Journal of Experimental Psychology: General
1994, Vol. 123, No. 4, 354-373

0096-3445/94/$3.00

Working Memory and Retrieval: A Resource-Dependent Inhibition Model

Andrew R. A. Conway and Randall W. Engle

Four experiments examined individual differences in working memory (WM) capacity and how those differences affect performance on retrieval from both primary and secondary memory. The results showed that WM differences appear only in retrieval from primary memory and then only under conditions that lead to interference or response competition within the task. This suggests that WM capacity is important to retrieval that is based on controlled effortful search but not search that is based on automatic activation. A view is presented suggesting that individual differences in attentional resources lead to differences in the ability to inhibit or suppress irrelevant information. The paradigm also allowed more general comparisons between the processes involved in retrieval from primary and secondary memory. As expected, it was found that retrieval from primary memory was a function of set size. However, for sets larger than 2 items, retrieval from secondary memory was independent of set size.

In 1890 William James distinguished *primary memory* from the rest of the memory system. By his definition, primary memory is equated with the current contents of consciousness. His concept of primary memory dealt mainly with perceptual processes and can be distinguished from *secondary memory*, or memory proper, which consists of memory of the distant past that must be brought back into consciousness by some process. The notion that memory consists of two main compartments has been accepted in one form or another ever since (Craik & Levy, 1976).

However, many recent conceptions of human memory (i.e., Anderson, 1983; Cowan, 1988; McClelland & Rumelhart, 1986) assume that short-term memory (STM) is simply the activated portion of long-term memory (LTM). Although the relationship among the constructs STM, WM, and LTM is probably more complex (Cantor, Engle, & Hamilton, 1991), in this article we speak in terms of inactive and active portions of memory. For our purposes, we use James's terminology and define the inactive portion of LTM as secondary memory and the active portion of LTM as primary memory. We operationalize the term "active" below.

One purpose of this article is to examine individual differences in WM capacity and how those differences affect performance on retrieval tasks. There is considerable evidence that WM capacity plays an important role in a wide variety of tasks, including learning new information, following directions, taking notes, reasoning, and problem solving (Engle, in press). Less evidence has been provided regarding the role of WM in retrieval. Cantor and Engle (1993) recently found that high- and low-WM subjects significantly differed in performance on a fact-retrieval task. That effect could have occurred because of individual differences in retrieval from either primary memory, secondary memory, or both. Therefore, we will attempt to delineate the effect of limitations in WM capacity on retrieval from both primary and secondary memory.

This approach will also allow us to make more general comparisons of the processes involved in retrieval from primary memory and secondary memory. Following the logic of Wickens, Moody, and Dow (1981), we will compare reaction time (RT) performance in primary memory and secondary memory conditions in a memory-search task (i.e., Sternberg, 1966). On the one hand, it is possible that the processes involved in retrieval from primary memory are identical to those involved in retrieval from secondary memory. On the other hand, the processes involved in retrieval from primary and secondary memory may be different. Most experiments designed to study retrieval from memory have used tasks aimed at tapping retrieval from primary memory or secondary memory, but not both. If retrieval dynamics are not different in primary and secondary memory, one has to question the need for such a distinction. A demonstration of different retrieval dynamics for tasks putatively measuring primary and secondary memory, however, would provide strong support for the qualitative distinction between two states of memory activation.

Our concept of memory activation is taken from Anderson's ACT* model (Anderson, 1983). Therefore, we will proceed with a review of the basic assumptions of that model. We will then briefly review the literature on individual differences in WM capacity. The manipulation comparing retrieval from primary and secondary memory is motivated by a series of studies conducted by Wickens et al.

Andrew R. A. Conway and Randall W. Engle, Department of Psychology, University of South Carolina.

This research was submitted to the University of South Carolina in partial fulfillment of the requirements for the Master of Arts degree by Andrew R. A. Conway. This work was supported by Grant RO1-HD27490-01A1 from the National Institute of Child Health and Human Development and Grant F49620-93-1-0336 from the Air Force Office of Scientific Research. We thank Nelson Cowan, Robin Morris, Doug Wedell, and Jim Laughlin for their helpful comments.

Correspondence concerning this article should be addressed to Randall W. Engle, Department of Psychology, University of South Carolina, Columbia, South Carolina 29208.

(1981). As such, we will finish the introduction with a close examination of their methods and results.

LTM Activation

Anderson's ACT* model assumes that primary memory consists of information in secondary memory that has been stimulated or activated above some critical threshold (Anderson, 1983). Activation is considered to be a limited resource that automatically spreads among related concepts. A concept becomes active and accessible to cognitive processes when the amount of activation available to it reaches some critical threshold. As the activation level of a concept increases, so does its accessibility. Anderson (1976) designed the fact-retrieval paradigm in an attempt to measure the amount of activation available to LTM.

In the fact-retrieval task, subjects memorize a set of sentences that consist of a subject and a predicate (i.e., The plumber is in the park.). The number of predicates paired with each subject varies. This number is termed the *fan size* (Anderson, 1974). After the learning phase there is a speeded verification test in which the subject must distinguish between studied sentences and foil sentences. Reaction time and error rate are consistently found to be greater to sentences associated with a large fan size than to those associated with a small fan size (see Anderson, 1976, for a review). This phenomenon is termed the *fan effect*.

Jones and Anderson (1987) used an item-recognition task, a fact-retrieval task, and a hybrid precuing task to compare STM and LTM retrieval. They found that the factors of information load and relatedness had comparable effects on retrieval from STM and LTM. To explain their results, they proposed the indirect pathway model of memory retrieval, which is based on the associational, spreading activation assumptions of ACT*. It includes two important features. First, the pretrial activation level of any node referenced by a test probe varies. If the pretrial activation level of a node is relatively high, then the information pertaining to that node is assumed to be in active, or short-term, memory. In contrast, if the pretrial activation level of a node is low, then the information pertaining to that node is assumed to be in LTM. The second feature of the model is that a decision regarding a probe of information can sometimes be based on the retrieval of an indirect pathway in which the connection of probe elements is partially accomplished through the use of preexperimental associations.

LTM Activation and WM Capacity

Individual differences in WM capacity have been a topic of considerable inquiry in the last decade. Much of this research has focused on the positive relationship between WM capacity and comprehension ability. A number of hypotheses have been proposed to explain this relationship, including the task-specific hypothesis (Daneman & Carpenter, 1980), the strategic-allocation hypothesis (Carpenter & Just, 1989), and the general-capacity model (Engle, Cantor, & Carullo, 1992). Engle et al. (1992) conducted a series of experiments testing the unique predictions of each hypothesis and concluded that the general-capacity model was the most viable explanation of the relationship between measures of WM capacity and measures of reading comprehension.

The ideas behind the general-capacity model are similar to theories that have been proposed by Anderson (1983), Schneider and Detweiler (1987), and Cowan (1988). The model assumes that WM is the activated portion of long-term declarative memory. Therefore, WM capacity and the amount of activation available to LTM are equivalent. It is the level of activation that differs among individuals and manifests itself in a wide range of cognitive tasks. Furthermore, capacity or activation was argued to change very little with changes in knowledge structure. Therefore, an individual may develop a knowledge structure that allows for a dramatic increase in the amount of information he or she is able to store and recall. However, an analysis of retrieval would show a relatively small number of links at each level of the hierarchy (Ericsson & Staszewski, 1989). Therefore, according to the general-capacity model, even if high- and low-span subjects are equated for learning or for their level of knowledge in a task, they would still show retrieval differences.

A recent study conducted in our lab directly compared Anderson's notion of LTM activation with WM capacity (Cantor & Engle, 1993). There is considerable evidence that WM capacity differs among individuals and that this difference is revealed by a number of cognitive tasks (Engle, in press). Following the logic of the general capacity model, Cantor and Engle (1993) argued that individual differences in capacity exist because people differ in the amount of activation available to LTM. If this is true, high- and low-WM subjects should show a different pattern of results in a fact-retrieval task (Anderson, 1974, 1976).

In their Experiment 1, Cantor and Engle (1993) compared performance of high- and low-WM subjects on a fact-retrieval task. The sentences each subject learned were thematically unrelated. As fan increased, low-WM subjects showed a larger increase in RT than did high-WM subjects. Furthermore, when the slope of the fan effect was partialed out of the correlation between WM and verbal abilities, the relationship was no longer significant. This suggests that the limit on LTM activation, as measured by the fact-retrieval task, is functionally identical to WM capacity.

In summary, a positive relationship exists between WM-capacity and comprehension ability. The general-capacity model of WM is best able to explain this relationship. In further support of the model, Cantor and Engle (1993) found that limits on LTM activation, as measured by a fact-retrieval task, statistically accounted for the relationship between WM and comprehension.

Comparing Retrieval From STM and LTM

Wickens et al. (1981) proposed the following equation for RT for retrieval from primary memory in a memory scanning (i.e., Sternberg, 1966) task: $RT_p = a + b(m)$. The

intercept, or *a* term represents the probe encoding process and does not vary with set size. The slope, or *b* term, however, represents the comparison stage and does vary with set size. Wickens et al. also proposed an equation for RT from secondary memory in a memory scanning task: $RT_s = [a + b(m)] + R$. Notice that the term in brackets is identical to the equation for retrieval from primary memory. Thus, the difference between retrieval from primary and secondary memory is represented by the *R* term. *R* refers to the processing that makes a memory set available to active or primary memory. This action must occur if the scanning process is to be accomplished. As it is written here, *R* does not interact with set size. Therefore, this equation assumes that the time required to make a set of four items available to primary memory is no greater than the time required to make a set of two items available to primary memory.

Wickens et al. (1981) conducted two experiments to determine whether the placement of the *R* term in the above equation is correct. Their first experiment was a test of retrieval from secondary memory. A memory set of either two or four words was in a vertical array presented for 3 s. Then a random three-digit number appeared on the screen. The subjects were instructed to count backward by threes from the three-digit number for 12 s. The screen then became blank for 2 s, indicating to the subjects that they could stop counting and prepare for the probe item. A probe word was then presented, and the subjects were asked to press a key to indicate whether the probe was one of the words studied on that trial. Reaction time to the probe was recorded.

The second experiment was the same as the first, except there was no distractor task. Immediately after the words were shown, the subjects received a probe for recognition. Therefore, the second experiment was a simple item-recognition task that tested primary memory. By comparing RT performance on the two tasks, Wickens et al. (1981) were able to determine whether retrieval from secondary memory fit the second equation and did not interact with set size. Two outcomes were possible. If primary–secondary memory condition did not interact with set size, then Wickens et al.'s original placement of the *R* term was correct. However, if memory condition did interact with set size, then the *R* term cannot be independent of set size in their equation, meaning the second equation is wrong.

Consistent with previous findings (Sternberg, 1966), Wickens et al. (1981) found a linear relationship between set size and RT in the primary memory condition. They found the same relationship, with the same slope, in the secondary memory condition. The only difference between the primary and secondary memory conditions was for the y-intercept. There was no interaction between set size and memory condition. Therefore, the equation Wickens et al. (1981) used for retrieval from secondary memory, $RT_s = [a + b(m)] + R$, appears to be correct.

