Journal of Experimental Psychology:
Learning, Memory, and Cognition
1993, Vol. 19, No. 5, 1101–1114

0278-7393/93/$3.00

Working-Memory Capacity as Long-Term Memory Activation: An Individual-Differences Approach

Judy Cantor and Randall W. Engle

One explanation of the correlation often observed between working-memory span scores and reading comprehension is that individuals differ in level of activation available for long-term memory units. Two experiments used the fan manipulation to test this idea. In Experiment 1, high- and low-working-memory Ss learned a set of unrelated sentences varying in the number of shared concepts (fan) and then performed speeded recognition for those sentences. Low-working-memory Ss showed a larger increase in recognition time as fan increased. When the slope of the fan effect was partialed out of the relationship between working-memory span and verbal abilities, the relationship was reduced to nonsignificance. In Experiment 2, Ss learned thematically related sentences that varied in fan. Low-span Ss showed the positive fan effect typically found with thematically unrelated sentences, whereas high-span Ss showed a negative fan effect. The results are discussed in terms of a general capacity theory.

Most models of human memory distinguish between previously learned knowledge that resides in long-term memory (LTM) and the temporary activation of that knowledge, which represents the contents of current thought. Working memory (WM) is often defined as that part of permanent LTM that is temporarily active above some critical threshold and that can be recognized and manipulated by ongoing cognitive processes. Within the information-processing system, WM is the arena where processing and storage interact (Anderson, 1983a; Bower, 1975; Cantor, Engle, & Hamilton, 1991; Card, Moran, & Newell, 1986; Cowan, 1988; Engle, Cantor, & Carullo, 1992; Just & Carpenter, 1992; Shiffrin, 1975).

General Capacity Model

The general capacity model of WM was recently developed by Engle et al. (1992) to account for individual differences in WM. It assumes a single working memory capacity that transcends a variety of cognitive representations and is implicated whenever information must be temporarily maintained (Cantor et al., 1991; Engle et al., 1992; Engle, Nations, & Cantor, 1990). Shute (1991) demonstrated that quantitative, verbal, and spatial WM tasks reflected a single cognitive factor and that this factor was a better predictor of learning than processing speed, general knowledge, or technical skill. There is also evidence for a common capacity that underlies auditory and visual WM spans (Daneman & Carpenter, 1980) and for those that require problem solving, reasoning, or reading (Kyllonen & Christal, 1990; Salthouse, Mitchell, Skovronek, & Babcock, 1989; Turner & Engle, 1989). According to the general capacity model, these results indicate that WM is a single unitary resource.

It is also assumed that the contents of WM is information in LTM that has been stimulated or activated above some critical threshold (Anderson, 1983a; Bower, 1975; Cowan, 1988; Just & Carpenter, 1992). As with Anderson's (1983a, 1983b, 1983c) ACT* model, activation is a resource that automatically spreads or otherwise divides among related concepts and is limited. The activation associated with a single concept can build or sum, and is not all or none. As the activation level of a concept increases, so does its accessibility. At some point, this activation will increase to a level that allows the corresponding concept to be recognized by cognitive processes or procedures. This concept has then crossed the critical threshold and is said to reside in WM (Anderson, 1983a, 1983b, 1983c; Just & Carpenter, 1992). It is not necessarily at the level of consciousness or in the focus of attention, but it is accessible to procedures and processes and can therefore be operated on.

Throughout LTM, concepts vary in their level of activation. There may be many concepts that have above-resting-state activation levels without crossing the critical threshold into WM. Although these items cannot be operated on, they are consuming some of the activation available to the system. Because activation is limited, the amount of this resource that is distributed among subthreshold items will influence the number of concepts that can be activated above threshold. The major premise of this model is that LTM activation is limited and that this limit drives what is commonly called WM capacity.

Once an item is active above threshold, it can be used to initiate a cognitive procedure. This relationship between processing and above-threshold information functions much like

This research was submitted to the University of South Carolina in partial fulfillment of the requirements for the Doctor of Philosophy degree by Judy Cantor. This work was partially supported by Grant HD27490 from the National Institute of Child Health and Human Development. We thank Heather Carlson for testing subjects and Robin Morris, Gloria Miller, Cristine McCormick, and Jerome L. Myers for their helpful comments.

Correspondence concerning this article should be addressed to either Judy Cantor, AT&T Bell Laboratories, 101 Crawfords Corner Road, Room 2D602, Holmdel, New Jersey, 07733-3030, or Randall W. Engle, Department of Psychology, University of South Carolina, Columbia, South Carolina 29208.

the recognition-act cycles described by Just and Carpenter (1987) or like Anderson's pattern matching mechanism. A procedure will operate or fire when all the knowledge necessary to meet its conditions is above threshold. The only demand for activation that occurs during these operations is the stimulation of information to be temporarily maintained, such as the partial and final products of processes, goals, gists generated during reading, or mental models derived about particular situations. However, the development of new procedures or processes, or those that are poorly learned, will consume activation because they are novel and are thus represented in declarative memory.

Nature of Individual Differences

There is now substantial evidence that WM differs among people, and that this difference is manifest in a wide range of ecologically important cognitive tasks. Problem solving, reasoning, reading comprehension, acquiring new vocabulary words, learning to spell, following directions, and taking notes in the classroom are all influenced by individual differences in WM (Engle, in press).

The first measure to show important differences in WM was Daneman and Carpenter's (1980) reading-span task. Subjects read a series of unrelated sentences while attempting to maintain the last word of each sentence. Span size was defined as the largest number of sentences from which at least 50% of the last words were recalled. Daneman and Carpenter found that these scores significantly predicted subjects' performance on a variety of other reading-related tasks and indicated that individual differences in WM are driven by differences in processing skill. When reading, good readers have fast and efficient reading processes that require less WM capacity than those of poor readers. Thus, good readers have functionally more capacity in reading-related tasks. We call this approach the *task-specific view*.

This task-specific view was challenged when Turner and Engle (1986, 1989) showed that performance on a nonreading-WM span also predicted reading measures. They developed a task in which subjects solved simple math problems while maintaining unrelated words. This operation–word span predicted comprehension and verbal skills (Verbal Scholastic Aptitude Test; VSAT) to the same extent as the reading span even when math skills (Quantitative Scholastic Aptitude Test) were partialed out. These results, and many subsequent replications, indicate that individual differences in WM are not driven by task-specific processing skills (Engle et al., 1992; Kyllonen & Christal, 1990; La Pointe & Engle, 1990; Salthouse et al., 1989; Shute, 1991).

A more recent version of the task-specific view assumes that there are general skills that are used in any task requiring the manipulation of language (Daneman, 1987; Daneman & Tardif, 1987; Just & Carpenter, 1992). Still, however, the focus is on processing efficiency, not on storage. Carpenter and Just (1989) examined eye fixations of subjects while they performed the reading-span task. One of their measures was lexical access, a parameter derived from the slope of the function relating gaze time and word frequency. High-WM-span subjects showed significantly faster lexical access times than low-span subjects, especially as the WM load increased. According to Carpenter and Just, high-span subjects are more efficient at retrieving information from LTM and at allocating their resources to meet the demands of the task. As a result, high-span subjects have more residual capacity to store the words to be remembered in the span task. They found that high-span subjects spent more time gazing at words to be remembered than low-span subjects, which supported their view.

Carpenter and Just (1989) argued that processing efficiency is the important parameter of the WM spans that drives the correlation between span tasks and more global tasks. However, they did not test whether their lexical access parameter statistically captured all of the covariance between WM span and other complex tasks, and a more direct test of this notion found little support for the processing-efficiency view (Engle et al., 1992).

Engle et al. (1992) conducted a series of experiments that tested whether the time subjects spent processing the WM spans accounted for the relationship between the number of words recalled in these tasks and VSAT. They examined a version of the reading span and the operation–word span developed by Turner and Engle (1989). Subjects paced themselves through components of each task by means of a moving window technique. The time between keypresses was recorded. After moving through several sentence-word or operation-word items, subjects attempted to recall the words. Also included in these experiments were storage-free tasks that did not require subjects to maintain words for subsequent recall.

Engle et al. (1992) found that high-span subjects spent more time on the first and last components of WM tasks and on the to-be-remembered words. A similar pattern was found for the storage-free processing tasks. However, the important test for the processing-efficiency model was whether the time spent on the sentences or operations accounted for the correlation between span (i.e., number of subsequent words recalled) and VSAT. The correlation between span and VSAT did not change when the time spent on the storage-free measures was removed. Thus, processing efficiency alone did not capture the individual differences in WM. More important, when they looked at the span tasks (those that required both processing and storage) they also found that the correlation between span and VSAT did not change when processing time was partialed out.

