The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t716100704

The role of attention during retrieval in working-memory span: A dual-task study

M. Karl Healey [a]; Akira Miyake [b]
[a] University of Toronto, Toronto, Ontario, Canada [b] University of Colorado at Boulder, Boulder, CO, USA

First Published on: 21 August 2008

To cite this Article Healey, M. Karl and Miyake, Akira(2008)'The role of attention during retrieval in working-memory span: A dual-task study',The Quarterly Journal of Experimental Psychology,62:4,733 — 745

To link to this Article: DOI: 10.1080/17470210802229005

URL: http://dx.doi.org/10.1080/17470210802229005

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2009, 62 (4), 733–745

Psychology Press
Taylor & Francis Group

The role of attention during retrieval in working-memory span: A dual-task study

M. Karl Healey
University of Toronto, Toronto, Ontario, Canada

Akira Miyake
University of Colorado at Boulder, Boulder, CO, USA

We tested the hypothesis that retrieving target words in operation span (OSpan) involves attention-demanding processes. Participants completed the standard OSpan task and a modified version in which all equations preceded all target words. Recall took place under either full attention or easy versus hard divided-attention conditions. Recall suffered under divided attention with the recall decrement being greater for the hard secondary task. Moreover, secondary-task performance was disrupted more by the standard OSpan task than by the modified version with the hard secondary task showing the larger decrement. Finally, the time taken to start recalling the first word was considerably longer for the standard version than for the modified version. These results are consistent with the proposal that successful OSpan task performance in part involves the attention-demanding retrieval of targets from long-term memory.

Keywords: Working memory capacity; Long-term memory retrieval; Complex span tasks; Focus of attention; Short-term memory.

Complex working-memory (WM) span tasks, such as reading span (Daneman & Carpenter, 1980) and operation span (OSpan; Turner & Engle, 1989), have been used to inform many theories of WM (Conway et al., 2005). Yet, a thorough understanding of the cognitive processes involved in performing these tasks remains elusive. The current study tests a novel hypothesis about these processes: that retrieving target words is attention demanding.

Complex span tasks require information to be stored for a period in the face of ongoing processing before being retrieved. For example, OSpan (Turner & Engle, 1989) requires participants to solve a series of equations, remember a word presented after each equation, and then recall the words. As such, complex span task performance involves three major components: processing, storage, and retrieval of the targets.

Correspondence should be addressed to M. Karl Healey, Department of Psychology, University of Toronto, 100 St. George Street, Toronto, Ontario M5S 3G3, Canada. E-mail: karl@psych.utoronto.ca

We thank Krista Smart, Caitlin Mahy, and Brian Hoessler for their help in collecting and compiling the data. We also thank Lynn Hasher and Morris Moscovitch for their insightful comments on earlier versions of the manuscript.

http://www.psypress.com/qjep
DOI:10.1080/17470210802229005

Until recently, most accounts of complex span task performance have made the assumption that all target words are maintained actively in WM rather than being stored in and retrieved from long-term memory (LTM). From this perspective it could be argued that the processing and storage components are of the most theoretical interest, and that because targets are maintained in an easily accessible state, retrieval is relatively automatic. Accordingly, while much research has focused on the interplay of processing and storage (see Friedman & Miyake, 2004, for a brief review), retrieval has been relatively neglected.

Questioning active maintenance

However, there is reason to question the assumption that complex span tasks are relatively pure measures of active maintenance with little contribution from LTM. It is well accepted that working (or short-term) memory and LTM are interacting systems, and that LTM can supplement WM capacity in many cognitive tasks (see Ericsson & Kintsch, 1995, for one account). Other theories suggest that even in prototypical working-memory tasks performance is at least partially governed by the same mechanisms that underlie LTM performance (e.g., cue-driven retrieval; Nairne, 2002; Nairne & Neath, 2000, and proactive interference; Lustig, May, & Hasher, 2001). In addition to theoretical arguments, there is empirical evidence that LTM contributes to WM task performance. Consider the example of the Brown-Peterson task, which was originally believed to be a relatively pure measure of short-term memory but was later shown to be highly influenced by proactive interference within LTM (Gardiner, Craik, & Birtwistle, 1972), especially when a distractor task prevents immediate recall of targets (Keppel & Underwood, 1962) as is the case in complex span tasks. More recently, the active maintenance assumption has come under close examination within the complex span task literature. For example, Unsworth and Engle (2007) argue that both active maintenance and retrieval from LTM contribute to span task performance.