The absence of an interaction between set size and whether retrieval was from primary or secondary memory is an extremely important finding. It provides insight into the nature of retrieval from secondary memory, suggesting that the initial memory activity consists of the retrieval of the address for a list rather than the retrieval, one by one, of the items on that list. Furthermore, the time required to activate a list from secondary into primary memory is identical for a list of four items and a list of two items.

Some aspects of the procedure used by Wickens et al. (1981) merit close scrutiny. First, in the secondary memory condition, each subject was presented with a blank screen for 2 s after counting and before presentation of the test item. This is ample time for the subject to activate the memory set from secondary memory into primary memory. The results suggest that this may have occurred for some subjects, because the measures of variance were much larger for the secondary memory condition than for the primary memory condition. Another problem with Wickens et al.'s task is that they used memory sets of only two and four items, which is well within the memory span of the average college student. Perhaps the conclusion that activation into primary memory is independent of the size of the set being activated is true only for subspan-sized sets. The task we used in the present research was designed to measure retrieval from primary and secondary memory independently while avoiding the potential shortcomings of Wickens et al.'s (1981) procedure.

The Current Paradigm

One of the concerns about the Wickens et al. (1981) study was that they used set sizes of only 2 and 4 items, leading to the possibility that their results were specific to sets of subspan length. In two of the studies reported in this article (Experiments 1 and 3) we used a search task with sets of 2, 4, 6, or 8 letters. In the other two studies (Experiments 2 and 4) we used sets of 2, 4, 6, 8, 10, or 12 words.

Another concern about the Wickens et al. (1981) procedure was that items were presented once and then tested. Thus, the possibility exists that the items in the two-item set and in the four-item set were learned to different levels and that retention was worse in the secondary memory condition because of the delay and interference. Interpretation of findings from a speeded search task such as the memory scanning (Sternberg, 1966) procedure relies on the level of accuracy across conditions being high and relatively similar. Ideally, the researcher is measuring speed of performance, not level of learning. To avoid this problem, we required subjects to memorize all the memory sets before performing the item recognition task.

Once they had memorized the memory sets to criterion, subjects performed a speeded verification task. Half the trials in the task were a test of primary memory, and half were a test of secondary memory. In a primary-memory trial, a number, designating one of the memory sets (e.g., 6), was presented for 1 s before the probe item appeared. Therefore, the subjects knew which memory set would be tested before the probe appeared. This meant that subjects could have the appropriate memory set activated and represented in primary memory when the probe appeared. In a secondary-memory trial, the digit corresponding to the memory set number and the probe item were presented

simultaneously. Therefore, the subjects did not know which memory set would be tested until the probe appeared. At probe onset of the secondary-memory trials, the appropriate set was presumed to be inactive, or in other words, represented in secondary memory. It is important to note that the time to access a memory set in secondary memory is not reflected by averaging the RTs for the secondary-memory trials. This is because a secondary-memory trial required the subjects to first activate the appropriate set from secondary memory into primary memory and then search that set to determine whether the probe was a match. (We should note here that the term *search* is used without regard to the issue of whether search is serial or parallel.)

Therefore, we assume that two processes were required of the subjects in the secondary-memory trials. The first process reflects the time required to access a set from secondary memory, and the second reflects the time to search the memory set for the probe item. The primary-memory trials also reflect the time to search the memory set. Therefore, the time required to access a set in secondary memory can be derived by subtracting the time required for primary-memory trials from the time required for secondary-memory trials. In sum, the measure of retrieval time from primary memory is the average of RTs on primary-memory trials. The measure of retrieval time from secondary memory is the difference between the secondary-memory trials and the primary-memory trials.

It was also our goal to explore the role of WM capacity in retrieval. As mentioned above, Cantor and Engle (1993) found an interaction between WM span and fan size in a fact-retrieval task. As fan size increased, low-span subjects showed a greater increase in RT than did high-span subjects. We designed the tasks and stimuli in the present studies to be analogous to those used in Cantor and Engle (1993). In their Experiment 1, each subject memorized a set of sentences that consisted of a subject and a predicate (i.e., "The plumber is in the park," see our Figure 1). The number of predicates paired with each subject varied. We varied the number of items associated with a certain memory set in an analogous manner. Therefore, Cantor and Engle (1993) studied RT as a function of fan size, and the present studies measured RT as a function of memory set size.

There is another feature of the materials used by Cantor and Engle (1993; as well as by much of the fan effect literature) that needs to be noted. Each predicate was associated with two subjects. For instance, a subject might learn "The plumber is in the park" and "The teacher is in the park." Therefore, for our Experiments 1 and 2, the memory sets were constructed such that each memory set item was a member of two different sets. This feature might lead to a level of interference or response competition during the speeded verification task. If so, perhaps the reason Cantor and Engle (1993) found a difference between high- and low-span subjects was because of their differing abilities to handle response competition. If that were true, then high- and low-span subjects would not differ in a fact-retrieval task in which each predicate is unique or in which each item belongs to one and only one memory set. To test for this possibility, we constructed the materials for Experiments 3 and 4 so that each memory set item was a member of only one set.

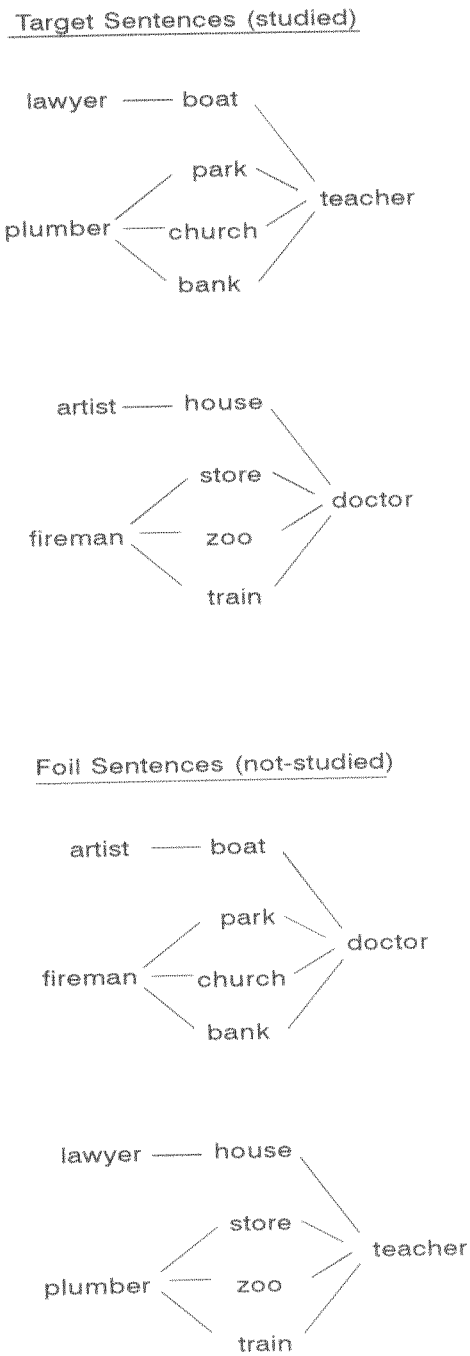

Figure 1. Materials used in Cantor and Engle (1993, Experiment 1).

In sum, we conducted four experiments. Letters were used as memory set items in Experiments 1 and 3, and words were used as memory set items in Experiments 2 and 4. In Experiments 1 and 2 the memory sets were analogous to the materials used by Cantor and Engle (1993) in that each memory set item was a member of two different sets.

In Experiments 3 and 4, each memory set item was unique to one set.

This series of experiments addressed two main issues. First: What is the difference between retrieval from primary memory and retrieval from secondary memory? It seems to be an established fact in cognitive psychology that retrieval from primary memory is a function of set size. Does the same relationship occur for retrieval from secondary memory? As mentioned above, the measure of retrieval from primary memory is the average RT on primary-memory trials, whereas the measure of retrieval time from secondary memory is the difference between secondary- and primary-memory trials. Therefore, if retrieval from secondary memory is a function of set size, then we would expect to see a steeper slope for the secondary-memory trials than for primary-memory trials. In contrast, if retrieval from secondary memory is not a function of set size, then the functions for the primary-memory trials and the secondary-memory trials should be parallel.

The second issue relevant to this set of experiments is the role of WM capacity in retrieval tasks. Cantor and Engle (1993) found that, in a fact-retrieval task, low-span subjects showed a larger fan effect than high-span subjects. Given this finding, one might conclude that high- and low-span subjects differ in retrieval from LTM or secondary memory. However, this finding may indicate that high- and low-span subjects differ in retrieval time from either secondary memory, primary memory, or both. The fact-retrieval task might incorporate retrieval from primary memory as well as retrieval from secondary memory. When a subject is asked to verify having studied "The plumber is in the park," he or she might first activate "plumber" and all the places associated with "plumber." This would be the equivalent of bringing the set "places associated with plumber" into primary memory. There then might occur a process equivalent to the search of primary memory. Therefore, like our secondary-memory trials, the fact-retrieval task would require two processes. The first process is the encoding of the subject term and the activation of the set of places associated with that subject. This could of course be done one place at a time, or it could be done at the level of the entire set. The second process would be the search of the set in active or primary memory to determine whether the probed place was, in fact, one of the places associated with the subject.

It could be that high- and low-span subjects differ in the time required to perform the first process, the second process, or both. In other words, it could be that high- and low-span subjects differ in the amount of time required to retrieve the set information from secondary memory, the individual probe information from primary memory, or both.

It is also possible that high- and low-span subjects do not differ in retrieval time from either primary memory or from secondary memory. As mentioned above, the materials from Cantor and Engle (1993) were designed in such a way that interference may play a role in the task. It could be that this interference factor accounts for the difference in fan effect between high- and low-span subjects. If that is the case, then we should see a difference in the set-size effect for high- and low-span subjects in Experiments 1 and 2 but not in Experiments 3 and 4. If WM capacity plays a role in retrieval, independent of interference, then we might expect to see span differences in the set-size effect for all four experiments.

In sum, there are four possible outcomes for the set-size effects for high- and low-span subjects; they may differ in retrieval from primary memory, secondary memory, both, or neither. These four possibilities are illustrated in Figure 2. The four panels show hypothetical results of RT in the speeded verification task as a function of set size,

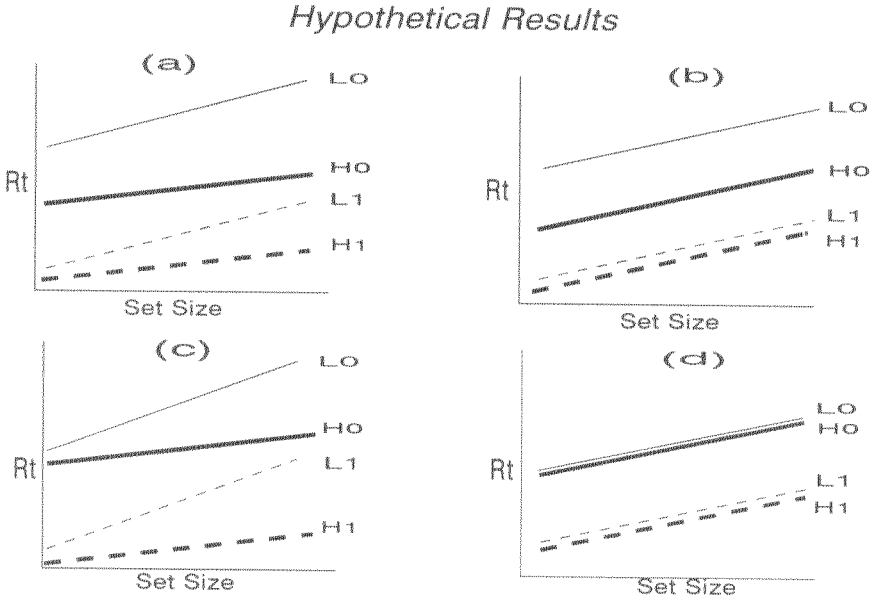

Figure 2. Possible outcomes in terms of the difference between high- and low-WM subjects.