According to Engle et al. (1992), individual differences in WM are not driven by processing efficiency. Instead, people differ in the total amount of activation available to retrieve information from LTM. This difference will be manifest in any task that makes at least moderate demands for such activation. High- and low-capacity subjects, as indexed by the WM spans, actually differ in their activation limits. The WM spans are considered to be a good measure of this limit because they require a switching of attention between reading or solving math problems and maintaining words. It is this attention-switching feature that is a critical component of the WM spans. In the operation–word span task, for example, subjects solve several math problems before they attempt to recall the words. It is assumed that for each operation–word pair of a trial, subjects attend to an operation and then shift attention to the to-be-remembered word. At this time, they may also reactivate other members of the to-be-remembered set. The presentation of a new operation–word item requires that attention be shifted again, this time away from the word set to the new operation. As a result, activation of the word set drops below the WM threshold. Final recall then involves reaccessing this set from LTM. This attention switching characterizes all of the WM tasks reported to date and provides a way to measure the activation limits that constrain retrieval from LTM.

LTM Activation: Empirical Findings and Implications for WM

In the literature on retrieval from LTM, one ubiquitous finding is that activating a concept in LTM will result in the automatic activation of related concepts (Meyer & Schvaneveldt, 1971; Neely, 1991). Within the architecture of the general capacity model, this is an important finding. It implies that to the extent that activation is limited, the number of related associations automatically activated in LTM will determine the amount of activation available to bring any one concept above threshold into WM. Therefore, increasing the number of associations in LTM should decrease the speed or probability that any one item will be retrieved. An elegant test of this notion is Anderson's (1974, 1976) fan effect.

In the typical fan-effect procedure, subjects learn simple sentences such as "the LAWYER is in the PARK." The number of sentences that share concepts (LAWYER and PARK) is experimentally manipulated. By varying the number of common concepts, it is assumed that the size of the propositional networks developed in LTM will also vary. Anderson (1974) called this the propositional fan size. After learning sentences, subjects are given a speeded verification task in which studied sentences and related foils are tested. Anderson (1974, 1976; Lewis & Anderson, 1976) showed that reaction times (RTs) and error rates are greater to sentences associated with a large fan size than to those associated with a small fan. This is the fan effect. It should be noted that these results are not due to learning differences across sentences. Even with extended practice, over 25 days, the fan effect remains (Pirolli & Anderson, 1985).

The original ACT model placed the locus of fan effects on the rate of spreading activation (Anderson, 1974, 1976; Lewis & Anderson, 1976; King & Anderson, 1976; Neely, 1991; Thorndyke & Bower, 1974). Activation spread slower when propositional fan size was large, because there were more concepts that automatically received some of the limited activation. However, more recent data demonstrate that retrieval time does not reflect the speed with which activation spreads. In a priming study, Ratcliff and McKoon (1981) found that with less than a 100-ms delay, there was no priming for either strongly or weakly related primes. For both types of primes, the onset of facilitation was about 100 ms, which indicates that activation spread at the same rate independent of the strength of association between primes and targets. If spreading activation caused priming, then stronger primes should have shown facilitation earlier than weaker primes. What seemed to determine differences in retrieval time was the total amount of activation an item received once activation had spread, not the speed of this spread.

Anderson (1983c) used these results to reinterpret the fan effect. He argued that as fan size increases, each related proposition receives less activation. Cognitive processes or procedures operate on active information, and when activation is low, pattern matching will be more difficult and will require more time. Most important, what drives the rate and probability of retrieval is not the rate of spreading activation but the total amount of LTM activation that is available to the system.

The fan effect also represents an interesting paradox in information processing. It suggests that the more an individual knows about a topic, the harder it will be to retrieve this knowledge (Smith, Adams, & Schorr, 1978). Therefore, experts should have more difficulty retrieving information in their area of expertise than would novices. Clearly, this is not the case (Chase & Simon, 1973; Ericsson & Staszewski, 1989). How then do we explain the fan effect?

Typically, sentences in the fan-effect task share concepts but are not related by theme and cannot be integrated in any coherent way. In this case each concept probably represents an individual proposition, and the fan effect occurs in the way described by Anderson (1983c). However, when the materials to be studied are thematically related, different results obtain. Either there is no fan effect at all or the fan effect is negative. Retrieving a sentence from a large propositional network may actually be faster than retrieving a sentence from a small network when the stimulus sentences are thematically related (Myers, O'Brien, Balota, & Toyofuku, 1984; Reder, 1982; Reder & Anderson, 1980; Reder & Ross, 1983; Smith et al., 1978; Whitlow, Smith, & Medin, 1982).

One way to account for attenuated fan effects with related materials is that subjects can produce a higher order LTM unit, much like a schema or general theme, and they can make their verifications on the basis of how consistent the test sentence is with the theme (Reder, 1982; Reder & Anderson, 1980; Reder & Ross, 1983). Rather than waiting for activation to spread among individual propositions, verification is based on the activation of a single unit. Reder and Anderson showed that with thematically related material, there was no typical fan effect, but there was a theme effect. The RTs were greater to sentences that were consistent with many themes than to sentences that were consistent with few themes. In another test, Reder and Anderson developed foil sentences that were not presented during study but were consistent with previously studied themes. It was assumed that such foils would force subjects to search beyond the level of a general schema and to search each studied sentence. Under these conditions, the fan effect reappeared, thereby supporting Reder and Anderson's *consistency judgment* hypothesis.

Although consistency judgments did account for several reports of attenuated fan effects, there is contradictory evidence that casts doubt on a consistency interpretation. For example, Radvansky and Zacks (1991) presented sentences that could co-occur in the real world, such as "the display case is in city hall" and "the potted palm is in city hall." These sentences are not likely to generate schemas, yet Radvansky and Zacks found attenuated fan effects.

Further evidence against the consistency judgment explanation comes from a series of studies by Myers et al. (1984). They varied the temporal and causal relationship among sentences, while using the kinds of controls that Reder and Anderson (1980) argued would eliminate consistency judgments. Foil sentences were consistent with the theme of studied sentences, and there were some sentences represented in more than one theme. Nonetheless, with strongly related materials, Myers et al. obtained negative fan effects.

Myers et al. (1984) suggested that for material that is temporally and causally related, elaborate connections are created among propositions, and these connections allow for faster retrieval. Essentially, subjects create short-cut paths for activation to spread among material that can be integrated. However, in the Radvansky and Zacks (1991) studies, sentences were not temporally or causally related, and still there were no fan effects. It is not at all clear why short-cut paths would be produced for their sentences.

As an alternative, Radvansky and Zacks (1991) posited that if the material to be studied represents a single situation of co-occuring events, subjects will build an integrated mental model representation of this information in LTM. Mental models are constructed around a particular situation and are built from individual propositions in memory. They therefore contain all of the information that their underlying propositions would contain (Johnson-Laird, 1983). In terms of activation, however, verification of a fact could rely on a single mental model representation without activation being distributed among individual propositions. As a result, mental models afford faster retrieval with less demand for activation; thus, sentence verification from a single mental model is faster than verification from a proposition network, even when that network is relatively small. The Myers et al. (1984) data are easily explained with this view. Causal and temporal relatedness allow for the development of integrated mental models (Radvansky & Zacks, 1991). The more related the material, the stronger and more coherent is the mental model representation. Moreover, having many sentences to describe a situation will result in a stronger mental model than having only a few sentences; therefore, the Radvansky and Zacks interpretation predicts negative fan effects.

The fan effect has many appealing properties as a measure of total activation. Subjects learn novel sentences so that the difference in a priori knowledge between people is minimized. This task also affords a direct measure of activation limits. People with less activation should show longer verification times, especially as propositional fan size increases. Most relevant to the general capacity model, the fan-effect procedure provides an opportunity to test activation limits independent of the attention-switching feature of the WM spans. Engle et al. (1992) argued that activation differences drive individual differences on WM tasks. In the span tasks, attention switching simply allows to-be-remembered information to drop below threshold. It must then be reaccessed. If the general capacity model is correct, then the typical fan-effect procedure should capture the same important individual differences as the WM spans. Because low-span subjects are those with less overall LTM activation, each concept in a propositional network will receive less activation, and verification time will be slower. Therefore, compared with high-span subjects, low-WM-span subjects should show exaggerated fan effects. Sentence verification time should be longer for these subjects, particularly as propositional fan size increases. We test this prediction in Experiment 1.

In Experiment 2, high- and low-WM-span subjects were presented with fan-effect material that varied in thematic relatedness. From the perspective of the general capacity model, it was predicted that high-span subjects could simultaneously represent all of the necessary information in WM. They should show the negative fan effects indicative of integrated mental models. Because of more severe activation limits, low-span subjects would have difficulty representing all of the information necessary to generate a well-integrated mental model and should therefore produce less negative or possibly positive fan effects.

Experiment 1

Method

Subjects

Eighty subjects were recruited from the departments of psychology, anthropology, and economics at the University of South Carolina. All subjects received course credit for participating. To ensure a wide range of abilities, eight VSAT intervals were defined, and students were chosen so that there were 10 subjects in each of the following categories: 200–300, 310–350, 360–400, 410–450, 460–500, 510–550, 560–600, and 610 and above.