A detailed examination of the requirements of complex span task performance in light of prominent attention-based models of WM (Cowan, 1999; Engle, 2002; Oberauer, 2002) provides further reason to believe the active maintenance assumption is probably invalid. These models describe WM as a subset of LTM with attention being required both for maintaining a limited number of LTM traces in a highly accessible state and for processing information. Considerable evidence suggests that the maximum number of items that can be maintained is as few as one (Oberauer, 2002) and at most five (Cowan, 1999). In the case of OSpan, even assuming a capacity of five items, it seems unlikely that as many as five or six targets can be actively maintained while leaving enough capacity to successfully process the equations. This analysis suggests that only a few targets can be maintained on each trial and that the remaining targets must be retrieved by other means. In these models, traces that are not being actively maintained reside in LTM (Cowan, 1999, 2000; Oberauer, 2002). It follows that any targets that are not being actively maintained will need to be retrieved from LTM during recall.

The importance of retrieval

The possibility that retrieval in complex span tasks is partially from LTM (e.g., Cowan et al., 2003; Miyake & Friedman, 2004; Unsworth & Engle, 2006, 2007) suggests that retrieval processes are more important than earlier theories assumed. Retrieval processes in LTM are likely to be more complex than those in WM and may be an important part of what complex span tasks measure. Therefore, beginning to delineate the nature of the retrieval processes involved in complex span tasks is of great importance. The present study takes a first step by testing the hypothesis that retrieval processes in complex span tasks are attention demanding: If target words are actively maintained in a highly accessible state, the attentional demands of retrieval should be minimal (Cowan, 1999; Engle, Tuholski, Laughlin, & Conway, 1999b), whereas if the targets are stored in a less accessible state within LTM, attention-demanding retrieval processes will be required (Kane & Engle, 2000).

The claim that LTM retrieval is attention demanding is well supported by evidence from divided-attention studies. These studies show that when participants are required to recall word lists while performing a secondary task, recall accuracy is typically unimpaired but secondary-task performance suffers relative to baseline, suggesting that recall and secondary-task performance draw on the same limited attentional resources (Craik, Govoni, Naveh-Benjamin, & Anderson, 1996; Craik, Naveh-Benjamin, Ishaik, & Anderson, 2000; Fernandes & Moscovitch, 2000; Guez & Naveh-Benjamin, 2006; Naveh-Benjamin & Guez, 2000).[1]

If LTM retrieval processes do indeed operate in complex span tasks they are likely to be particularly attention demanding because the tasks are free-recall situations in which no external cues are presented, and participants must self-generate cues to reactivate the target traces; generating such cues is likely to require attention. For example, Craik et al. (1996) found that the extent to which concurrent retrieval decreased secondary-task performance depended on whether free recall, cued recall, or recognition testing was used: Attentional demand was highest for free recall and lowest for recognition. The authors suggested that the attentional demand of retrieval is partially determined by the amount of "environmental support" available. Essentially, external support, such as a cue word, reduces the need for self-initiated retrieval processes such as cue generation and memory search. Indeed, cue elaboration and memory search processes have been found to be amongst the most attention-demanding aspects of retrieval in a cued retrieval task (Guez & Naveh-Benjamin, 2006; Naveh-Benjamin & Guez, 2000). Elaboration and search processes should be even more difficult and attention demanding in a span task in which no external cues are provided, and encoding of the targets into LTM takes place under impoverished conditions in which rapid presentation followed by immediate processing demands allows little time for mnemonic strategies such as rehearsal or elaborative encoding, resulting in traces that are weakly encoded. Moreover, because participants must recall only targets from the current trial and ignore the targets from previous trials, considerable proactive interference builds up over the course of successive trials. This interference must be overcome at the time of retrieval, a process that is believed to rely on attention (Engle, Kane, & Tuholski, 1999a; Hasher, Lustig, & Zacks, 2007; Hasher & Zacks, 1988; Kane & Engle, 2000).