WM capacity, and type of trial. Primary-memory trials are denoted by a 1, indicating that there is a 1-s delay between the onset of the digit indicating the particular set being tested and the letter or word being probed on that trial. Secondary trials are denoted by a 0, reflecting the fact that the digit cuing the particular set and the probe for that trial appeared simultaneously.

Panel (a) of Figure 2 shows the results we would expect if high-span subjects are faster at retrieval from both primary and secondary memory. The slopes for high- and low-span subjects would differ at both delay conditions. Furthermore, the difference between the two delay conditions (representing the time required to activate a set from secondary memory into primary memory) would be greater for low-span subjects than for high-span subjects.

If high- and low-span subjects differ in retrieval time from secondary memory but not from primary memory, the results should look like those in Panel (b) of Figure 2. The functions for the primary-memory trials should be equivalent, but the difference between the secondary-memory trials and the primary-memory trials should be greater for low-span subjects.

Panel (c) of Figure 2 shows the results we would expect if WM capacity is reflected in retrieval from primary memory but not from secondary memory. The slopes for the primary-memory trials should differ, but the difference between the primary-memory trials and the secondary-memory trials would be equivalent for high- and low-WM-capacity subjects. If high- and low-span subjects do not differ in retrieval time from either primary memory or secondary memory, the results would look like those in Panel (d).

The reader should reflect carefully on Panels (c) and (d), because, at the risk of divulging the punch line to the entire article, the results of Experiments 1 and 2 look very much like Panel (c), and the results of Experiments 3 and 4 look very much like Panel (d).

Subject Screening

Subjects were screened for WM capacity on the basis of their performance on an operation–word span task similar to that used by LaPointe and Engle (1990).

Operation–Word Span Task

This task requires a subject to solve simple mathematical operations while trying to remember words. We used the same operations as in previous studies conducted in our lab (Cantor & Engle, 1993; LaPointe & Engle, 1990). Each operation began with the multiplication or division of two integers, the result of which was added to or subtracted from a third integer (i.e., (8/4) + 2 = 4). An answer, which was correct approximately half the time, was provided after the operation. Each operation was presented with a high-frequency, one-syllable word (taken from LaPointe & Engle, 1990). For each subject, a pool of 66 mathematical operations was randomly paired with a pool of 66 to-be-remembered words. During the task, subjects were presented with operation–word pairs (i.e., (8/4) + 2 = 4 BIRD). Each subject was to read the operation aloud, say "yes" or "no" to indicate if the provided answer was correct or incorrect, and then say the word. After the subject said the word, the experimenter pressed a key, and another operation–word pair was presented. This process continued until a question mark was presented, which cued the subject to recall the to-be-remembered words. The subject was then asked to write the to-be-remembered words on a response sheet. The number of operation–word pairs per series varied from two to six. Three series of each size (two–six) were performed. Therefore, a total of 15 (3 × 5) series were presented, and their order of presentation was randomized. Three additional series, each consisting of two items, were provided as practice for the subject. A subject's span score was the sum of the correctly recalled words for trials that were perfectly recalled (and in correct order). This score was originally reported by Turner and Engle (1989), and consistently correlates with performance on the Verbal Scholastic Aptitude Test (VSAT; Cantor & Engle, 1993; Cantor, Engle, & Hamilton, 1991; Engle, Cantor, & Carullo, 1992; LaPointe & Engle, 1990). Each subject's accuracy on the operations was also recorded. If accuracy was below 85%, the subject was not included in any of the following experiments. If a subject's span score was 20 or higher, he or she was classified as high span. If a subject's span was 12 or lower, he or she was classified as low span. These cutoffs are based on the upper and lower quartiles from previous work in our lab (Cantor & Engle, 1993, Experiment 1, N = 80) and are consistent with our current quartile boundaries, which are based on approximately 300 subjects.

Experiment 1

Method

Subjects

Twenty high-span and 20 low-span subjects, as determined by their performance on the operation–word span task described above, were used as subjects. The mean span score for the 20 high-span subjects was 23.75. The mean span score for the 20 low-span subjects was 4.90. All subjects were undergraduates from the University of South Carolina, were tested individually, and received course credit for participation.

Materials

Each subject memorized four sets of letters, which consisted of 2, 4, 6, and 8 letters respectively. Each letter appeared as a target in two sets. Only consonants (excluding "Y") were used in the study. The letter sets for each subject were predetermined by a random process. The process involved randomly ordering the 20 consonants in the alphabet. Because each letter appeared as a target in two different sets, only the first 10 consonants (after randomization) were used ([2 + 4 + 6 + 8]/2 = 10). The first 8 letters were designated as Set 8. These 8 letters were distributed across the other three sets such that 1 of the 8 letters was in Set 2, 3 of the 8 letters were in Set 4, and 4 of the 8 letters were in Set 6. The two

letters used that were not members of Set 8 appeared as targets in Set 6. One of those 2 also appeared as a target in Set 2 and the other as a target in Set 4. This design is illustrated in Appendix A, along with the items from the other experiments. The goal of the design was to ensure that a letter set was not simply a subset of a larger set.

The experiment was run on an IBM PS/2 computer, with Micro Experimental Laboratory (MEL) software (Schneider, 1988). Stimuli were presented on a VGA monitor. Subjects responded using an IBM AT keyboard, using the "1" and "3" keys.

Procedure

Subjects were asked to complete a learning stage and a verification stage in the same session. The session lasted approximately an hour and a half.

Learning stage. Each letter set appeared on the computer screen for a fixed amount of time. The algorithm n(10sec) + 10sec, where *n* equals the set size, determined how long each set remained on the screen. The order of set presentation was random. After viewing each set twice, the subjects were asked to recall each set separately. Each subject recalled a set by saying the letters aloud to the experimenter, who recorded accuracy. After recall, the subject studied each set again. After study, the subject was asked to recall again, and so on. When the subject recalled a set correctly three consecutive times, that set was dropped from the study–recall cycle. After all sets had dropped out, each set was presented once more for study.

Verification stage. Each subject performed 384 speeded verification trials. Four levels of set size, two levels of delay, and two response types allowed for 16 different trial types. Therefore, 24 trials of each type were presented (16 × 24 = 384). A letter appeared as a target *n* times, where *n* = 24/set size. Possible foils for each trial included all the letters used in the experiment that were not members of the target set. There was an equal number of targets and foils. Trials were random, such that the subject had no way to predict what set size or delay would be used in the next trial. Each trial proceeded as follows. The word "READY" appeared on the screen for 1,000 ms. Then a number (either 2, 4, 6, or 8) appeared in the center of the screen, to designate one of the letter sets. The probe letter appeared either simultaneously with, or 1,000 ms after, the onset of the number.[1] The letter was presented on the line below the number. Both the number and letter remained on the screen until the subject responded by pressing a key. The "1" and "3" keys on the numeric keypad were designated as "yes" and "no" keys, respectively.

Design

Learning stage. This was a 4 × 2 mixed factorial design. The within-groups variable was set size (2, 4, 6, 8), and the between-subjects variable was WM capacity (high span or low span). The dependent measure was the number of study–recall cycles needed to reach criterion.

Verification stage. This was a 4 × 2 × 2 × 2 mixed factorial design. The within-groups variables were set size (2, 4, 6, 8), delay (0 and 1,000 ms), and response type (target or foil). The between-subjects variable was WM capacity (high span or low span). Two dependent measures were recorded: RT and error rate.

Results

Learning Stage

The mean numbers of study–recall cycles are reported in Table 1. Low-span subjects needed slightly more cycles to reach criterion than did high-span subjects. This was supported by a span (high, low) by set size (2, 4, 6, 8) analysis of variance (ANOVA) on the mean number of cycles needed to reach criterion. The main effect for span was marginal, $F(1, 38) = 3.62$, $p = .065$, $MS_e = 3.866$. No other effects were significant (for all, $p > .10$).

Verification Stage

In a typical memory-search task (i.e., Sternberg, 1966), a subject is presented with a memory set and then a probe in the same trial. Therefore, to respond to a foil, the subject must compare the probe with the memory set items (either serially or in parallel). In our paradigm the subjects memorize the sets before they perform the verification phase. Therefore, it is not clear what the subjects do when they are presented with a foil. They could either compare the foil with all the items of the probed set, or they could use a strategy that is based on their knowledge of the sets other than the one probed. Indeed, in all of our experiments, the error variance for foils is much greater than that for targets, suggesting that a greater variety of strategies was used to respond to foils than to respond to targets. For this reason, we chose to analyze target and foil data separately and to restrict our discussion to the target data (foil analyses are reported in Appendix B). Block medians for correct responses were used in all the RT analyses.

Reaction time and accuracy data were analyzed separately. As can be seen in Figure 3, RTs increased with set size. Also, RTs were faster in the 1-s delay condition than in the 0-s delay condition, and low-span subjects were slower to respond than high-span subjects. More important, as set size increased, low-span subjects showed a larger increase in RT than high-span subjects. This suggests that high- and low-span subjects differ in retrieval from primary memory. The difference between the 0-s delay function and the 1-s delay function was not different for high- and low-span subjects. This suggests that high- and low-span subjects do not differ in retrieval from secondary memory. Finally, the difference between the 0-s delay function and the 1-s delay function is not constant across set size. The difference between the 0-s delay condition and the 1-s delay condition

[1] We conducted a pilot study manipulating the delay between the presentation of the set number and probe item. We used either a 0-, 200-, 600-, or 1,000-ms delay. We found that RTs became quicker as delay increased. Specifically, RT with no delay was significantly slower than RT with a 200-ms delay. Reaction time with a 200-ms delay was significantly slower than RT with a 600-ms delay. However, there was no significant difference between the 600- and 1,000-ms delay conditions. Therefore, we concluded that 1,000 ms is more than enough time for a subject to activate a set from secondary memory.