Materials and Procedure

There were two tasks, both presented by means of an IBM PS/2 model 55sx and a color monitor set to black and white. The WM task was an operation-word span similar to that used by Engle et al. (1990). The operations were developed by La Pointe and Engle (1990). Each operation began with the multiplication or division of two integers, the result of which was to be added to or subtracted from a third integer, for example, (6/2) − 2 = 1. An answer to the math problem was also provided, and the answer was correct about half the time. The words for this task were taken from LaPointe and Engle's high-frequency, one-syllable word pool (e.g., *bird* or *door*). For each subject, a pool of 81 mathematical operations was randomly paired with a pool of 81 to-be-remembered words. Therefore, each subject received a different combination of operations and words. The stimuli used to construct this task are presented in Appendix A. During the task, subjects were presented with simple mathematical operations, each of which was paired with a to-be-remembered word, for example, (6/2) − 2 = 1 DOG. Operation–word items were presented one at time on the screen. Subjects read the operation out loud, orally verified whether the answer provided was correct, and then read the word. An experimenter pressed a key as soon as the word was read, and a new operation–word pair was displayed. This pattern continued until a question mark cued subjects to write down all of the words that they could remember, in the order of their presentation. Guessing was encouraged, and recall was not timed. The number of operation–word pairs presented prior to the recall cue varied from 2 to 7. There were three trials at each of these levels, and the order of their presentation was randomized. An additional three trials were presented at the beginning of the task, and each trial contained 2 operation–word pairs. These were used as practice and were not scored. Subjects' span score was the sum of the correctly recalled words for trials that were perfectly recalled (excluding practice). This score was originally reported by Turner and Engle (1989), and consistently correlates with VSAT (Cantor et al., 1991; Engle et al., 1992; Engle et al., 1990; La Pointe & Engle, 1990).

The materials for the fan manipulation were taken from several fan-effect studies (Anderson, 1974, 1983b); Myers et al., 1984; Whitlow, 1984). Sixteen target sentences were generated that all took the form "the SUBJECT is in the PLACE" (e.g., "The LAWYER is in the BOAT"). These sentences are schematically represented in the top half of Figure 1. Each PLACE term was associated with two SUBJECT terms. Each SUBJECT term was associated with one, three, or four PLACE terms. Thus, propositional fan size was 1, 3, or 4. As can be seen in Figure 1, two sets of target sentences were created. The corresponding foil sentences are shown in the bottom half of Figure 1. They were generated by switching SUBJECT terms between sets of target sentences so that they still reflect the same propositional fan size. For example, subjects studied the target sentence "the LAWYER is in the BOAT." The foil for this sentence was "the LAWYER is in the HOUSE."

The procedure for the fan task occurred in two phases. In the first phase, subjects were presented with groups of target sentences for study. Once they had learned these sentences they were given the second phase of the task, which was speeded verification of targets and foils.

Learning phase. The to-be-learned sentences (targets) were presented in groups so that all of the target sentences associated with a single SUBJECT term were presented together. For example, "the plumber is in the park," "the plumber is in the church," and "the plumber is in the bank" were all presented in the same display. The algorithm $n(10\text{ s}) + 10\text{ s}$, where n equaled the number of sentences to be displayed, determined how long each group remained on the screen (Myers et al., 1984; Reder & Anderson, 1980). After each of the groups of sentences had been presented for study, subjects were tested so that we could determine whether they had learned the material. A SUBJECT term was displayed on the screen, and subjects were told to orally recall all of sentences associated with that term. A group of sentences was scored as correct only if all of the appropriate sentences were recalled and if no additional incorrect sentences were generated. After the subjects had been tested on all of the groups of sentences, the sentences were presented again individually for study. The order of presentation was randomized and was not the same as any other study order. The subjects were then tested again for learning. Any group of sentences that had been correctly recalled on three consecutive study–test cycles was then dropped from further study and test. When all of the groups met the criterion, each was presented once more for study, and the learning phase of this task was terminated. At this point, we assumed that all of the material had been learned and that subjects had established LTM propositional networks like those described by Anderson (1983a).

Verification phase. For this phase, target and foil sentences were presented one at a time. Subjects pressed one of two keys to indicate whether the sentence presented had also been studied. The "1" and "3" keys on the numeric keypad were used for *yes* and *no* responses, respectively. Each target and foil was randomly tested three times. RT was measured from the onset of a sentence to the time of the subject's response.

All subjects were first administered the fan-effect task and then the operation–word span. Subjects were tested individually in a single session that lasted approximately 1 hr.

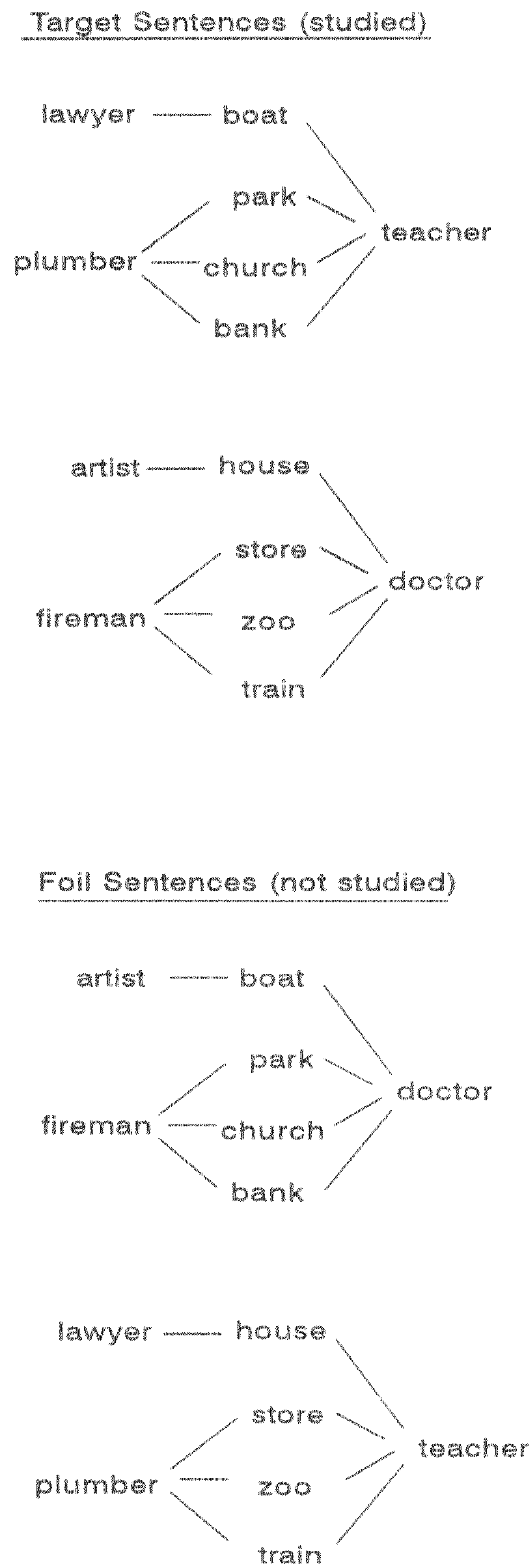

Figure 1. The relationship among SUBJECT and PLACE terms for Experiment 1.

Results

The goal of this study was to test whether high- and low-WM-capacity subjects also differed in their fan effects. Therefore, data for a total of 40 subjects who scored in the upper and lower quartiles of the WM span were selected for initial analyses. These subjects were considered to be high- and low-WM capacity and had mean span scores of 28.25 and 7.61, respectively. It should be noted that all post hoc tests were computed using a Tukey honestly significant difference (HSD) test, and when reported, they reflected a significance level of .05 or better.

Study–Recall Cycles

There are many measures that can be examined from the fan task. The first is the number of study–recall cycles required to meet the learning criterion. This measure provides an indication of whether high- and low-span subjects differed in the amount of time that they needed to learn the target sentences. The mean number of cycles at each fan level was submitted to a Span (high or low) × Fan Size (1, 3, or 4) analysis of variance (ANOVA). The only significant effect reflected an increase in the number of learning cycles as fan size increased, $F(2, 76) = 15.57$, $p < .001$, $MS_e = 0.35$. There was no overall effect of span, $F(1, 38) = 2.78$, $p > .10$, $MS_e = 1.95$, and span did not interact with fan size, $F(2, 76) = 1.02$, $p > .36$. These results show no differences between subjects in the amount of time they needed to learn target sentences.

