58) = 0.41$. The interaction between the two main effects was not reliable, $F(4,116) = 0.37$.

Discussion

For the recognition RTs, the interaction between SOA and the retrieval manipulation was not reliable, suggesting an additive effect of these two factors. However, visual inspection of the results (see Figure 5) suggests some deviation from pure additivity. To clear up the discrepancy, we performed a post hoc comparison of the old1 and old5 conditions at the shortest SOA. According to the hypothesis that recognition-retrieval occurs in parallel with response selection in Task 1, the R2 slowing with reduced SOA indicates postponement of response-selection processing in Task 2. With the large amount of observed slowing, there should have been enough time for recognition in Task 2 to occur completely in parallel with Task 1 processes at the shortest SOA (refer to Figure 2) Therefore, the retrieval manipulation effect should have been eliminated at the shortest SOA. However, there was a reliable effect of the retrieval manipulation at this SOA, $t(29) = 3.69$, $p < .005$. This result is not consistent with the parallel retrieval hypothesis, but would be expected from the response selection-retrieval bottleneck hypothesis.

There were three reliable effects on RT1: (a) RT1 increased as the SOA was reduced, (b) The retrieval manipulation affected RT1, and (c) Task 1 accuracy decreased as the SOA was reduced. As in Experiment 1, a tentative post hoc account of the first effect is that participants engaged in conjoint responding when the SOA was small. Conjoint responding could conceivably account for the other two effects as well. If R1 is sometimes delayed until R2 is ready, then RT1 should be affected by Task 2 manipulations. Also, delaying R1 may sometimes cause the correct response to be forgotten, thus slightly increasing the error rate. However, less than half of the participants showing R1 slowing also showed the patterns of interresponse intervals (IRIs) expected from conjoint responding, namely an increasing number of short IRIs (we arbitrarily selected less than 200 ms as the criterion) at the short SOAs. Therefore, conjoint responding did not account for all participants' patterns of RT1 effects. Nevertheless, using standard PRP instructions with speed emphasis on R1 might have reduced (but perhaps not eliminated) the RT1 effects without affecting the RT2 effects.

It may be argued that some of the dual-task interference in this experiment was due to competition between responses within the manual modality. (This was not a problem in Experiment 1 because two different response modalities were used there.) This argument is supported by the fact that Task 1 responses in this experiment were slower than in Experiment 1 and that there was a larger effect of SOA on RT1s in this experiment than in the last one. However, response interference would most naturally have arisen during response selection, response execution, or both. If so, and if the parallel retrieval model was correct, then underadditive effects should have been obtained. Furthermore, to account for the pattern of R2 slowing, interference effects would have to have arisen in the early processing of Stimulus 2, which seems unlikely if the interference was due to some kind of response competition.

General Discussion

The slowing of a memory retrieval task by a concurrently performed task was not solely due to interference between response-related processes (response selection and response execution) in the two tasks. Experiment 1 showed considerable slowing of RTs in a cued recall task even when a concurrent task was unlikely to have overlapped with response-related processing in the cued recall task. Similar results were found for speeded yes-no recognition in Experiment 2. Furthermore, in both experiments, the effect of varying the duration of memory retrieval was additive with the effect of SOA. These results argue that memory retrieval was postponed as a result of the overlapping unrelated task.

Capacity Sharing: An Alternative to the Bottleneck?

Up to this point, we have suggested that memory retrieval was postponed by response selection in Task 1. However, there are other ways in which memory retrieval and response selection could have interfered with each other. One common explanation of dual-task interference is that processing stages might simultaneously draw mental energy or capacity from a limited pool of capacity (cf. Kahneman, 1973). Memory retrieval and response selection would operate in parallel, but each would take longer to complete. This could conceivably account for both Task 1 and Task 2 slowing in Experiments 1 and 2, as well as the additivity of the retrieval manipulations and SOA.

However, the interference between response selection and memory retrieval observed here probably arose from the same causes as response-selection interference observed in PRP experiments using tasks that did not involve memory retrieval. In the latter, there are two pieces of evidence that favor postponement over capacity sharing. First, Pashler (1994a) asked participants to place equal emphasis on two simple sensorimotor tasks in a PRP experiment. The distributions of IRIs from a condition with a 0-ms SOA mostly fell into one of two types. Participants who apparently grouped their responses showed a narrow-peaked distribution centered around 0 ms. In contrast, a second group of participants showed a bimodal distribution, indicating that one response virtually always came several hundred milliseconds before the other, consistent with a bottleneck model of interference and questioning the possibility of capacity sharing. Second, the bottleneck model predicts that at long SOAs, when there is no queuing, the speed of R1 should not influence the speed of R2. In contrast, at short SOAs, when there is queuing, the speed of R1 should influence the speed of R2. One way to evaluate this prediction is to divide all RT1s into bins and then for each bin to compute the mean RT2 for that subset of trials. At the short SOAs, there should be a stronger dependence of RT2 on RT1 than at the long SOAs. Indeed, this pattern of results has been demonstrated in recent PRP studies (e.g., Pashler, 1989).