Table 1
Mean Number of Study–Recall Cycles as a Function of Set Size and Span Group for All Experiments

Experiment	Memory set size[a] 2	4	6	8	10	12
1						
High span	5.00	5.20	5.35	5.35		
Low span	5.00	5.75	6.10	6.10		
2						
Session 1						
High span	5.10	5.70	6.00	6.55	7.10	6.60
Low span	5.10	6.10	6.90	7.80	9.20	8.90
Session 2						
High span	4.00	4.00	4.00	4.15	4.15	4.10
Low span	4.05	4.05	4.15	4.40	4.55	4.25
3						
High span	5.05	5.20	5.35	5.45		
Low span	5.00	5.40	5.30	5.85		
4						
Session 1						
High span	5.00	5.05	5.35	5.45	5.70	5.40
Low span	5.00	5.45	6.05	6.60	6.85	6.60
Session 2						
High span	4.00	4.00	4.00	4.00	4.00	4.00
Low span	4.00	4.10	4.05	4.00	4.10	4.20

Note. Perfect recall required five study–recall cycles except in Experiments 2 and 4, Session 2, which required only four study–recall cycles for perfect recall.
[a] Number of items.

is smaller at a set size of two items than at the other set sizes. However, when the analysis was performed on all set sizes except that of two items, the difference between the 0-s delay function and the 1-s delay function was not significantly different across set size. Therefore, beyond Set 2, the difference between the 0-s delay function and the 1-s delay function was constant across set size.

These conclusions were supported by our statistical analyses. We submitted the median RTs to a span (high, low) by set size (2, 4, 6, 8) by delay (0, 1,000) ANOVA. There were main effects for set size, $F(3, 114) = 69.17$, $p < .05$, $MS_e = 160{,}532$; delay, $F(1, 38) = 537.00$, $p < .05$, $MS_e = 38{,}504$; and span, $F(1, 38) = 9.16$, $p < .05$, $MS_e = 1{,}225{,}812$. More important, there was a span by set size interaction, $F(3, 114) = 3.06$, $p < .05$. The span by delay interaction was not significant, $F(1, 38) = 1.75$, $p > .10$. The set size by delay interaction was marginal, $F(3, 114) = 2.76$, $p = .046$, $MS_e = 31{,}026$. An analysis excluding Set 2 revealed that the set size by delay interaction was not significant, $F(2, 76) < 1$.

Error rates were low and did not differ between high- and low-span subjects. Subjects became less accurate as set size increased. As with the RT data, mean error rates were submitted to a span (high, low) by set size (2, 4, 6, 8, 10, 12) by delay (0, 1,000) ANOVA. There were main effects for set size, $F(3, 114) = 12.11$, $p < .05$, $MS_e = .006$, and delay, $F(1, 38) = 8.23$, $p < .05$, $MS_e = .002$. There was not a main effect for span, $F(1, 38) < 1$. None of the interactions were significant (for all, $p > .10$).

Discussion

Low-span subjects required slightly more study–recall cycles to reach criterion in the learning stage than high-span subjects. Otherwise, there were no significant effects in the learning data.

The more interesting data come from the verification stage. Consider the two main issues we addressed in the introduction. First, what is the difference between retrieval from primary memory and retrieval from secondary memory? Our measure of retrieval from primary memory is the median RT on trials in the 1-s delay condition. The slope of this function changed as set size changed. Therefore, retrieval from primary memory is dependent on set size. Our measure of retrieval from secondary memory is the difference between RTs in the 0-s delay condition and those in the 1-s delay condition. Contrary to Wickens et al. (1981), we found an interaction between set size and delay. This interaction seemed to be caused by a set size of 2. The difference in RT between the 1-s delay condition and the 0-s delay condition was much smaller at a set size of 2. When we conducted an analysis that excluded a set size of 2, the delay by set size interaction was not significant. Therefore, for set sizes beyond two items, the time required to activate a set from secondary memory to primary memory is constant.

The second issue we addressed in the introduction regards the difference between high- and low-span subjects. Cantor and Engle (1993) found a span by fan size interaction, such that as fan size increased, low-span subjects' RTs increased more than those of high-span subjects. Similarly, we found a span by set size interaction. However, we did not find a span by delay interaction. Therefore, the difference between the 0-s delay function and the 1-s delay function is the same for high- and low-span subjects. These results most resem-

Figure 3. Mean verification times and error rates for Experiment 1 as a function of set size, delay, and WM span. H and L refer to high and low span subjects. 0 and 1 refer to the 0-s and 1-s delay conditions.

ble the hypothetical results shown in Figure 2 (Panel c). We see a difference between high- and low-span subjects in retrieval from primary memory but not from secondary memory.

In sum, we found that WM capacity plays an important role in retrieval from primary or active memory but not from secondary or inactive memory. To replicate this finding and to further generalize our conclusions, we conducted Experiment 2, in which we used words as memory set items.

Experiment 2

Method

Subjects

Twenty high-span and 20 low-span subjects, as determined by the operation–word span task described previously, were used as subjects. The mean span score for the 20 high-span subjects was 25.35. The mean span score for the 20 low-span subjects was 7.58. Two subjects (both low span) were unable to perform the learning stage and were therefore replaced. All subjects were undergraduates from the University of South Carolina, were tested individually, and received course credit for participation.

Materials

Each subject memorized six sets of words, which consisted of 2, 4, 6, 8, 10, and 12 words respectively. Each word appeared as a target in two sets. The word sets for each subject were predetermined by a random process. The process involved randomly ordering 42 words. Because each word appeared as a target in two different sets, only the first 21 words (after randomization) were used ([2 + 4 + 6 + 8 + 10 + 12]/2 = 21). The words and a typical arrangement are illustrated in Appendix A. The goal of the design was to ensure that one word set was not simply a subset of a larger set. This was done by assigning the first 12 words to Word Set 12 and then distributing the rest of the words evenly across sets, to minimize overlap between sets. Twenty of these orderings were generated, one for each low- and high-span subject.

The computers and software used were the same as in Experiment 1.

Procedure

Subjects performed the experiment in two sessions. Each session lasted approximately an hour and a half, and the second session occurred within 48 hr of the first session. The subjects performed the learning stage in the first session. In the second session they performed an additional learning stage and then the verification stage.

Session 1

Each word set appeared on the computer screen for a fixed amount of time. The algorithm n(10sec) + 10sec, where *n* equals the set size, determined how long each set remained on the screen. The order of set presentation was random. After viewing each set twice, the subjects were asked to recall each set separately. The subjects recalled a set by writing the words on a response sheet. They were allowed to recall the words in any order. They would then read their response to the experimenter, who recorded accuracy. After recall, the subjects studied each set again. After study, the subjects were asked to recall again, and so on. When a subject had recalled a set correctly three consecutive times, that set was dropped from the study–recall cycle. After all sets had dropped out, each set was presented once more for study.

Session 2

Learning stage. The learning stage for Session 2 was exactly the same as that for Session 1, with one exception. In Session 1, each word set was presented twice before the subjects were asked to recall. In Session 2, because the subjecs were already familiar with the sets, they were asked to recall after one presentation.

Verification stage. Each subject performed 576 speeded verification trials. Six levels of set size, two levels of delay, and two response types allowed for 24 different trial types. Therefore, 24 trials of each type were presented (24 × 24 = 576). A word appeared as a target *n* times, where *n* = 24/set size (for Word Set 10, four words were presented as a target three times, and six words were presented as a target twice). There were equal numbers of targets and foils. A foil was 1 of the 21 words used in the experiment that was not a member of the set being tested. Trials were random, such that the subjects had no way to predict what set size or delay would be used in the next trial. Each trial proceeded in the same way as in Experiment 1.

Design

Learning stage. This was a 6 × 2 mixed factorial design. The within-groups variable was set size (2, 4, 6, 8, 10, 12), and the between-subjects variable was WM capacity (high span or low span). The dependent measure was the number of study–recall cycles needed to reach criterion.

Verification stage. This was a 6 × 2 × 2 × 2 mixed factorial design. The within-groups variables were set size (2, 4, 6, 8, 10, 12), delay (0 and 1,000 ms), and response type (target or foil). The between-subjects variable was WM capacity (high span or low span). The dependent measures were RT and error rate.

Results

Three sets of data were analyzed from Experiment 2: the first learning stage, the second learning stage, and the verification stage. The dependent measure for the learning stages was the number of study–recall cycles. Reaction time and accuracy were the dependent measures in the verification phase.

Session 1: Learning Stage

The mean numbers of study–recall cycles are reported in Table 1. Low-span subjects needed more cycles to reach criterion than high-span subjects. Also, the number of cycles required to reach criterion increased as set size increased. Furthermore, low-span subjects showed a larger increase than high-span subjects in the number of cycles required to reach criterion as set size increased.

These conclusions were supported by a span (high, low) by set size (2, 4, 6, 8, 10, 12) ANOVA on the mean number of cycles needed to reach criterion. There were significant

effects for span, $F(1, 38) = 8.17$, $p < .05$, $MS_e = 9.855$, and set size, $F(5, 190) = 34.20$, $p < .05$, $MS_e = 1.561$. Furthermore, there was an interaction between span and set size, $F(5, 190) = 5.36$, $p < .05$.

Session 2: Learning Stage

Although the differences were very small, low-span subjects required slightly more cycles to reach criterion than high-span subjects. Also, the number of cycles needed to reach criterion increased as set size increased.

Again, these conclusions were supported by a span (high, low) by set size (2, 4, 6, 8, 10, 12) ANOVA on the mean number of cycles needed to reach criterion. The main effect for span was marginal, $F(1, 38) = 3.96$, $p = .054$, $MS_e = .464$. There was a main effect for set size, $F(5, 190) = 2.99$, $p < .05$, $MS_e = .249$. The interaction between span and set size was not significant ($p > .10$).

Verification Stage

Reaction time and accuracy data were analyzed separately, but both are illustrated in Figure 4.

As can be seen in Figure 4, RTs increased with set size until a set size of 10 and then dropped off at a set size of 12. Reaction times were faster in the 1-s delay condition than in the 0-s delay condition, and low-span subjects were slower to respond than high-span subjects. Beyond that, there are three important findings. First, low-span subjects showed a larger increase in RT, as set size increased, than high-span subjects. This suggests that high- and low-span subjects differ in retrieval from primary memory. Second, the difference between the 0-s delay function and the 1-s delay function was not different for high- and low-span subjects. This suggests that high- and low-span subjects do not differ in retrieval from secondary memory. Third, the difference between the 0-s delay function and the 1-s delay function was constant across set size. This suggests that retrieval from secondary memory is independent of set size.

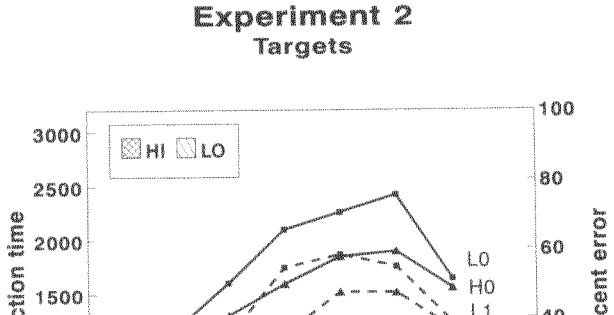

Figure 4. Mean verification times and error rates for Experiment 2 as a function of set size, delay, and WM span.