Reaction Times

The most important measures in this task are the RTs and error rates. The general capacity model predicts differences between high- and low-span subjects in the magnitude of their fan effects. Mean RTs (in milliseconds) for correct responses were submitted to a Span (high or low) × Fan Size (1, 3, or 4) × Trial Type (target or foil) ANOVA. It should be noted that error rates were quite low, averaging less than 5%. As can be seen in Figure 2, low-span subjects showed larger verification times, especially as propositional fan size increased. This was supported by the interaction between span and fan size, $F(2, 76) = 8.93$, $p < .001$, $MS_e = 861{,}702$, indicating significantly longer RTs for low-span subjects at Fan Levels 3 and 4. As predicted, low-span subjects produced much larger fan effects than high-span subjects. This pattern was similar for both targets and foils, and thus no interaction was produced between span, fan size, and trial type, $F(2, 76) < 1.0$.

It should be noted that although high-span subjects produced relatively small fan effects, the slope of their fan effect was significantly different from zero, $t(19) = 2.12$, $p < .05$. Moreover, the magnitude of this effect is quite similar to those demonstrated by Anderson's (1974) Stanford University subjects. He found that the difference between Fan Sizes 1 and 3 was approximately 150–200 ms. The corresponding effect with our high-span subjects was approximately 200 ms.

Overall, high-span subjects were faster than low-span subjects, $F(1, 38) = 6.86$, $p < .02$, $MS_e = 2{,}536{,}240$, and target verification was faster than that associated with foils, $F(1, 38) = 8.96$, $p < .01$, $MS_e = 661{,}792$. Span and trial type did not interact, $F(1, 38) = 2.28$, $p > .14$, and nor did trial type and fan, $F(2, 76) < 1.0$.

Errors

Mean error data were submitted to the same analysis as the RT data. As can be seen in Figure 2, there was only one significant effect. Errors increased as fan size increased, $F(2, 76) = 3.17$, $p < .05$, $MS_e = 51$. The remaining nonsignificant effects are reported in Appendix A.

Does LTM Activation Equal WM Capacity?

In general, there are many ways that high- and low-WM-capacity subjects differ (Engle, in press), including reading and lexical access times. However, these differences do not account for the relationship between WM span and comprehension (Engle et al., 1992). If the general capacity model is correct, then the magnitude of the fan effect and the WM spans reflect the same underlying mechanism. They are both mediated by limits in LTM activation. Therefore, the well-established correlation between WM span and VSAT should be fully captured by differences in the fan effect.

To test this prediction, we computed a regression analysis using the data from all subjects, including mid-span subjects who were not used in the RT and error rate analyses. Fan-effect slopes were computed to reflect increases in RT as a function of fan size. Both target and foil data were used, as the ANOVAs showed that their general pattern did not interact with span. In the original construction of targets, two sets of sentences were generated (see top panel of Figure 1). As a reliability check, a slope relating RT to fan size was computed for each of these target sets and their corresponding foils. Cronbach's alpha indicated that the slopes were consistent ($\alpha = .86$). Therefore, a single slope was recomputed for all of the data. Multiple regression analysis indicated that the slope measure was a strong predictor of VSAT ($R = .48$, $p < .001$) and that this relationship was not significantly improved by entering subjects' WM span (change in $R = .03$, $p > .06$). The corresponding partial correlation provides another way to demonstrate this effect. The zero-order correlation between span and VSAT ($r = .42$, $p < .001$) was substantially reduced ($r = .19$, $p > .06$) when the fan effect slope was controlled. Therefore, individual differences in the fan effect accounted for the ubiquitous correlation between WM span and VSAT. The magnitude of the fan effect appears to reflect the same mechanism as the WM spans.

Experiment 2

We designed Experiment 2 to test whether high- and low-WM-capacity subjects will show the negative fan effect that is indicative of a single integrated mental model. The fan materials were identical to those used by Myers et al., (1984).

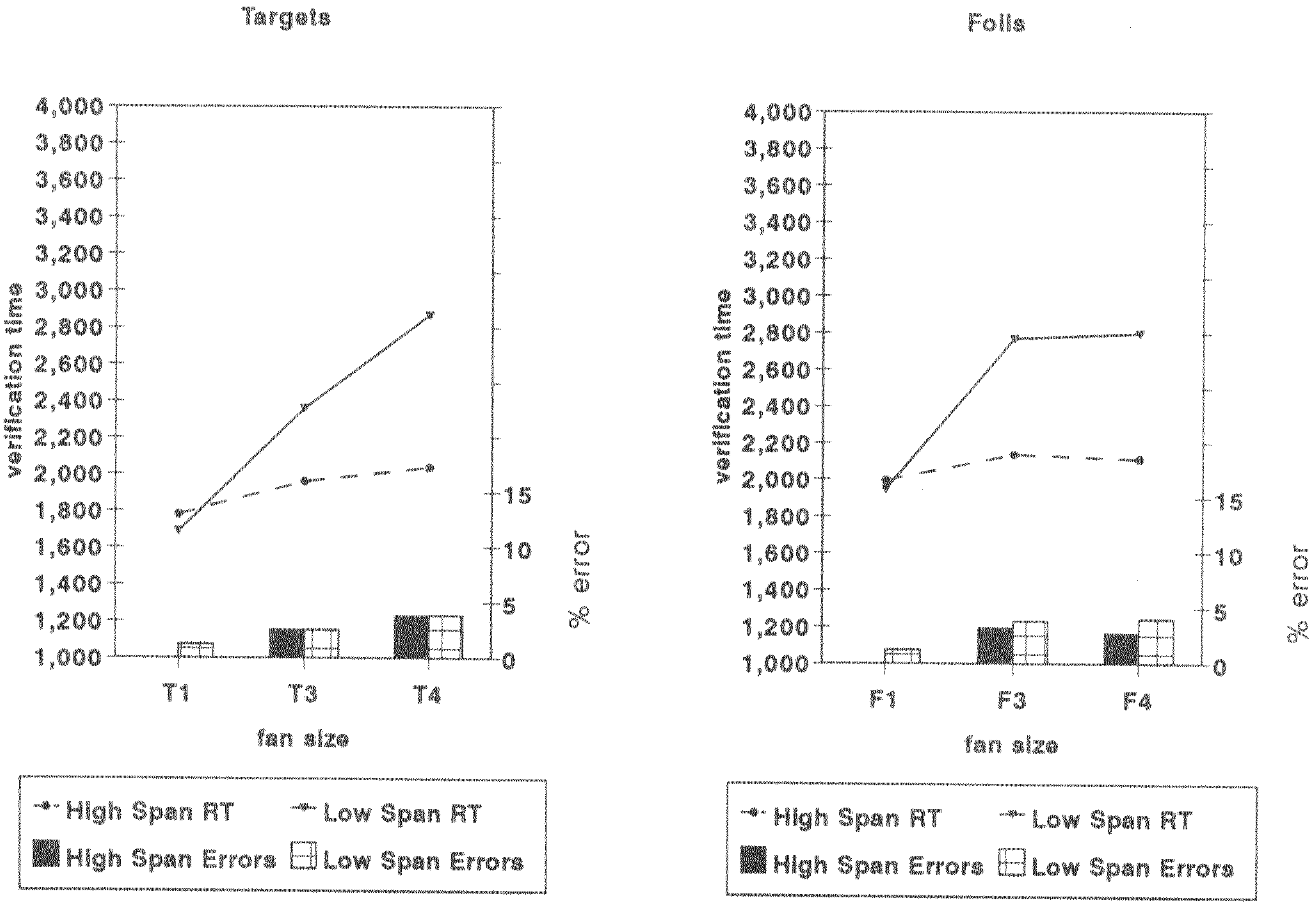

Figure 2. Mean verification times (in milliseconds) and errors for Experiment 1 as a function of trial type (targets or foils) and working-memory span size (high or low). (RT = reaction time; T1 = target sentences with one shared concept; T2 = target sentences with two shared concepts; T3 = target sentences with three shared concepts; F1 = foil sentences with one shared concept; F3 = foil sentences with three shared concepts; F4 = foil sentences with four shared concepts.)

There were groups of three or six sentences, and the groups of six varied in their causal and temporal relatedness. On the basis of the general capacity model, it was predicted that high-span subjects could simultaneously represent the concepts of six sentences in WM and would therefore integrate these sentences into a single mental model representation. These sentences should produce negative fan effects like those demonstrated by Myers et al. By contrast, low-span subjects should have difficulty activating all of the concepts related to a group of six sentences. As a result, these sentences should show either less-negative fan effects, or produce the positive fan effect that is typically found with thematically unrelated materials.

Method

Subjects

Subjects were not the same as those in Experiment 1. As part of ongoing projects in the Working Memory Lab at the University of South Carolina, hundreds of subjects are screened on the operation-word span task. The results of several previous studies have shown that the upper quartile on this task corresponds to a score of approximately 20, and the lower quartile corresponds to a score of 12. From this pool, 80 subjects were chosen to participate in the fan-effect task. There were 40 high-span subjects whose WM span scores averaged 26.20 and 40 low-span subjects with an average score of 7.88. Subjects received course credit for their participation.