Reevaluating Data Supporting the Parallel Retrieval Hypothesis

How does the memory retrieval-response selection bottleneck hypothesis account for the results supporting the parallel retrieval hypothesis? The main evidence supporting the parallel retrieval hypothesis was the finding that retrieval accuracy was unaffected by increasing the complexity of a concurrently performed task (Baddeley et al., 1984). One alternative interpretation is that participants were not performing the tasks in parallel but were rapidly switching between the tasks. When there was a gap in processing in one task, processing in the other task could occur. This strategy might have disrupted the timing of memory retrievals, but not necessarily their accuracies. Some data from the Baddeley et al. study are consistent with this hypothesis. When the experimenters measured recognition RTs, increasing secondary task difficulty significantly increased recognition latencies.

A second alternative explanation of Baddeley et al.'s (1984) data is that load complexity during retrieval had little effect because recall performance was near maximum. For example, in their Experiment 1 participants first studied words with either a low or a high concurrent card-sorting load and then freely recalled the words with either load. In the low-encoding-load/high-retrieval-load condition, participants recalled on average only 4.6 words from a 12-word list. Because the words had been studied in a dual-task situation, and because the participants in this experiment—as in all of the experiments in their study—were relatively old (mean age = 47 years), 4.6 may have been close to the number of words that were stored in memory in these conditions and thus close to the maximum recall score. It would not be surprising, then, that easing the retrieval load by moving to the low-encoding-load/low-retrieval-load condition should produce no statistically reliable improvement in the number of words recalled.

Other Accounts of Attentional Limits in Retrieval

Jacoby (1991) suggested that there are two kinds of recognition: one that is based on familiarity, which is automatic and not subject to dual-task interference, and one that is based on conscious recognition, which is subject to interference. In support of this distinction, he reported an experiment in which participants either read words aloud or generated words as solutions to anagram puzzles (e.g., *yodrw* -» *dowry).* Reading words aloud was presumed to contribute to the familiarity of the items, whereas generating words was presumed to allow for elaborate retrieval of the item. Later, participants performed a word recognition task while concurrently performing an auditory number-monitoring task. Dual-task conditions during recognition reduced recognition accuracy of words encoded by means of the anagram task but had no effect on accuracy of words encoded by means of the naming task.

Within Jacoby's (1991) framework, our results would imply that consciously-based recognition is postponed by processing in an overlapping task. However, a possible alternative explanation of Jacoby's original data may be that familiarity-based recognition is faster than recognition based on relatively elaborate retrieval processes. If so, then the recognition of the named items might have been more amenable to a task-switching strategy than would the recognition of the generated items. Faster recognition processes would be more likely to fit into the temporal gaps that occur during processing in Jacoby's dual-task conditions. Discriminating between these two possibilities will require further research.

Craik (1983) made the interesting suggestion that retrieval cannot occur in parallel with other cognitive processes whenever it requires the use of self-initiated retrieval cues. Self-initiated cues are those that must be generated by the participant. Recognition relies on few such cues because the stimulus is presented in full. Free recall, however, depends on many self-initiated cues because the participant must mentally reconstruct the original encoding experience. Tasks that rely on few self-initiated cues should be unaffected in dual-task conditions, whereas tasks that rely on more cues should suffer from interference. Craik and his colleagues (cited in Craik, 1983) reported results that are consistent with this prediction. In one experiment, participants performed a cued recall test while concurrently performing a card-sorting task. As expected, cued-recall accuracy was reduced in the dual-task condition when compared with a single-task control. In another experiment, participants performed a recognition test while performing the card-sorting task. In this case, recognition accuracy was unaffected by the dual-task manipulation. Despite these results, Craik's hypothesis has difficulty accounting for Jacoby's (1991; Jacoby et al., 1989) data, which clearly provide cases of dual-task interference on recognition accuracy. In addition, Craik's hypothesis cannot account for our results because they show that recognition processes are postponed by processes in an unrelated nonretrieval task.

The Generality of Retrieval Interference

In implicit memory tasks, it is not necessary for the participant to refer to the previous encoding episode for memory to be evident. In explicit tasks, in contrast, the participant is specifically asked to remember a prior episode (e.g., recall and recognition; Graf & Schacter, 1985). Implicit retrieval may be able to occur in parallel with other cognitive processes, even if explicit retrieval cannot. A study by Jacoby et al. (1989) seems to favor this hypothesis. They had participants read names, and later gave two kinds of tests of the names. One test was a yes-no recognition test and the other was a fame judgment test. In the fame judgment test, participants made fame judgments (famous-nonfamous) on previously read names and on new names. Memory on this test was evident when, by virtue of their previous presentation, nonfamous names were judged to be famous. Though recognition performance was impaired when participants concurrently performed an auditory number-monitoring task, memory as measured in the fame judgment task was unaffected. Though these data are suggestive, they too are susceptible to an alternative account: that implicit retrieval is faster than standard memory retrieval and therefore more amenable to being "spliced" into the other task with task switching. Obviously, further investigation of implicit tasks would be warranted; the PRP design should make it possible to get around the task-switching problem.