These conclusions were supported by a statistical analysis based on the median RTs in a span (high, low) by set size (2, 4, 6, 8, 10, 12) by delay (0, 1,000) ANOVA. Span group was the only between-subjects variable. There were main effects for set size, $F(5, 190) = 55.91$, $p < .05$, $MS_e = 223{,}888$; delay, $F(1, 38) = 216.47$, $p < .05$, $MS_e = 100{,}466$; and span, $F(1, 38) = 3.43$, $p = .072$, $MS_e = 3{,}026{,}209$. More important, there was a span by set size interaction, $F(5, 190) = 3.11$, $p < .05$. There was no significant interaction between span and delay, $F(1, 38) = 1.26$, $p > .10$, nor was there an interaction between set size and delay, $F(5, 190) = 1.78$, $p > .10$. Finally, the three-way interaction was not significant ($p > .10$).

Error rates were relatively low and did not differ between high- and low-span subjects. Subjects were generally less accurate as set size increased. As with the RT data, mean error rates were submitted to a span (high, low) by set size (2, 4, 6, 8, 10, 12) by delay (0, 1,000) ANOVA. There were main effects for set size, $F(5, 190) = 27.11$, $p < .05$, $MS_e = .007$, and delay, $F(1, 38) = 17.89$, $p < .05$, $MS_e = .002$. There was no main effect for span, $F(1, 38) < 1$. Also, none of the interactions were significant (for all, $p > .10$).

Discussion

Low-span subjects required more study–recall cycles to reach criterion in the learning stages than high-span subjects. Also, as set size increased from 2 to 10, more study–recall cycles were needed to reach criterion. Surprisingly, fewer cycles were needed to memorize Set 12 than Set 10.

Again, consider the two main issues we addressed in the introduction. First, what is the difference between retrieval from primary memory and retrieval from secondary memory? As in Experiment 1, we found that retrieval from primary memory is dependent on set size. Unlike in Experiment 1, we did not find a delay by set size interaction, including a set size of 2. Therefore, retrieval from secondary memory is independent of set size, meaning that large sets of words are retrieved from secondary memory as quickly as small sets.

The second issue we addressed in the introduction regards the difference between high- and low-span subjects. As in Experiment 1, we found a span by set size interaction but *not* a span by delay interaction. Therefore, the difference between the 0-s delay function and the 1-s delay function is the same for high- and low-span subjects. Again, these results most resemble the hypothetical results shown in Figure 2 (Panel c). We see a difference between high- and low-span subjects in retrieval from primary memory but not from secondary memory.

In both Experiments 1 and 2 we found that WM capacity plays an important role in retrieval from primary or active memory but not from secondary or inactive memory. As mentioned above, however, the materials in Experiments 1 and 2 were designed to incorporate interference into the

task. If the difference between high- and low-span subjects is due to this interference component, then we should not see a difference between high- and low-span subjects in Experiments 3 and 4, in which there was no overlap in set membership. If WM capacity plays a more general role in retrieval, then in Experiments 3 and 4 we should still see a difference between high- and low-span subjects.

Experiment 3

Method

Subjects

Twenty high-span and 20 low-span subjects, as determined by the operation–word span task described previously, were used as subjects. The mean span score for the 20 high-span subjects was 30.15. The mean span score for the 20 low-span subjects was 8.63. All subjects were undergraduates from the University of South Carolina, were tested individually, and received course credit for participation.

Materials

As in Experiment 1, each subject memorized four sets of letters, which consisted of 2, 4, 6, and 8 letters respectively. Unlike in Experiment 1, each letter appeared as a target in only one set. Only consonants (excluding "y") were used in the study. The letter sets for each subject were predetermined by a random process. The process involved randomly ordering the 20 consonants in the alphabet and designating the first 2 as Letter Set 2, the next 4 as Letter Set 4, and so on. Twenty of these orderings were generated, 1 for each high- and low-span subject. An example of one of these orderings is shown in Appendix A.

The computers and software used were the same as in Experiments 1 and 2.

Procedure and Design

The procedure and design were the same as in Experiment 1.

Results

Learning Stage

The mean numbers of study–recall cycles are reported in Table 1. High- and low-span subjects did not significantly differ in the number of study–recall cycles needed to reach criterion. The number of cycles needed to reach criterion increased as set size increased. These conclusions were supported by a span (high, low) by set size (2, 4, 6, 8, 10, 12) ANOVA. There was a main effect for set size, $F(3, 114) = 5.35$, $p < .05$, $MS_e = .489$. There was not a main effect for span, $F(1, 38) < 1$, nor was there an interaction, $F(3, 114) < 1$.

Verification Stage

Again, RT and accuracy data were analyzed separately. As can be seen in Figure 5, RTs increased with set size and were faster in the 1-s delay condition than in the 0-s delay condition. Also, low-span subjects responded more slowly overall than high-span subjects. However, high- and low-span subjects' RT functions increased with set size at the same rate. Therefore, even though low-span subjects were slower overall, the rate at which they searched primary memory is equal to that of high span subjects. Furthermore, the difference between the 0-s delay function and the 1-s delay function is the same for high- and low-span subjects. As in the previous two experiments, this suggests that high- and low-span subjects do not differ in retrieval from secondary memory. Finally, as in Experiment 1, the difference between the 0-s delay function and the 1-s delay function is not constant across set size. Again, it appears that the difference between the 0-s delay condition and the 1-s delay condition is smaller at a set size of 2 than at the other set sizes. When the analysis was performed on all set sizes except 2, the difference between the 0-s delay function and the 1-s delay function was constant across set size.

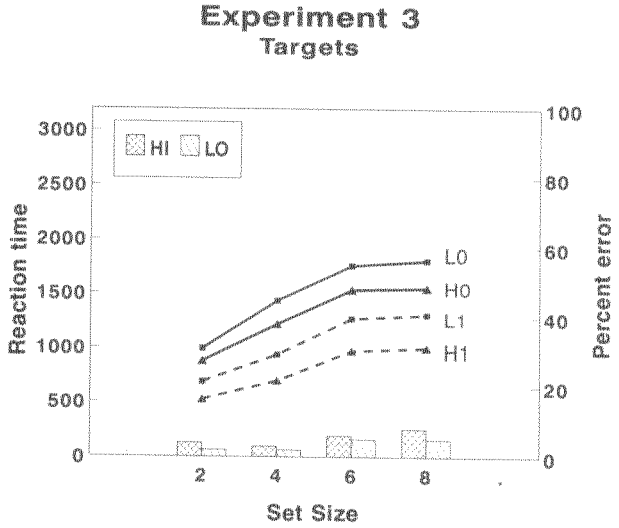

Figure 5. Mean verification times and error rates for Experiment 3 as a function of set size, delay, and WM span.

These conclusions were supported by our statistical analysis. We submitted the median RTs to a span (high, low) by set size (2, 4, 6, 8) by delay (0, 1,000) ANOVA. There were main effects for set size, $F(3, 114) = 76.16$, $p < .05$, $MS_e = 96{,}289$, and delay, $F(1, 38) = 349.67$, $p < .05$, $MS_e = 51{,}184$. There was also a main effect for span, $F(1, 38) = 5.11$, $p < .05$, $MS_e = 807{,}307$. However, there was no span by set size interaction, $F(3, 114) < 1$. Also, there was no span by delay interaction, $F(1, 38) = 1.01$, $p > .10$. There was a set size by delay interaction, $F(3, 114) = 9.66$, $p < .05$, $MS_e = 19{,}496$. However, when the set size of 2 was excluded from the ANOVA, the interaction between set size and delay was not significant, $F(2, 76) < 1$.

Error rates were low and did not differ between high- and low-span subjects. Subjects were less accurate as set size increased. As with the RT data, the mean error rates were submitted to a span (high, low) by set size (2, 4, 6, 8) by

delay (0, 1,000) ANOVA. There were main effects for set size, $F(3, 114) = 14.07$, $p < .05$, $MS_e = .003$, and delay $F(1, 38) = 11.13$, $p < .05$, $MS_e = .001$. There was no main effect for span, $F(1, 38) = 2.876$, $p > .05$. There was an interaction between delay and span, $F(1, 38) = 5.05$, $p < .05$. No other interactions were significant (for all, $p > .10$).

Discussion

Both high- and low-span subjects needed more study–recall cycles to reach criterion as set size increased. However, high- and low-span subjects did not differ in the number of cycles needed to reach criterion.

In terms of our retrieval question, the results are similar to those from Experiment 1. We found a delay by set size interaction, which was caused by a set size of 2. When we excluded the set size of 2 from the analysis, the delay by set size interaction was not significant. Therefore, consistent with both Experiments 1 and 2, the time required to activate a set from secondary memory is constant beyond Set 2.

In terms of our WM question, the results of Experiment 3 are quite different from those of Experiments 1 and 2. We found no difference between high- and low-span subjects in retrieval time from primary memory *or* from secondary memory. These results most resemble the hypothetical results shown in Figure 2 (Panel d). Thus, we replicated the finding from Experiments 1 and 2, that high- and low-span subjects do not differ in retrieval time from secondary memory. Furthermore, when the memory sets are not overlapping, high- and low-WM-span subjects do *not* differ in retrieval time from primary memory. To replicate this finding, we conducted Experiment 4, in which we used words as memory set items.

Experiment 4

Method

Subjects

Twenty high-span and 20 low-span subjects, as determined by the operation–word span task described previously, were used as subjects. The mean span score for the 20 high-span subjects was 28.60. The mean span score for the 20 low-span subjects was 8.40. One low-span subject was unable to perform the learning stage and was therefore replaced. All subjects were undergraduates from the University of South Carolina, were tested individually, and received course credit for participation.

Materials

As in Experiment 2, each subject memorized six sets of words, which consisted of 2, 4, 6, 8, 10, and 12 words respectively. Unlike in Experiment 2, each word appeared as a target in only one set. The word sets for each subject were predetermined by a random process. The process involved randomly ordering the 42 words and designating the first 2 as Set 2, the next 4 as Set 4, and so on. Twenty of these orderings were generated, 1 for each high- and low-span subject. An example of one ordering is shown in Appendix A.

The computers and software used were the same as in the previous experiments.

Procedure and Design

The procedure and design were the same as in Experiment 2.

Results

Three sets of data were analyzed from Experiment 4: the first learning stage, the second learning stage, and the verification stage. The dependent measure for the learning stages was the number of study–recall cycles. Reaction time and accuracy were the dependent measures in the verification phase.

Session 1: Learning Stage

The mean numbers of study–recall cycles are reported in Table 1. Low-span subjects needed more cycles to reach criterion than high-span subjects. Also, the number of cycles needed to reach criterion increased as set size increased.

These conclusions were supported by a span (high, low) by set size (2, 4, 6, 8, 10, 12) ANOVA on the mean number of cycles to reach criterion. There were significant effects for span, $F(1, 38) = 12.77$, $p < .05$, $MS_e = 3.315$, and set size, $F(4, 152) = 7.84$, $p < .05$, $MS_e = .786$. The interaction between span and set size was not significant, $F(4, 152) = 1.60$, $p > .10$.

Session 2: Learning Stage

Most subjects had little trouble with the learning stage. However, low-span subjects required slightly more cycles to reach criterion than high-span subjects. This was supported by a span (high, low) by set size (2, 4, 6, 8, 10, 12) ANOVA on the mean number of cycles needed to reach criterion. The effect for span was marginal, $F(1, 38) = 3.67$, $p = .063$, $MS_e = .138$. The main effect for set size was not significant, $F(3, 114) < 1$, nor was the span by set size interaction, $F(3, 114) < 1$.