Materials and Procedure

The fan materials were the same as those used by Myers et al. (1984) and are provided in Appendix B. There was a total of eight groups of sentences, and each group described the actions of a different SUBJECT (ACTOR, BANKER, LAWYER, TEACHER, SAILOR, PLUMBER, DOCTOR, or FIREMAN). Table 1 represents an example of the sentences associated with TEACHER. For each SUBJECT term, a baseline group of four thematically related sentences were generated (see Table 1). From these baseline sentences, three additional versions of text were constructed, and it was these versions that were actually used in the experiment. In Version 1, three of the four baseline sentences were used to create the Fan Size 3 condition. It should be noted that Fan Size 3 materials were thematically related, but their level of integration was not manipulated in this study. The only manipulation occurred in the Fan Size 6 sentences. Version 2 represents the

Table 1
Example Sentences for Experiment 2

Baseline sentences
The teacher took a table near a window
The teacher read the menu
The teacher checked his wallet
The teacher placed an order
Version 1: Fan Size 3 sentences
The teacher read the menu
The teacher checked his wallet
The teacher placed an order
Version 2: Fan Size 6: high integration
The teacher took a table near a window
The teacher read the menu
The teacher found that the prices were high
The teacher checked his wallet
The teacher decided he could afford a steak
The teacher placed an order
Version 3: Fan Size 6: low integration
The teacher took a table near a window
The teacher read the menu
The teacher asked for a glass of water
The teacher checked his wallet
The teacher munched on a breadstick
The teacher placed an order

high-integration Fan Size 6 condition. For this version, two sentences were added to the original baseline four sentences (see Table 1). These additional sentences were temporally and causally related to the other sentences in the group. For example, in Table 1, the temporally and causally related sentences were "the teacher found that the prices were high" and "the teacher decided that he could afford a steak." Version 3 of the sentences represents the Fan Size 6 low-integration version. In this case, the two sentences that were added to the baseline set of four were not temporally or causally related to the other sentences in the same version. In Table 1, these low-integration sentences are "the teacher asked for a glass of water and "the teacher munched on a breadstick." Thus, for each SUBJECT term, there were three versions of text, a Fan Size 3 version and one version each for high- and low-integration Fan Size 6.

Myers et al. (1984) also attempted to control for consistency judgments during sentence verification. Reder and Anderson (1980) argued that when material is thematically related, sentence verification will be based on whether a sentence is consistent with an overall theme similar to that represented in studied sentences. In Table 1, for example, any sentence that relates to events in a restaurant would be judged as one of the sentences that were studied. There were several ways that Myers et al. attempted to avoid these kinds of judgments. The first reflected the way that foils were created. As noted earlier, there were eight sets of target sentences generated, and each was associated with a different SUBJECT. To control for consistency judgments, Myers et al. developed four general themes (restaurant, ball game, racetrack, and bar). Each of these themes was used in the sentences associated with two SUBJECTS. For example, both the TEACHER and the LAWYER sentences were based on a restaurant theme. Foils were then created by switching the SUBJECT terms for sentences that reflected the same theme. Thus, the foil associated with "the teacher read the menu" was "the doctor read the menu." Reder and Anderson showed that when foil sentences were related to the same theme as targets, the fan effect reappeared, which indicated that consistency judgments could no longer be used. Another way that Myers et al. controlled for consistency judgments was to have sentences that were used for more than one SUBJECT term. For example, in Table 1, ". . . checked his wallet" was associated with the TEACHER. However, this sentence was also associated with the DOCTOR (see Appendix B). For each SUBJECT term, there were two sentences that were also used for other SUBJECTS. Moreover, SUBJECTS that shared the same theme (e.g., LAWYER or TEACHER) could not share the sentences.

In this study, subjects were shown only one version of text associated with each SUBJECT. Versions were counterbalanced across subjects, resulting in four conditions or groups of stimuli. In Condition 1, the groups of sentences related to the ball game and restaurant were presented in their Fan Size 3 version, and groups about the bar and racetrack were presented in their low-integration Fan Size 6 version. For Condition 2, Fan Size 3 items were the same as condition 1, but now the groups related to the bar and racetrack were presented in their Fan Size 6 high-integration versions. In Conditions 3 and 4, the ball game and restaurant groups of sentences were presented in their high- and low-integration Fan Size 6 versions respectively, and the bar and racetrack groups were used for Fan Size 3. There was a total of 80 subjects; 40 were high span and 40 were low span. The counterbalancing conditions were equally represented in each of these span groups.

The procedure for this task was identical to that of Experiment 1. Each group of sentences was displayed individually for study. The sentences remained on the screen for $n(10\text{ s}) + 10$ s, where n equaled the number of sentences to be displayed. After all of the groups were studied, subjects were tested for learning. A SUBJECT term was presented, and subjects were to orally recall the sentences associated with that SUBJECT. After all groups were tested, they were presented for study again. Each group was dropped from these study–test cycles when the group had been perfectly recalled on three consecutive cycles. Once all of the groups had been recalled to criterion, they were presented one last time.

After the learning phase, subjects participated in the verification task. All studied sentences and foils were presented one at a time, and subjects decided whether the sentence displayed had also been studied. Each of these sentences was randomly presented three times, and the order was rerandomized for each subject. The RTs were recorded from the onset of a sentence to the time of the subjects' responses. The entire procedure lasted approximately 1.5 to 2 hr, and subjects were tested individually.

Results

Preliminary analyses showed no effect of counterbalancing order ($F < 1$). Therefore this variable was collapsed for the remaining analyses. As a result, there were four groups of subjects. Two of these groups included high-WM-span subjects, and two groups included low-span subjects. Within these, half of the subjects received high-integration Fan Size 6 sentences, and half received low-integration Fan Size 6 sentences. The Fan Size 3 materials did not differ in their level of integration.

Study–Recall Cycles

The mean number of cycles at each fan level was submitted to a Span (high or low) × Fan Size (3 or 6) × Integration (high or low) ANOVA. This time, high-span subjects required fewer study–recall cycles (M = 4.2) than did low-span subjects (M = 4.9), $F(1, 76) = 12.7$, $p < .01$, MS_e = 1.00. It should be noted, however, that all subjects did eventually learn the sentences to the same criterion. More important, the analysis on errors during verification (reported later) did not show differences between subjects in the amount of material learned. The remaining nonsignificant effects are reported in Appendix B.

Errors

As previously described, there were some sentences that were used for more than one SUBJECT term. These sentences controlled for consistency judgments and were therefore the only sentences identified for RT and error analyses. There were 16 target sentences and 16 related foils. Each was tested three times during the sentence verification phase. The mean number of errors was submitted to a Span (high or low) × Fan Size (3 or 6) × Integration (high or low) × Trial Type (target or foil) ANOVA. As can be seen in Figure 3, all errors were below 5%, and there were no effects that reached statistical significance. Most important was the lack of difference between high- and low-span subjects, $F(1, 76) = 0.70$, $p > .40$, $MS_e = 8.63$.

Although the analysis on cycles to criterion did show initial differences in the time subjects needed to learn the material, the error analysis shows that all of the material was learned. None of the remaining effects were significant, and they are therefore reported in Appendix B.

RT

The most important measures in this task are the verification times. It was predicted that high-span subjects should show negative fan effects, especially in the high-integration condition. Low-span subjects were expected to show fewer negative effects and possible positive fan effects. Mean RT was submitted to the same analysis as the error data. As can be seen in Figure 3, the overall pattern clearly supports the general capacity model. Regardless of the relatedness of sentences, high-span subjects produced negative fan effects. Low-span subjects, the other hand, produced only positive fans.

This was supported by the interaction between span and fan size, $F(1, 76) = 103.40$, $p < .001$, $MS_e = 82{,}090$. Interestingly, increasing integration apparently reduced RTs, thereby producing an overall effect of integration, $F(1, 76) = 4.56$, $p < .04$, $MS_e = 1{,}310{,}115$, but this variable did not interact with span, $F(1, 76) = 0.03$, $p > .85$, or with fan size, $F(1, 76) = 0.34$, $p > .56$. This latter finding is particularly surprising because Fan Size 3 sentences were the same in the high- and low-integration conditions. However, subjects who had highly integrated sets of six sentences also had faster RTs to the Fan Size 3 sentences. Apparently, subjects imposed