Conclusions

Our data show that memory retrieval was postponed by response selection occurring in a concurrently performed choice RT task. This finding is consistent with a response selection-memory retrieval bottleneck hypothesis. The bottleneck may arise because memory retrieval and response selection require the same mental mechanism or because these operations, though they involve different machinery, inhibit each other. Regardless, the response selection-retrieval bottleneck hypothesis leads to other general questions. First, are there kinds of memory retrieval that are not part of this central bottleneck? Implicit memory retrieval may be one such kind of retrieval. Second, is it the case that two or more memory retrievals cannot be performed simultaneously, as the response selection-memory retrieval hypothesis predicts? In various experimental situations, such as during free recall, one might assume that more than one item is being retrieved at a given time. This possibility deserves further empirical examination.

References

Baddeley, A., Lewis, V., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General, 113,* 518-540.

Balota, D. A., Black, S. R., & Cheney, M. (1992). Automatic and attentional priming in young and older adults: Reevaluation of the two-process model. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 485-502.

Borger, R. (1963). The refractory period and serial choice-reactions. *Quarterly Journal of Experimental Psychology, 15,* 1-12.

Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society of London, B, 302,* 341-359.

Craik, F. I. M., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 474-479.

Fagot, C., & Pashler, H. (1992). Making two responses to a single object: Exploring the central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1058-1079.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 501-518.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513-541.

Jacoby, L. L., Woloshyn, V., & Kelley, C. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General, 118,* 115-125.

Johnston, W. A., Greenberg, S. N., Fisher, R. P., & Martin, D. W. (1970). Divided attention: A vehicle for monitoring memory processes. *Journal of Experimental Psychology, 83,* 164-171.

Johnston, W. A., Wagstaff, R. R., & Griffith, D. (1972). Information-processing analysis of verbal learning. *Journal of Experimental Psychology, 96,* 307-314.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice-Hall.

Keele, S. W. (1973). *Attention and human performance.* Pacific Palisades, CA: Goodyear.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Macht, M. L., & Buschke, H. (1983). Age differences in cognitive effort in recall. *Journal of Gerontology, 38,* 695-700.

Martin, D. W. (1970). Residual processing capacity during verbal organization in memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 391-397.

Martin, D. W., & Kelly, R. T. (1974). Secondary task performance during directed forgetting. *Journal of Experimental Psychology, 103,* 1074-1079.

Martin, D. W., Marston, P. T., & Kelly, R. T. (1973). Measurement of organizational processes within memory stages. *Journal of Experimental Psychology, 98,* 387-395.

McCann, R. S., & Johnston, J. C. (1992). Locus of the single-channel bottleneck in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 471-484.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287-330.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General, 118,* 43-71.

Moscovitch, M. (1992). A neuropsychological model of memory and consciousness. In L. R. Squire & N. Butters (Eds.), *Neuropsychology of memory* (pp. 5-22). New York: Guilford Press.

Moscovitch, M., & Umilta, C. (1990). Modularity and neuropsychology: Modules and central processes in attention and memory. In M. F. Schwartz (Ed.), *Modular deficits in Alzheimer-type dementia* (pp. 1-59). Cambridge, MA: MIT Press.

Moscovitch, M., & Umilta, C. (1991). Conscious and nonconscious aspects of memory: A neuropsychological framework of modules and central systems. In R. G. Lister & H. J. Weingartner (Eds.), *Perspectives on cognitive neuroscience* (pp. 229-266). New York: Oxford University Press.

Park, D. C., Smith, A. D., Dudley, W. N., & Lafronza, V. N. (1989). Effects of age and a divided attention task presented during encoding and retrieval on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1185-1191.

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 358-377.

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. (1994a). Graded capacity-sharing in dual-task interference? *Journal of Experimental Psychology: Human Perception and Performance, 20,* 330-342.

Pashler, H. (1994b). Overlapping mental operations in serial performance with preview. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 47A,* 161-191.

Pashler, H., & Johnston, J. C. (1989). Chronometric evidence for central postponement in temporally overlapping tasks. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 41A,* 19-45.

Schweickert, R. (1978). A critical path generalization of the additive factor method: Analysis of a Stroop task. *Journal of Mathematical Psychology, 18,* 105-139.

Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist, 57,* 421-457.

Trumbo, D., & Milone, F. (1971). Primary task performance as a function of encoding, retention, and recall in a secondary task. *Journal of Experimental Psychology, 91,* 273-279.

Welford, A. T. (1952). The "psychological refractory period" and the timing of high speed performance—A review and a theory. *British Journal of Psychology, 43,* 2-19.

Welford, A. T. (1980). The single-channel hypothesis. In A. T. Welford (Ed.), *Reaction time* (pp. 215-252). New York: Academic Press.

Received April 20,1993
Revision received October 24,1994
Accepted October 27,1994 ■