Verification Stage

Reaction time data and accuracy data were analyzed separately. However, both RT and accuracy data are shown in Figure 6. As in Experiment 2, RTs increased with set size until a set size of 10 and then dropped off at a set size of 12. Reaction times were faster in the 1-s delay condition than in the 0-s delay condition. Overall, high-span subjects were not significantly faster than low-span subjects. Furthermore, the RT functions for high- and low-span subjects increased with set size at the same rate. This suggests that high- and low-span subjects did not differ in retrieval from primary memory. As in the three previous experiments, the difference between the 0-s delay function and

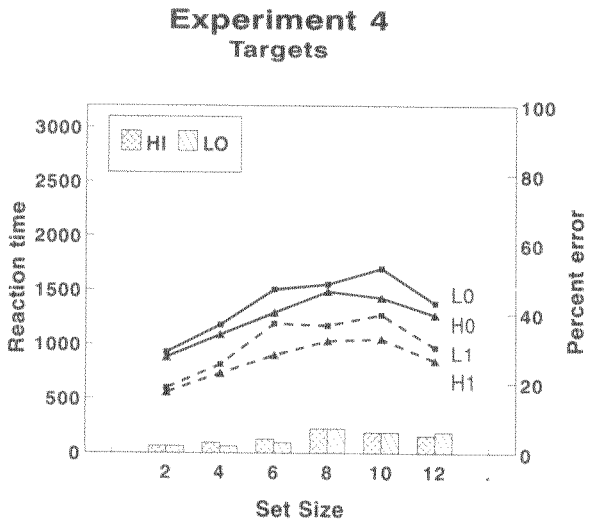

Figure 6. Mean verification times and error rates for Experiment 4 as a function of set size, delay, and WM span.

the 1-second delay function was not different for high- and low-span subjects. Again, this suggests that high- and low-span subjects do not differ in retrieval from secondary memory. Finally, unlike in Experiment 2, the difference between the 0-s delay function and the 1-s delay function is not constant across set size. The difference between the 0-s delay condition and the 1-s delay condition was smaller at a set size of 2 than at the other set sizes. When we performed the analysis on all set sizes except 2, the difference between the 0-s delay function and the 1-s delay function was constant across set size.

These conclusions were supported by a statistical analysis. Median RTs were submitted to a span (high, low) by set size (2, 4, 6, 8, 10, 12) by delay (0, 1,000) ANOVA. Span group was the only between-subjects variable. There were main effects for set size, $F(5, 190) = 54.48$, $p < .05$, $MS_e = 81{,}582$, and delay, $F(1, 38) = 355.54$, $p < .05$, $MS_e = 48{,}619$. The main effect for span was not significant, $F(1, 38) = 2.97$, $p = .093$. More important, there was no span by set size interaction, $F(5, 190) = 1.87$, $p > .10$. The span by delay interaction was not significant, either, $F(1, 38) < 1$. Unlike in Experiment 2, we found a set size by delay interaction, $F(5, 190) = 2.42$, $p = .038$, $MS_e = 11{,}716$. However, we conducted another ANOVA excluding the set size of 2, and the set size by delay interaction was not significant, $F(4, 152) = 1.39$, $p > .10$.

Error rates were relatively low and did not differ between high- and low-span subjects. Subjects were less accurate at the larger set sizes. As with the RT data, error rates were submitted to a span (high, low) by set size (2, 4, 6, 8, 10, 12) by delay (0, 1,000) ANOVA. There were main effects for set size, $F(5, 190) = 16.06$, $p < .05$, $MS_e = .002$, and delay, $F(1, 38) = 16.87$, $p < .05$, $MS_e = .002$. There was no main effect for span, $F(1, 38) < 1$. Also, none of the interactions were significant (for all, $p > .10$).

General Discussion

Let us first summarize the results of the four studies reported here: (1) With the exclusion of a set size of 2, the time to activate a set from secondary to primary memory was independent of the set's size. (2) WM capacity had no effect on the retrieval time from secondary memory. (3) WM capacity did have an effect on the time to search the set that was active in primary memory, but (4) only under conditions in which there was overlap in set membership, which thus engendered a level of response competition or interference. Surprisingly, in all four experiments the set size functions had significant quadratic trends as well as linear trends (for all, $p < .05$). In a typical memory-search task (e.g., Sternberg, 1966) the set-size function is linear. At first glance, we thought that the quadratic trend may have been due to the number of possible foils. For example, in Experiment 1 there were only two possible foils for Set 8. Therefore, it could be that subjects were responding to Set-8 targets by searching this "set" of two foils instead of comparing the probe with the members of Set 8. However, in Experiment 3, there were 12 possible foils for Set 8, and we found the same quadratic trend. It should also be noted that quadratic trends were also present in the learning data. It is not clear to us why we found these trends.

Are the Retrieval Characteristics From Primary and Secondary Memory the Same?

Wickens et al. (1981) argued that retrieval from secondary memory is a two-stage process, with the first step being the selection and activation of information about the entire set or list. The second step would involve the search or retrieval of information from the active set. Furthermore, Wickens et al. argued that the extra process necessary for retrieval from secondary memory required the same amount of time when the test item was from either a two- or four-item set. These conclusions were supported by their finding of no interaction between set size (two and four items) and memory condition (primary or secondary).

In three of our four experiments, we did find a set size by delay condition interaction when a set size of 2 was included in the analyses. When Set 2 was excluded from the analyses, however, there was no interaction between set size and delay condition in any of the four experiments. Therefore, beyond Set 2, our data support Wickens et al.'s (1981) conclusion that the processes involved in retrieval from primary memory are distinct from those involved in retrieval from secondary memory. The time required to make a set of 4 items available to primary memory is equal to the time required to make a set of 12 items available to primary memory. This is an important finding, because 4 items is presumably subspan, and 12 items is presumably supra-span for most of our subjects. The fact that the time to make a memory set available to primary memory did not interact with group (high–low span) further reinforces this finding and suggests that the retrieval from secondary memory is a result of an automatic process, not a controlled limited-

capacity process. This finding also speaks to the very nature of retrieval of a list or set of information from secondary or inactive memory. The process of bringing set information into a state of activation or primary memory seems to be more of an indexing or addressing of the list, rather than a process of retrieval of all of the individual list items. For example, it will not take any more time to access the set "50 states" than it will to access the set "12 months." The size of the set is not a factor until the individual items are searched in primary memory.

Why were the results of a set size of 2 different from the other set sizes? We remind the reader that, whereas Wickens et al. (1981) presented their sets one time, our subjects had considerable learning of the set items. The RTs for our set size of 2, no-delay condition, were extremely fast. It seems reasonable that this would occur if the subjects had been making their responses on the basis of a simple stimulus–response (S–R) connection rather than a search of the memory set. Instead of activating Set 2 and searching for a match, a subject may have simply remembered to press the "yes" key when a certain item was paired with the number 2. This seems more plausible for a set size of 2 than for the other sets, because it is easy for subjects to memorize a two-item set without making elaborate associations within that set. Therefore, when faced with the verification task of "Is *bird* a member of set 2?" they can rely on the connection 2–bird–yes instead of activating the entire set and searching for *bird*. As the memory sets become larger, it is less economical for subjects to use this strategy, because they would need to keep track of a large number of S–R connections. Therefore, the subjects would form associations among the memory set items and connect the digit cue to the set cue. Then, in the verification task, when faced with the question "Is *bird* a member of set 8?" the subjects activate the entire set, and activation is divided among all eight memory set items.

Thus, we argue that retrieval from primary memory and retrieval from secondary memory are quite different. Retrieval from primary memory is a function of set size, even with well-learned set information, whereas retrieval from secondary memory is not a function of set size. This lends support to the notion of a distinction between the two states of memory that is based on qualitative distinctions in level of activation. Primary memory is not just one end of a continuum of activation with secondary memory at the other end. There appears to be a qualitative and discrete distinction between retrieval from primary memory and retrieval from secondary memory.

Does Working Memory Capacity Influence Retrieval?

The second issue we address in this article is the role of WM in retrieval. First, in all four experiments, the delay by span group interaction was not significant. Low-span subjects were as fast as high-span subjects at performing the process necessary for retrieval from secondary memory. This is especially interesting because that conclusion holds for all four experiments. Therefore, according to our results, WM capacity does not play a role in the indexing or addressing of well-learned set information in secondary memory.

Where then, does WM capacity play a role in retrieval? The results are clear: We found a span by set size interaction when set items appeared as targets in two sets (Experiments 1 and 2) but not when they appeared in one set (Experiments 3 and 4). Therefore, high- and low-span subjects differ in their ability to retrieve information from primary memory when set membership overlaps but not when sets are unique. Why?

To answer this question, we must first consider how a subject performs a retrieval task when each memory set item appears as a target in only one set. Because high- and low-span subjects differ only in retrieval from primary memory, let us consider only primary-memory trials. Furthermore, assume that the word *dog* was a member of the set size of 6. During a trial that tested retrieval from primary memory, the memory set number 6 was presented 1 s before the probe item *dog*. Therefore, when the memory set number appears, activation spreads from the memory representation for 6 to the cue for the set of six items, leading to indexing of the set. Then, when the probe item *dog* appears, activation spreads from the memory representation for 6 to the set items, including *dog*, and from *dog* back to the memory representation for 6, and a match occurs.

Now consider a similar trial in the experiments in which each memory set item appeared as a target in two different sets. Assume that the word *dog* appeared in Set 6 and Set 10. Again, the memory set number is presented 1 s before the probe item. When the number 6 appears, this leads to the activation of the set information. Then activation spreads from the memory representation for 6 to the memory representations for all members of that set, including *dog*. Then, when the probe *dog* appears, activation spreads from the memory representation for *dog* back to the appropriate memory set number 6 and to the inappropriate set number 10. Therefore, when the sets overlap, activation spreads to irrelevant information.

How did this diffusion of activation lead to the differences between high- and low-span subjects in Experiments 1 and 2? The data from all four experiments were collapsed over delay and combined for the letters experiments (Experiments 1 and 3), shown in Figure 7, and the words experiments (Experiments 2 and 4), shown in Figure 8. The data in both figures show that high-span subjects were not as affected by the overlapping sets as were low-span subjects. The results are most clear in Figure 7, in which the curves for the high-span subjects in Experiments 1 and 3 are nearly identical. The high-span subjects were relatively unaffected by the manipulation of overlap or uniqueness of set membership. Figure 7 also shows that the RT function for low-span subjects in Experiment 3—the study with no overlap in set membership—was also nearly identical to the high-span functions for the two experiments. The real outlier in this figure is the function for the low-span subjects in Experiment 1. Low-span subjects were markedly slowed by

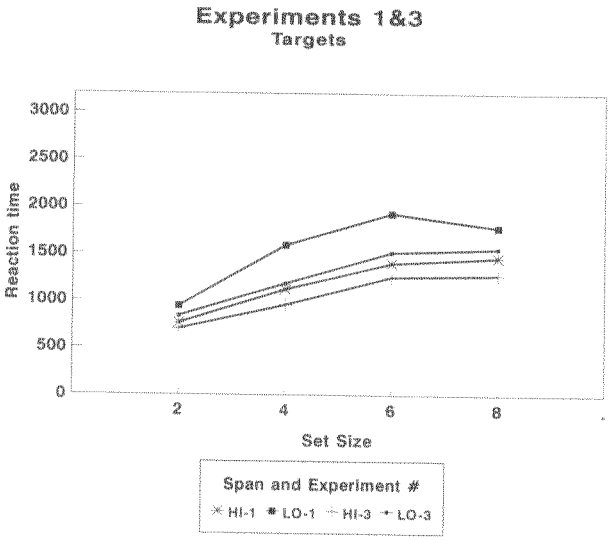

Figure 7. Mean verification times for Experiments 1 and 3 as a function of group, set size and experiment. H and L refer to high and low span subjects. 1 and 3 refer to Experiments 1 and 3.

the overlap in set membership, and the effect increased with set size.