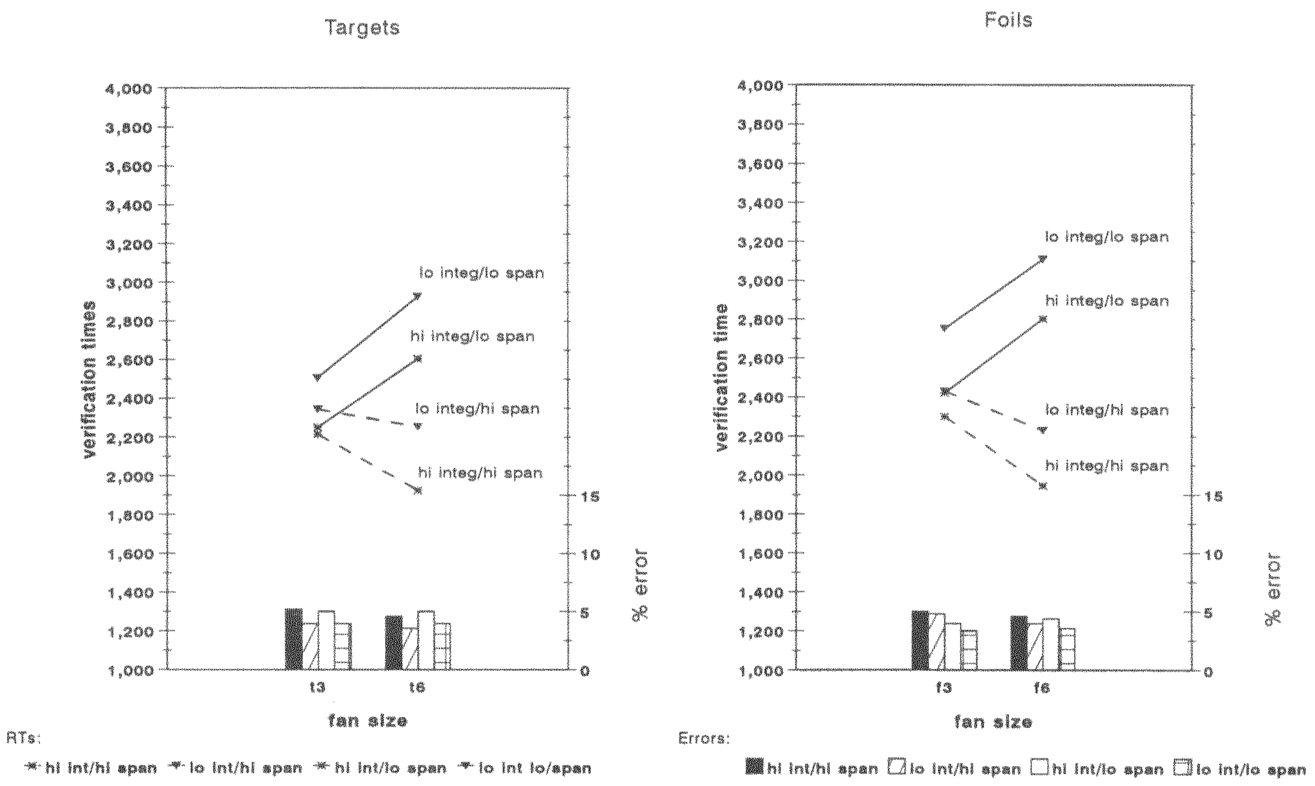

Figure 3. Mean verification times (in milliseconds) and errors for Experiment 2 as a function of trial type (targets or foils), span size (high or low), and integration condition (high or low). (Lo integ/lo span = low integration/low span; hi integ/lo span = high integration/low span; lo integ/hi span = low integration/high span; hi integ/hi span = high integration/high span. t3 = target sentences with three shared concepts; t6 = target sentences with six shared concepts; f3 = foil sentences with three shared concepts; f6 = foil sentences with six shared concepts.

integration on the Fan Size 3 sentences when those sentences occurred in the context of high-integration Fan Size 6 material. Of course, given the nature of the task, subjects always received many more Fan Size 6 sentences than they did Fan Size 3 sentences. Subjects who received high-integration Fan Size 6 sentences apparently imposed some additional level of integration on the Fan Size 3 sentences as well. It should be noted, however, that this overall improvement in verification time did not produce differences in the magnitude of the fan effect. If the result of increasing integration was simply the construction of integrated mental models, then low-span subjects would not have produced positive fan effects in the high-integration condition. Instead, increasing the temporal and causal relatedness of materials seems to have increased the strength of association between items. It is assumed that if these associations are increased, then activation will spread faster and RTs will speed up. No other effects interacted with span, so the remainder of this analysis is reported in Appendix B.

General Discussion

The main goal of these experiments was to test the relationship between WM capacity and LTM-activation limits. The general capacity model assumes that differences in the amount of available LTM activation drive differences in WM capacity and produce the well-established correlation between WM span tasks and verbal abilities. We therefore predicted that the increase in recognition time brought about by increases in fan size, which provides a measure of activation limits, would be highly related to measures of WM. In support of this prediction, low-capacity subjects showed larger effects of fan than did their high-capacity counterparts. Furthermore, individual differences in fan-effect slopes captured the significant covariance between WM span and VSAT. These results provide the first empirical evidence that WM capacity and LTM-activation limits reflect the same underlying construct. They clearly support the general capacity model as well as a variety of other models that include in their architecture the assumption that WM is equivalent to LTM activation (Anderson, 1983a; Cowan, 1988; Just & Carpenter, 1992).

WM and Mental Models

We assumed that the distribution of activation among individual propositions could be attenuated if subjects could retrieve a single higher order representation, like a mental model, that would contain all of the underlying propositional information. There is, however, an important caveat. The initial construction of a mental model requires that the underlying propositions be integrated in WM (Johnson-Laird, 1983). Thus, the subjects' ability to generate this kind of representation should be influenced by their WM capacity. With a large number of propositions, low-capacity subjects might not be able to maintain all of the necessary information in WM to construct a single integrated mental model.

In Experiment 2, we used fan-effect materials that varied in thematic relatedness. Myers et al., (1984) showed that fact retrieval from a set of highly related sentences produced a negative fan effect, and Radvansky and Zacks (1991) suggested that this effect reflects retrieval from a single mental model. Using the same materials and procedure as Myers et al., we tested whether high- and low-WM-span subjects equally showed the negative fan effects indicative of integrated mental models. We found that low-span subjects produced positive fan effects independent of the relatedness of sentences in a group. On the other hand, high-span subjects showed the negative fan effect, which suggests that their retrieval was from a single mental model.

There was one unexpected finding in Experiment 2. There was a context effect associated with the integration manipulation. In this experiment, all subjects received Fan Size 3 materials that did not vary in thematic relatedness. Nonetheless, subjects who studied high-integration Fan Size 6 sentences produced faster verifications to Fan Size 3 materials than did subjects who received low-integration Fan Size 6 sentences. Apparently, subjects who studied high-integration Fan Size 6 sentences also imposed some unitization of Fan Size 3 sentences, thereby decreasing verification times to these items. However, this context effect did not change the magnitude or pattern of fan effects. Only high-span subjects, those with large WM capacities, produced negative fan effects.

What do these results say about the role of WM capacity in forming or using mental models? It should be emphasized that our results do not imply that low-capacity subjects cannot create mental models. It is possible that the low-span subjects in Experiment 2 treated each statement as a separate mental model and did not unitize the statements into a higher level structure. Another possibility is that low-span subjects did group statements together but in smaller groups or structures than those used by high-span subjects. Johnson-Laird, Byrne, and Tabossi (1989) and Byrne and Johnson-Laird (1989) have shown that reasoning takes longer when more mental models are used. If our low-span subjects created more mental models, then we would also expect their verification times to show an increase with fan because activation spread among multiple mental models, not among individual propositions. Thus, low-span subjects could be less able to form mental models because of limited WM capacity; they could be limited in the number of propositions they could integrate into a single coherent unit, or they could be limited in the size and complexity of the mental model they are capable of using. The data we present here do not speak to these issues, but they do strongly suggest that WM capacity may be an important subject variable in the development and use of mental models.

Activation Versus Inhibition

Hasher and Zacks (1988) developed a model of inhibition deficits to explain age-related declines in WM. According to their view, older adults have difficulty preventing unnecessary information from entering WM, and often cannot purge information from WM once it is no longer needed. As a consequence, older people maintain too many concepts in an active state. Essentially, they overload their WM capacity. In addition, because the representation of concepts in WM is prolonged, older subjects will develop spurious connections between items. This will cause interference when these items are retrieved (Gerard, Zacks, Hasher, & Radvansky, 1991; Hamm & Hasher, 1992; Hasher, Stoltzfus, Zacks, & Rypma, 1991; Hasher & Zacks, 1988).

There is some evidence for age-related deficits in inhibition. For example, in selective attention tasks, young adults have difficulty responding to target letters that were used as distractors in previous trials. Apparently, these letters have been inhibited. Older adults show no such effect (Hasher et al., 1991). Older subjects also tend to maintain multiple interpretations of text in memory for a longer period of time than do their younger counterparts (Hamm & Hasher, 1992). Most relevant to the general capacity model are the differences between younger and older subjects in the magnitude of their fan effects. Gerard et al. (1991) showed that older subjects produce much larger fan effects. Gerard et al. argued that because of poor inhibition mechanisms, older subjects prolong the representation of information in WM. As propositional fan size increases, this prolonged activation results in significant interference. Younger subjects will more quickly inhibit unnecessary items and therefore show much smaller fan effects.