We draw a similar conclusion from Figure 8. Although there does appear to be some slight effect of overlap in set membership for high-span subjects, the effect is much greater for the low-span subjects.

We assume that, in Experiments 3 and 4, in which there was no overlap in set membership, both high- and low-span subjects performed the verification task by relying totally on automatic processes. Because limited-capacity controlled attention was not required for this task, no differences were observed between high- and low-span subjects.

In Experiments 1 and 2 we assume that the automatic spread of activation occurred from probe set cue to the word or letter item and from the item to the probe set cue. Activation also spread from the item to the nontested competing set cue, as described above. This latter link would be weaker than the links between the probed cue and the item; thus, it would be susceptible to inhibition. We assume that high-span subjects use their available attentional resources to prevent or inhibit the activation of this weaker and irrelevant information.

In contrast, low-span subjects were unable to inhibit activation of the irrelevant and weaker information, which would ultimately allow that information to also come into an active state. This would explain why low-span subjects were more slowed in experiments in which each memory set item appeared as a target in two sets. Recall the instance in which *dog* was a member of Set 6 and also of Set 10. It is our argument that, because low-span subjects could not inhibit the spread from *dog* along the weaker link *dog*–10, they would ultimately be faced with the dilemma that the system would be providing conflicting information, namely, that *dog* was a member of both Set 6 and Set 10. This would possibly lead to the low-span subjects performing a serial search of Set 6, similar to what Sternberg (1966) originally proposed.

Therefore, high- and low-span subjects differ in retrieval from primary memory because of their differing ability to inhibit irrelevant information. It is important to note, however, that that ability occurs because we assume that inhibition is resource demanding and that subjects differ in the attentional resources available to them. Thus, subjects with greater attentional resources also have greater capacity for inhibiting information that is irrelevant to the task.

A Resource-Dependent View of Inhibition

Unfortunately, the term *inhibition* is a nebulous one that connotes a multitude of meanings. We would like to elucidate the manner in which we use that term in this article.

Bjork (1989) argued that there are two ways in which the retrieval of an item can affect an attempt to retrieve a different item later, and both are often referred to as inhibition. He referred to one mechanism as *suppression* and to the other as *blocking*. He described suppression as a stronger, more active type of inhibition in that it is directed at the to-be-inhibited information, and it is initiated to achieve some goal. It is also characterized by the actual lowering of activation of a weaker or irrelevant item. An example of this type of retrieval inhibition is illustrated by the work of Postman, Stark, and Fraser (1968). Using the A-B, A-D paradigm, Postman et al. showed that retroactive interference is a result of subjects' suppression of first-list responses during second-list learning. Importantly, they found suppression to occur for the entire set, not at the level of individual associations. Thus, active suppression of first-list responses during second-list learning facilitated learning of the second list. Therefore, the inhibition was initiated to

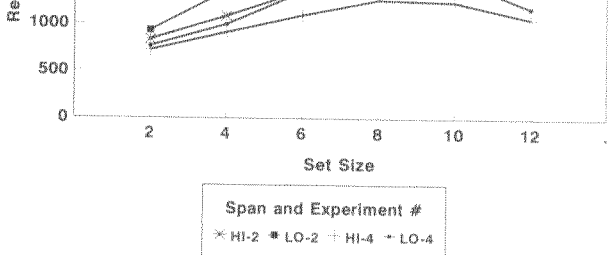

Figure 8. Mean verification times for Experiments 2 and 4 as a function of group, set size and experiment. HI and LO refer to high and low span subjects. 2 and 4 refer to Experiments 2 and 4.

achieve some goal, and it was directed at the to-be-inhibited information. We argue that this type of inhibition is resource demanding and would occur to the extent that a subject had available resources to do so. It would be less likely to occur under conditions that directed a subject's attentional resources elsewhere or when subjects themselves have smaller resource capacities.

Blocking was the other type of retrieval inhibition Bjork (1988) discussed. He described blocking as a by-product of the increased activation of other memorized information. Brown (1968) illustrated an example of blocking. In Brown's experiment, subjects who spent 5 min studying the names of 25 of the 50 states later recalled more of those 25 states but fewer of the other 25 states than did subjects in a control group that spent those 5 min doing light reading. This type of inhibition is characterized as blocking because the experimental subjects did not intend to inhibit the names of the 25 states that were not studied. In fact, the inhibition was counterproductive to the goal of recalling all 50 states. Retrieval difficulty results not from the actual lowering of the activation of nonstudied items but from the increase in activation of studied items. This type of inhibition also occurs in the search of associative memory (SAM) model of retrieval (Raaijmakers & Shiffrin, 1981).

We argue that the reason we find a difference between high- and low-span subjects in Experiments 1 and 2 is because of the subjects' differing ability to suppress irrelevant information. We feel that suppression is a better candidate for the retrieval inhibition than blocking for two reasons. First, suppression is directed at the to-be-inhibited information, and it is initiated to achieve some goal, whereas blocking is a by-product of the increased activation of other memorized information and is often counterproductive to the goal. In our task, subjects benefited from inhibiting a specific irrelevant set cue–item association. Therefore, if the inhibition mechanism is successful, it is directed and is initiated to achieve a goal.

Another reason we favor suppression over blocking is because the blocking hypothesis seems to be consistent with our original (and now clearly wrong) view that high- and low-span subjects simply differ in the level of activation available to the system (Engle, Cantor, & Carullo, 1992). Consider a blocking explanation of the present results. Again, assume that *dog* is a member of Set 6 and Set 10. When the set cue 6 is presented, activation automatically spreads from the representation for the set cue to the representations for all items of that set. When the probe *dog* is presented, activation spreads from the probe representation back to the appropriate set cue as well as to the irrelevant set cue 10. If blocking occurs and high-span subjects are better blockers than low-span subjects, then we have to assume that the relevant set cue–item association 6–*dog* is more highly activated for high-span subjects than for low-span subjects. This assumption is necessary because blocking is a by-product of activated representations. Relevant representations will block irrelevant representations better if the relevant representations are more activated.

Therefore, if blocking occurs, then high-span subjects are better blockers than low-span subjects. This would mean relevant set cue–item associations become more highly activated for high- span subjects than for low-span subjects. This argument predicts that we should see retrieval differences between high- and low-span subjects even when there are no associations to be blocked (e.g., Experiments 3 and 4). We did not observe retrieval differences in Experiments 2 and 4.

For these reasons, we feel that suppression is a better candidate than blocking for the retrieval inhibition that occurred in Experiments 1 and 2. However, the present study does not rule out the possibility that blocking occurred. Future research will be needed to establish what conditions lead to blocking and what conditions lead to suppression.

The work presented here may have implications for domains of research beyond adult models of WM and retrieval. For instance, the resource-dependent view of inhibition is possibly also relevant to the literature on developmental differences in WM capacity. Hasher and Zacks (1988) argued that when WM, as well as selective attention, are functioning normally, inhibitory mechanisms "serve to limit entrance into working memory, information that is along the 'goal path' of comprehension" (p. 212). Therefore, people with inefficient inhibitory mechanisms will allow information that is off the goal path to enter WM. Hasher and Zacks (1988) provided considerable evidence that WM deficits in aging individuals result from reduced ability to inhibit irrelevant information. It is possible that the reduced ability to inhibit as we get older is a result, in turn, of reduced attentional resources. It will no doubt be a matter of debate whether suppression is a result of the automatic spreading of inhibition or is a result of controlled, limited-capacity attention.

The work reported in this article is consistent with Ackerman's (1988) theory of skill acquisition. The theory predicts that general abilities such as WM capacity cease to be a factor after extensive learning if the task involves a consistent relationship between the stimulus elements and the necessary responses. General abilities remain important, even with extensive learning, only when task conditions lead to a variable relationship between stimulus and response elements. In contrast, our earlier view was that individuals differ in their overall level of activation, which should not depend on whether a task was performed under automatic or controlled conditions. Our results show that individual differences in WM capacity were important to retrieval of well learned information only when the task caused a degree of conflict or interference. The activation notion predicted that WM-capacity differences would be important even in the absence of such interference conditions. Quite clearly, the Ackerman theory has been supported by our research, and the theory that individuals differ in overall level of activation is wrong.

Our new view, that individual differences in WM capacity reflect differences in attentional resources, appears consistent with Ackerman's (1988) theory. Both views argue that individual differences in WM capacity will be important to a task only during phases that require the task be performed under controlled, limited-capacity attention.

Consistent with this view, we did not find individual differences in WM capacity to be a factor in Experiments 3 and 4. Therefore, we argue that performance in these experiments required automatic processing only. However, some memory search studies have suggested that automaticity is achieved when the slope of the set-size function becomes flat (Shiffrin & Schneider, 1977). The slope of the set-size functions in Experiments 3 and 4 are not flat. Therefore, one might argue that limited-capacity controlled attention should play a role in these tasks. However, parallel-search models (Anderson, 1983; Ratcliff, 1978) define search as an automatic spread of activation from the memory set cue to the memory set items. According to this view, set-size effects will still obtain while a subject is engaged in automatic processing. Set-size effects occur because activation is divided among the number of memory set items, such that less activation will be available per item for larger sets. The slope of the set-size effect will become flat only when a subject responds on the basis of a simple S–R association such that when the subject encounters a set cue and a single item it automatically leads to the response of "yes" (Logan, 1988). When this occurs, the subject is no longer searching the memory set. We argue that this process is qualitatively different from responding by means of spreading activation, yet we view both processes as automatic.

We would be remiss here if we did not mention the relationship of our findings to Baddeley's (1986) theory of WM. The view with which we end seems quite consistent with the idea of a central executive as being a general attentional system. We would add that the general attentional system is important to the inhibition of superfluous information as well as for the activation and maintenance of information relevant to the task. Thus, those knowledge units represented in the active portion of memory would be affected not only by the ability to keep information active but also by the ability to prevent irrelevant and competing information from interfering with active memory.

References

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 3,* 288–318.

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 6,* 451–474.

Anderson, J. R. (1976). *Language, memory, and thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Baddeley, A. D. (1986). *Working memory.* New York: Oxford University Press.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness* (pp. 309–330). Hillsdale, NJ: Erlbaum.