The general capacity model and Hasher and Zacks's (1988) framework equally predict age-related differences in the fan effect. However, there is another effect reported by Gerard et al. (1991) that is not easily accommodated by the general capacity view. During the sentence verification phase of their task, older subjects produced substantially more false recognitions than did younger subjects. This is particularly interesting because foil sentences were constructed by repairing concepts that were not originally studied together. According to Gerard et al., because older subjects have difficulty inhibiting information, they will maintain many concepts in WM that were not actually studied together. Consequently, older subjects incorrectly develop spurious connections between these concepts and falsely recognize them as original sentences. The general capacity model does not assume that some subjects have difficulty inhibiting information, and therefore it cannot account for elevated false recognitions.

The inhibition model is fairly new, and it was developed to account for age-related declines in WM. Nonetheless, this view is a potential alternative to models that posit some form of capacity deficit. Although Hasher and Zacks (1988) do not attempt to extend their model to individual differences in younger adults, such a generalization seems reasonable (cf. Brainerd & Reyna, 1989; Dempster, 1991; Gernsbacher & Faust, 1991). There appears to be only one study that has directly tested the relationship between WM tasks and inhibition. MacDonald, Just, and Carpenter (1992) measured reading times for ambiguous and unambiguous sentences in younger adults. Ambiguous sentences are those with more than one possible interpretation, and they should therefore take longer to read if all interpretations are maintained in WM. Interestingly, this ambiguity effect was much larger for high-WM-span subjects than for low-span subjects. If low-span subjects had activated too much related information into WM, then they should have shown the largest ambiguity effects. Although inhibition deficits may drive age-related differences in WM, the results of MacDonald et al. (1992) suggest that they do not mediate individual differences among younger adults.

To the extent that elevated false recognitions indicate inhibition deficits, our own data also do not support an extension of Hasher and Zacks's (1988) model to individual differences among younger adults. We found that low-span subjects did not produce more false recognitions than misses, nor did they show more false recognitions than did high-capacity subjects. However, we do not wish to completely eliminate inhibition as a potential source of individual differences in WM. The validity of this hypothesis should not be judged on the rejection of a single prediction. The fan task, as typically constituted, may involve considerable response competition (Brainerd & Reyna, 1989). A given SUBJECT term like *doctor* is associated with a variety of PLACE terms like *park* and *zoo*, and each PLACE is generally associated with at least one other SUBJECT. Thus, if low-span subjects are more vulnerable to interference than are high-span subjects, the fan task could reflect that difference. It remains to be seen whether the differential fan effect observed between high- and low-span subjects in the present studies would be found when the fact-retrieval task involved minimal interference.

Working memory is necessary for any task that involves the temporary storage of information and has been implicated in such ecologically important tasks as reading, problem solving, reasoning, and understanding spatial relations (Baddeley, 1986; Daneman & Carpenter, 1980; Carpenter, Just, & Schell, 1990; Kyllonen & Christal, 1990; Shute, 1991). In this article we have focused on the general capacity model, which defines WM as information in LTM that has been activated above some resting state to a level that makes it accessible to cognitive processes or procedures. Essentially, WM provides a window that contains information to be attended to or operated on. The size of this window is assumed to reflect the amount of activation that is available to increase the accessibility of LTM knowledge. Finally, there are important individual differences in WM. They are not driven by the processing efficiency associated with a particular task or by more global language-based efficiency parameters (Engle et al., 1992). Rather, individual differences in WM appear to reflect the total amount of capacity or LTM activation available to subjects, independent of the nature of the task. Our data also suggest that capacity limitations are important beyond the initial learning of information. Capacity limitations will influence how well this information can be subsequently retrieved and the extent to which it can be integrated into a single higher order mental model.

References

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

Anderson, J. R. (1983b). Retrieval of information from long-term memory. *Science, 220,* 25–30.

Anderson, J. R. (1983c). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22,* 261–295.

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

Balota, D. A. (1983). Automatic semantic activation and episodic memory encoding. *Journal of Verbal Learning and Verbal Behavior, 22,* 88–104.

Bower, G. H. (1975). Cognitive psychology: An introduction. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 1 pp. 25–80). Hillsdale, NJ: Erlbaum.

Brainerd, C. J., & Reyna, V. F. (1989). Output-interference theory of dual task deficits in memory development. *Journal of Experimental Child Psychology, 47,* 1–18.

Byrne, R. M. J., & Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory and Language, 28,* 564–575.

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

Card, S. K., Moran, T. P., & Newell, A. (1986). An engineering model of human performance. In K. B. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance.* New York: Wiley.

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.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review, 97,* 404–431.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55–81.

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.

Daneman, M. (1987). Reading and working memory. In J. R. Beech & A. M. Colley (Eds.), *Cognitive approaches to reading* (pp. 57–86). New York: Wiley.

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

Daneman, M., & Tardif, T. (1987). Working memory and reading skill re-examined. In M. Coltheart (Ed.), *Attention & performance* (Vol. 12, pp. 491–508). Hillsdale, NJ: Erlbaum.

Dempster, F. N. (1991). Inhibitory processes: A neglected dimension of intelligence. *Intelligence, 15,* 157–174.

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.

Gerard, L., Zacks, R. T., Hasher, L., & Radvansky, G. A. (1991). Age deficits in retrieval: The fan effect. *Journal of Gerontology, 46,* 131–136.

Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 245–262.

Hamm, V. P., & Hasher, L. (1992). Age and the availability of inferences. *Psychology and Aging, 7,* 56–64.

Hasher, L., Stoltzfus, E. R., Zacks, R. T., & Rypma, B. (1991). Age and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 163–169.

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). San Diego, CA: Academic Press.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

Johnson-Laird, P. N., Byrne, R. M. J., & Tabossi, P. (1989). Reasoning by model: The case of multiple quantification. *Psychological Review, 96,* 658–673.

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension.* Needham Heights, MA: Allyn & Bacon.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149.

King, D. R. W., & Anderson, J. R. (1976). Long-term memory search: An intersecting activation process. *Journal of Verbal Learning and Verbal Behavior, 15,* 587–605.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity. *Intelligence, 14,* 389–433.

La Pointe, 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.

Lewis, C. H., & Anderson, J. R. (1976). Interference with real world knowledge. *Cognitive Psychology, 8,* 311–335.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology, 24,* 56–98.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227–234.

Myers, J. L., O'Brien, E. J., Balota, D. A., & Toyofuku, M. L. (1984). Memory search without interference: The role of integration. *Cognitive Psychology, 16,* 217–242.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Pirolli, P. L. & Anderson, J. R. (1985). The role of practice in fact retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 136–153.

Radvansky, G. A., & Zacks, R. T. (1991). Mental models and the fan effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 940–953.

Ratcliff, R., & McKoon, G. (1981). Does activation really spread? *Psychological Review, 88,* 454–462.

Reder, L. M. (1982). Plausibility judgments versus fact retrieval: Efficient strategies for question-answering. *Psychological Review, 89,* 250–280.

Reder, L. M., & Anderson, J. R. (1980). A partial resolution of the paradox of interference: The role of integrating knowledge. *Cognitive Psychology, 12,* 447–472.

Reder, L. M., & Ross, B. H. (1983). Integrated knowledge in different tasks: The role of retrieval strategy on fan effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 55–72.

Salthouse, T. A., Mitchell, D. R. D., Skovronek, E., & Babcock, R. L. (1989). Effects of adult age and working memory on reasoning and spatial abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 507–516.

Shiffrin, R. M. (1975). The locus and role of attention in memory systems. In P. Rabbit & S. Dornic (Eds.), *Attention & Performance* (Vol. 5, pp. 168–193). Hillsdale, NJ: Erlbaum.

Shute, V. J. (1991). Who is likely to acquire programming skills? *Journal of Educational Computing Research, 7,* 1–24.

Smith, E. E., Adams, N., & Schorr, D. (1978). Fact retrieval and the paradox of interference. *Cognitive Psychology, 10,* 438–464.

Thorndyke, P. W., & Bower, G. H. (1974). Storage and retrieval processes in sentence memory. *Cognitive Psychology, 5,* 515–543.

Turner, M. L., & Engle, R. W. (1986). Working memory. *Proceedings of the Human Factors Society, 30,* 1273–1277.

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

Whitlow, J. W., Jr. (1984). Effects of precueing on focused search in fact retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 733–744.

Whitlow, J. W., Jr., Smith, E. E., & Medin, D. L. (1982). Retrieval of correlated predicates. *Journal of Verbal Learning and Verbal Behavior, 21,* 383–402.

Appendix A

Stimuli and Remaining Analyses: Experiment 1

Stimuli for Experiment 1

Word Pool

add	dream	knife	skill
aid	dress	knee	snake
arm	dust	lock	stay
back	east	meal	sum
beach	fact	miss	talk
beat	far	moon	tall
bible	fight	mouth	taste
bird	file	near	tool
blue	form	need	town
bomb	forth	nose	trade
brain	gas	out	tree
break	green	own	truck
buy	guest	pair	type
camp	guy	rain	wait
care	hall	rock	wire
cause	hard	roll	
close	head	scene	
curve	heat	score	
cut	help	sea	
dance	jump	send	
desk	key	set	
door	king	shall	

Operations Pool

(9 × 1) − 9 = 1	(4/1) − 1 = 5	(5/5) + 1 = 2
(8 × 1) + 8 = 16	(7 × 1) − 6 = 2	(4/1) − 4 = 2
(7 × 1) + 6 = 13	(8 × 1) + 5 = 13	(6 × 2) + 2 = 14
(10 × 2) + 3 = 23	(4/2) − 1 = 3	(3 × 2) − 1 = 6
(9 × 7) − 1 = 49	(4/2) − 2 = 2	(2 × 1) + 1 = 3
(6/2) − 3 = 2	(8/2) − 4 = 2	(10/1) − 5 = 7
(7 × 7) − 1 = 49	(6 × 3) − 2 = 17	(5 × 1) + 1 = 6
(8/1) − 5 = 5	(6/3) + 2 = 4	(9 × 3) + 2 = 29
(10/1) − 9 = 3	(10/1) + 9 = 19	(10/2) − 4 = 3
(10 × 6) + 1 = 61	(9/1) + 8 = 18	(9/3) + 3 = 6
(8 × 4) + 2 = 34	(4 × 2) + 1 = 9	(8/1) − 6 = 4
(10 × 5) + 2 = 52	(10/2) + 4 = 9	(7 × 2) − 1 = 14
(10/2) + 6 = 10	(4/2) − 1 = 3	(9/1) − 7 = 4
(9/1) + 1 = 10	(5/1) + 4 = 9	(8/4) + 2 = 4
(3/3) + 1 = 2	(6 × 4) + 1 = 25	(7 × 2) + 3 = 17
(8/4) − 2 = 2	(3 × 1) − 2 = 2	(9/3) + 1 = 4
(2/1) − 2 = 2	(6/2) + 1 = 4	(9/1) + 8 = 18
(6/2) + 1 = 4	(10/1) + 3 = 13	(4 × 2) + 1 = 9
(7 × 7) + 1 = 50	(7/1) + 6 = 12	(10/2) + 4 = 9
(5/1) − 1 = 6	(10/2) + 4 = 9	(4/2) − 1 = 3
(10 × 2) + 3 = 23	(8 × 2) − 4 = 13	(5/1) + 4 = 9
(3/1) − 1 = 4	(7/1) − 2 = 7	(6 × 4) + 1 = 25
(4 × 2) − 2 = 7	(10/2) − 4 = 3	(3/1) + 1 = 4
(3/1) + 3 = 6	(10/1) + 1 = 11	(3 × 2) − 1 = 6
(9 × 2) − 1 = 18	(10/1) + 3 = 13	(3/1) + 1 = 4
(4 × 4) + 1 = 17	(10/1) + 9 = 19	(9/1) + 5 = 14
(3/1) − 2 = 3	(2/2) + 2 = 2	(10/1) − 1 = 11
(6 × 1) − 6 = 1		

Remaining Analysis: Experiment 1

There was no effect of span, $F(1, 38) = 0.35$, $p > .55$, $MS_e = 125$, or trial type, $F(1, 38) = 0.43$, $p > .51$. There was no interaction between span and fan size, $F(2, 76) = 0.06$, $p > .94$, $MS_e = 51.41$, or between trial type and fan, $F(2, 76) = 0.26$, $p > .77$, $MS_e = 15.75$. There was no interaction of span and trial type, $F(1, 38) = 3.57$, $p > .07$, $MS_e = 15.67$, and no interaction between fan size, span, and trial type, $F(2, 76) = 1.35$, $p > .26$, $MS_e = 15.75$.

(*Appendixes continue on next page*)

Appendix B

Stimuli and Remaining Analyses: Experiment 2

In each group, Lines 1, 2, 4, and 6 represent the baseline condition. Lines 2, 4, and 6 were used as Version 1 Fan Size 3. The sentences in Lines 3 and 5, in italics, were used in the Fan Size 6 conditions and varied in temporal and causal relatedness. For Sentences 3 and 5, the left-hand side represents high integration and the right-hand side represents low integration.

The ACTOR
had a ticket for the Mets game
went to the ball game
sat down as the umpire yelled play ball/bought a hot dog from a vendor
saw the start of the ball game
found the first few innings boring/looked at his program
went home early

The BANKER
decided to see a baseball game
arrived at the ball park
found a crowd buying tickets/bought a souvenir pennant
waited in line
entered in time to see his team score/sat near the first base dug out
cheered loudly

The LAWYER
wanted a good meal
went to a new restaurant
found all the tables filled/saw there was a salad bar
waited in line
asked about the specialties/spoke to the hostess
made a selection

The TEACHER
took a table near a window
read the menu
found that the prices were high/asked for a glass of water
checked his wallet
decided he could afford a steak/munched on a breadstick
placed an order

The SAILOR
was in the corner saloon
chatted with the bartender
asked him to turn on the TV/saw his neighbor enter
saw the start of the ballgame
decided he'd better buy something/said hello to a friend
bought a beer

The PLUMBER
entered the barroom
sat down at the bar
thought he's like a martini/listened to a folk singer
placed an order
decided to have only one drink/watched some people play darts
went home early

The DOCTOR
liked to bet on horses
went to the racetrack
studied the odds/stood at the rail
made a selection
debated betting ten dollars/watched the horses at the gate
checked his wallet

The FIREMAN
entered the stands at the racetrack
watched the horses race
saw his choice win/looked for a seat
cheered loudly
felt his throat become dry/tore up his ticket
bought a beer

Remaining Analysis on the Number of Study–Test Cycles for Experiment 2

There was no effect of integration, $F(1, 76) = 2.63, p < .10, MS_e = 1.0$, or fan size, $F(1, 76) = 2.67, p > .10, MS_e = 0.21$. There were no significant interactions for the effects of span and integration, $F(1, 76) = 0.83, p > .36, MS_e = 1.0$; integration and fan size, $F(1, 76) = 2.40, p > .12, MS_e = 0.21$; span and fan size, $F(1, 76) = 0.19, p > .66, MS_e = 0.21$; or integration, span, and fan size, $F(1, 76) = 1.07, p > .30, MS_e = 0.21$.

Remaining Error Analysis for Experiment 2

There was no effect of integration, $F(1, 76) = 1.67, p > .20, MS_e = 8.63$; trial type, $F(1, 76) = 0.84, p > .36$; or fan size, $F(1, 76) = 0.67, p > .41, MS_e = 2.68$. There were no interactions between integration and span, $F(1, 76) = 0.01, p > .97, MS_e = 8.63$; integration and trial type, $F(1, 76) = 1.32, p > .25, MS_e = 2.13$; integration and fan size, $F(1, 76) = 0.38, p > .54, MS_e = 2.68$; span and fan size, $F(1, 76) = 1.68, p > .19, MS_e = 2.68$; span and trial type, $F(1, 76) = 1.32, p > .25, MS_e = 2.13$; trial type and fan size, $F(1, 76) = 0.84, p > .36$; integration, span, and trial type, $F(1, 76) = 0.0, p = 1.0, MS_e = 2.13$; integration span, and fan size, $F(1, 76) = 0.47, p > .49, MS_e = 2.68$; integration, trial type, and fan size, $F(1, 76) = 0.72, p > .40, MS_e = 2.51$; span trial type and fan size, $F(1, 76) = 0.50, p < .48, MS_e = 2.51$; or integration, span, trial type, and fan size, $F(1, 76) = 0.24, p > .62, MS_e = 2.51$.

Remaining RT Analysis for Experiment 2

High-span subjects were faster than low-span subjects, $F(1, 76) = 11.26, p < .002, MS_e = 1{,}310{,}115.40$, and there was no overall effect of fan size, $F(1, 76) = 3.43, p > .07, MS_e = 82{,}089$. There was no interaction between integration and trial type, $F(1, 76) = 0.10, p > .75, MS_e = 182{,}555$; span and trial type, $F(1, 76) = 0.75, p > .38, MS_e = 182{,}555$; integration, span, and trial type, $F(1, 76) = 0.02, p > .88, MS_e = 182{,}555$; integration, span, and fan size, $F(1, 76) = 0.01, p > .96, MS_e = 82{,}090$; trial type and fan size, $F(1, 76) = 2.63, p > .10, MS_e = 46{,}171$; integration, trial type, and fan size, $F(1, 76)$ 3.85, $p >= .06, MS_e = 46{,}171$; or span, trial type, and fan size, $F(1, 76) = 2.41, p > .12, MS_e = 46{,}171$. The four way interaction was also not significant, $F(1, 76) = 1.74, p > .19, MS_e = 46{,}171$.

Received August 13, 1992
Revision received January 21, 1993
Accepted January 22, 1993 ■