Brown, J. (1968). Reciprocal facilitation and impairment of free recall. *Psychonomic Science, 10,* 41–42.

Cantor, J., & Engle, R. W. (1993). Working memory capacity as long-term memory activation: An individual-differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1101–1114.

Cantor, J., Engle, R. W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence, 15,* 229–246.

Carpenter, P. A., & Just, M. A. (1989). The role of working memory in language comprehension. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 31–68). Hillsdale, NJ: Erlbaum.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin, 104,* 163–191.

Craik, F. I. M., & Levy, B. A. (1976). The concept of primary memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 4, pp. 133–171). Hillsdale, NJ: Erlbaum.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

Engle, R. W. (in press). Individual differences in memory and their implications for learning. In R. Sternberg (Ed.), *Encyclopedia of Intelligence.* New York: Macmillan.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 972–992.

Engle, R. W., Nations, J. K., & Cantor, J. (1990). Is working memory capacity just another name for word knowledge? *Journal of Educational Psychology, 82,* 799–804.

Ericsson, K. A., & Staszewski, J. J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 235–267). Hillsdale, NJ: Erlbaum.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–225). New York: Academic Press.

Jones, W. P., & Anderson, J. R. (1987). Short- and long-term memory retrieval: A comparison of the effects of information load and relatedness. *Journal of Experimental Psychology: General, 116,* 137–153.

LaPointe, L. B., & Engle, R. W. (1990). Simple and complex word spans as measures of working memory capacity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1118–1133.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 4,* 492–527.

McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition.* Cambridge, MA: MIT Press.

Postman, L., Stark, K., & Fraser, L. (1968). Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior, 7,* 672–694.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Schneider, W. (1988). Micro-Experimental Laboratory: An integrated system for IBM PC compatibles. *Behavior Research Methods, Instruments, and Computers, 20,* 206–217.

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). New York: Academic Press.

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.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153,* 652–654.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28,* 127–154.

Wickens, D. D., Moody, M. J., & Dow, R. (1981). The nature and timing of the retrieval process and of interference effects. *Journal of Experimental Psychology: General, 110,* 1–20.

Appendix A

Experiment 1 Memory Sets

2	4	6	8
V	Z	V	X
N	G	W	M
	R	G	B
	B	M	Z
		Q	N
		X	W
			Q
			R

Experiment 2 Memory Sets

2	4	6	8	10	12
snow	doll	feet	iron	legs	snow
bike	goat	lion	book	neck	doll
	ship	cake	body	wing	feet
	moon	food	head	duck	iron
		king	ball	head	legs
		bird	bird	cake	neck
			moon	ship	book
			king	bike	lion
				ball	goat
				food	wing
					body
					duck

Experiment 3 Memory Sets

2	4	6	8
C	W	J	D
S	G	R	Q
	H	N	K
	X	B	L
		F	M
		T	P
			V
			Z

Experiment 4 Memory Sets

2	4	6	8	10	12
snow	feet	book	head	king	bowl
doll	iron	lion	cake	hair	bean
	legs	goat	ship	toes	shoe
	neck	wing	bike	nose	corn
		body	ball	door	face
		duck	food	fish	tail
			moon	flag	eggs
			bird	lamp	frog
				eyes	pigs
				tree	cage
					meat
					girl

(*Appendixes continue on next page*)

Appendix B

Experiment 1: Foils Analyses

Table B1
Reaction time: Means as a Function of Span, Delay (in Milliseconds), and Set Size

	High span (ms)		Low span (ms)	
Set size[a]	0	1,000	0	1,000
2	1,022	619	1,324	763
4	1,583	994	2,005	1,434
6	2,189	1,649	2,882	2,069
8	1,979	1,467	2,222	1,810
Main Effects				
Span	$F(1, 38) = 3.49$, $p = .069$, $MS_e = 3{,}240{,}631$			
Set Size	$F(3, 114) = 42.422$, $p < .05$, $MS_e = 554{,}442$			
Delay	$F(1, 38) = 421.551$, $p < .05$, $MS_e = 57{,}427$			
Interactions				
Set Size × Span	$F(3, 114) = .789$, $p > .10$, $MS_e = 554{,}442$			
Delay × Span	$F(1, 38) = 2.122$, $p > .10$, $MS_e = 57{,}427$			
Set Size × Delay	$F(3, 114) = 3.893$, $p < .05$, $MS_e = 50{,}464$			
Set Size × Delay × Span	$F(3, 114) = 2.805$, $p = .043$			
Accuracy				
Main Effects				
Span	$F(1, 38) = 0.073$, $p > .10$, $MS_e = 0.015$			
Set Size	$F(3, 114) = 25.379$, $p < .05$, $MS_e = 0.005$			
Delay	$F(1, 38) = 0.147$, $p > .10$, $MS_e = 0.002$			
Interactions				
Set Size × Span	$F(3, 114) = 1.561$, $p > .10$, $MS_e = 0.005$			
Delay × Span	$F(1, 38) = .551$, $p > .10$, $MS_e = 0.002$			
Set Size × Delay	$F(3, 114) = .500$, $p > .10$, $MS_e = 0.002$			
Set Size × Delay × Span	$F(3, 114) = .748$, $p > .10$			

[a] Number of items.

Experiment 2: Foils Analyses

Table B2
Reaction time: Means as a Function of Span, Delay (in Milliseconds), and Set Size

	High span (ms)		Low span (ms)	
Set size[a]	0	1,000	0	1,000
2	1,135	668	1,289	835
4	1,479	979	1,806	1,231
6	1,940	1,439	2,535	1,976
8	2,385	1,747	2,724	2,243
10	2,629	1,901	3,032	2,351
12	1,894	1,553	2,500	1,816
Main Effects				
Span	$F(1, 38) = 2.772$, $p > .10$, $MS_e = 6{,}333{,}470$			
Set Size	$F(5, 190) = 48.304$, $p < .05$, $MS_e = 522{,}720$			
Delay	$F(1, 38) = 186.552$, $p < .05$, $MS_e = 195{,}081$			
Interactions				
Set Size × Span	$F(5, 190) = 0.747$, $p > .10$, $MS_e = 522{,}720$			
Delay × Span	$F(1, 38) = 0.293$, $p > .10$, $MS_e = 195{,}081$			
Set Size × Delay	$F(5, 190) = 1.61$, $p > .10$, $MS_e = 84{,}328$			
Set Size × Delay × Span	$F(5, 190) = 1.693$, $p > .10$			
Accuracy				
Main effects				
Span	$F(1, 38) = 0.01$, $p > .10$, $MS_e = 0.018$			
Set Size	$F(5, 190) = 33.225$, $p < .05$, $MS_e = 0.006$			
Delay	$F(1, 38) = 3.195$, $p = .082$, $MS_e = 0.003$			
Interactions				
Set Size × Span	$F(5, 190) = 1.474$, $p > .10$, $MS_e = 0.006$			
Delay × Span	$F(1, 38) = 0.210$, $p > .10$, $MS_e = 0.003$			
Set Size × Delay	$F(5, 190) = 0.392$, $p > .10$, $MS_e = 0.003$			
Set Size × Delay × Span	$F(5, 190) = 1.756$, $p > .10$			

[a] Number of items.

Experiment 3: Foils Analyses

Table B3
Reaction time: Means as a Function of Span, Delay (in Milliseconds), and Set Size

	High span (ms)		Low span (ms)	
Set size[a]	0	1,000	0	1,000
2	1,066	564	1,222	735
4	1,394	781	1,670	1,154
6	1,700	1,066	2,125	1,512
8	1,769	1,186	2,063	1,460

Main effects	
Span	$F(1, 38) = 4.582$, $p = .039$, $MS_e = 1{,}593{,}015.238$
Set Size	$F(3, 114) = 68.389$, $p < .01$, $MS_e = 136{,}696.660$
Delay	$F(1, 38) = 360.356$, $p < .01$, $MS_e = 71{,}826.096$
Interactions	
Set Size × Span	$F(3, 114) = 1.834$, $p > .10$, $MS_e = 136{,}696.66$
Delay × Span	$F(1, 38) = 0.223$, $p > .10$, $MS_e = 71{,}826.096$
Set Size × Delay	$F(3, 114) = 2.249$, $p = .086$, $MS_e = 26{,}937.945$
Set Size × Delay × Span	$F(3, 114) = 0.455$, $p > .10$

Accuracy	
Main effects	
Span	$F(1, 38) = 0.880$, $p > .10$, $MS_e = 0.014$
Set Size	$F(3, 114) = 30.327$, $p < .01$, $MS_e = 0.004$
Delay	$F(1, 38) = 8.101$, $p < .01$, $MS_e = 0.003$
Interactions	
Set Size × Span	$F(3, 114) = 0.255$, $p > .10$, $MS_e = 0.004$
Delay × Span	$F(1, 38) = 0.639$, $p > .10$, $MS_e = 0.003$
Set Size × Delay	$F(3, 114) = 3.051$, $p = .031$, $MS_e = 0.002$
Set Size × Delay × Span	$F(3, 114) = 0.696$, $p > .10$

[a] Number of items.

Experiment 4: Foils Analyses

Table B4
Reaction time: Means as a Function of Span, Delay (in Milliseconds), and Set Size

	High span (ms)		Low span (ms)	
Set size[a]	0	1,000	0	1,000
2	1,012	616	1,127	687
4	1,253	770	1,389	938
6	1,420	939	1,547	1,070
8	1,533	1,024	1,708	1,233
10	1,520	1,009	1,614	1,234
12	1,455	932	1,565	1,099

Main effects	
Span	$F(1, 38) = 2.22$, $p > .10$, $MS_e = 1{,}122{,}711$
Set Size	$F(5, 190) = 62.146$, $p < .05$, $MS_e = 47{,}646$
Delay	$F(1, 38) = 383.882$, $p < .05$, $MS_e = 67{,}893$
Interactions	
Set Size × Span	$F(5, 190) = 0.464$, $p > .10$, $MS_e = 47{,}646$
Delay × Span	$F(1, 38) = 0.562$, $p > .10$, $MS_e = 67{,}893$
Set Size × Delay	$F(5, 190) = 0.95$, $p > .10$, $MS_e = 18{,}245$
Set Size × Delay × Span	$F(5, 190) = 0.925$, $p > .10$

Accuracy	
Main effects	
Span	$F(1, 38) = 0.029$, $p > .10$, $MS_e = 0.01$
Set Size	$F(5, 190) = 16.805$, $p < .05$, $MS_e = 0.002$
Delay	$F(1, 38) = 2.957$, $p = .094$, $MS_e = 0.002$
Interactions	
Set Size × Span	$F(5, 190) = 1.293$, $p > .10$, $MS_e = 0.002$
Delay × Span	$F(1, 38) = 0.169$, $p > .10$, $MS_e = 0.002$
Set Size × Delay	$F(5, 190) = 1.355$, $p > .10$, $MS_e = 0.002$
Set Size × Delay × Span	$F(5, 190) = 1.178$, $p > .10$

[a] Number of items.

Received November 30, 1993
Revision received June 20, 1994
Accepted July 5, 1994 ■