To summarize, there is an emerging view that complex span task performance is an interplay between active maintenance in WM and retrieval from LTM (e.g., Cowan et al., 2003; Miyake & Friedman, 2004; Unsworth & Engle, 2006, 2007). Consistent with this view, we argue that participants probably attempt to actively maintain as many targets as possible, but the volume of targets and the processing demands overload WM's limited capacity, and targets are displaced into LTM. At the time of retrieval, any currently active targets are output relatively effortlessly, but the remaining targets must be retrieved through an attention-demanding process. Here we tested two specific hypotheses deriving from this view: first that attention is required during retrieval in span tasks and second that the amount of attention required varies with the putative likelihood that targets are active in WM.

We used a divided-attention paradigm to compare the attentional demands of retrieval

[1] It is typically found that dividing attention during retrieval leads to large decrements in secondary-task performance but has little effect on recall levels. In contrast, dividing attention during encoding severely impairs later recall but has only a small effect on secondary-task performance. This pattern suggests that both encoding and retrieval processes are highly attention demanding but that retrieval processes are obligatory (e.g., Craik et al., 1996). The key point for the present article is that LTM retrieval processes demand attention.

between two versions of the OSpan task: a standard version, in which interleaved processing and storage demands were likely to displace some targets from WM, and a modified, simple span-like version in which all of the processing was completed before the imposition of a memory load, making active maintenance of targets maximally likely. Participants recalled targets in the two span tasks either under full attention or while doing a demanding secondary task. Based on work by Cowan et al. (2003), who found that the durations of children's preparatory pauses before outputting the first target were much longer for complex than for simple span tasks, we recorded retrieval time as an indirect measure of attentional demand.

Because recall of actively maintained information is believed to be largely effortless and error free (Cowan, 2000), we predicted that recall accuracy in the modified simple span-like task would be very high overall and considerably higher than in the standard version. Given that dual-task costs in LTM retrieval typically manifest in secondary-task performance and not recall levels (Craik et al., 1996, 2000; Fernandes & Moscovitch, 2000; Guez & Naveh-Benjamin, 2006; Naveh-Benjamin & Guez, 2000), we predicted that performing the secondary task and retrieving the targets concurrently would lead to decreased secondary-task performance. Similarly, we predicted that retrieval times would be increased under dual-task conditions. Moreover, we predicted that the decrease in secondary-task performance and the increase in retrieval time would be considerably larger for the standard version of the OSpan task.

Method

Participants

The participants were 42 undergraduate students from the University of Toronto, who received course credit or monetary compensation for their participation. They all had normal colour vision and either were native speakers of English or had acquired English in early childhood.

Design

We used a 2 (OSpan: standard and modified word-span-like versions) $\times$ 3 (secondary task: location [easy], color [hard], and no secondary-task control) within-subjects design. Each of the resulting six conditions were administered in a separate block. Specifically, across the first three blocks, one of the span tasks was paired with each of the secondary tasks, and across the final three blocks the other span task was paired with the secondary tasks. The order of administering the two versions of the OSpan task was counterbalanced across participants. For each participant the order of the three secondary-task conditions was kept the same for the first three and the final three blocks, but the order of secondary-task conditions was fully counterbalanced across participants. In addition, participants performed the two secondary tasks (the location and colour tasks) by themselves for a baseline measurement of secondary-task performance.

Materials and Procedure

Operation span tasks. In the standard version of the OSpan task (Figure 1A), the equations and targets were presented in alternating order. In the modified version (Figure 1B), all of the equations were presented first, followed by all of the targets. Thus, the modified version was more analogous to a simple word span task. To ensure that the modified version relied principally on active maintenance, it was necessary to choose a set size that would not exceed the capacity for active maintenance.[2] Therefore, we chose to use a set size of 4 items on all trials to match the considerable empirical evidence that the maximum number of items that can be recalled in an immediate serial recall task using only active maintenance is about 3–4 items (e.g., Cowan, 2000). Thus, for

[2] Note that to the extent that targets are being actively maintained in the simple span-like version, recall performance is expected to be near ceiling. If recall in the simple span-like version were to be considerably below ceiling it would suggest that the active maintenance capacity of WM had been overwhelmed, which would increase the LTM contribution.

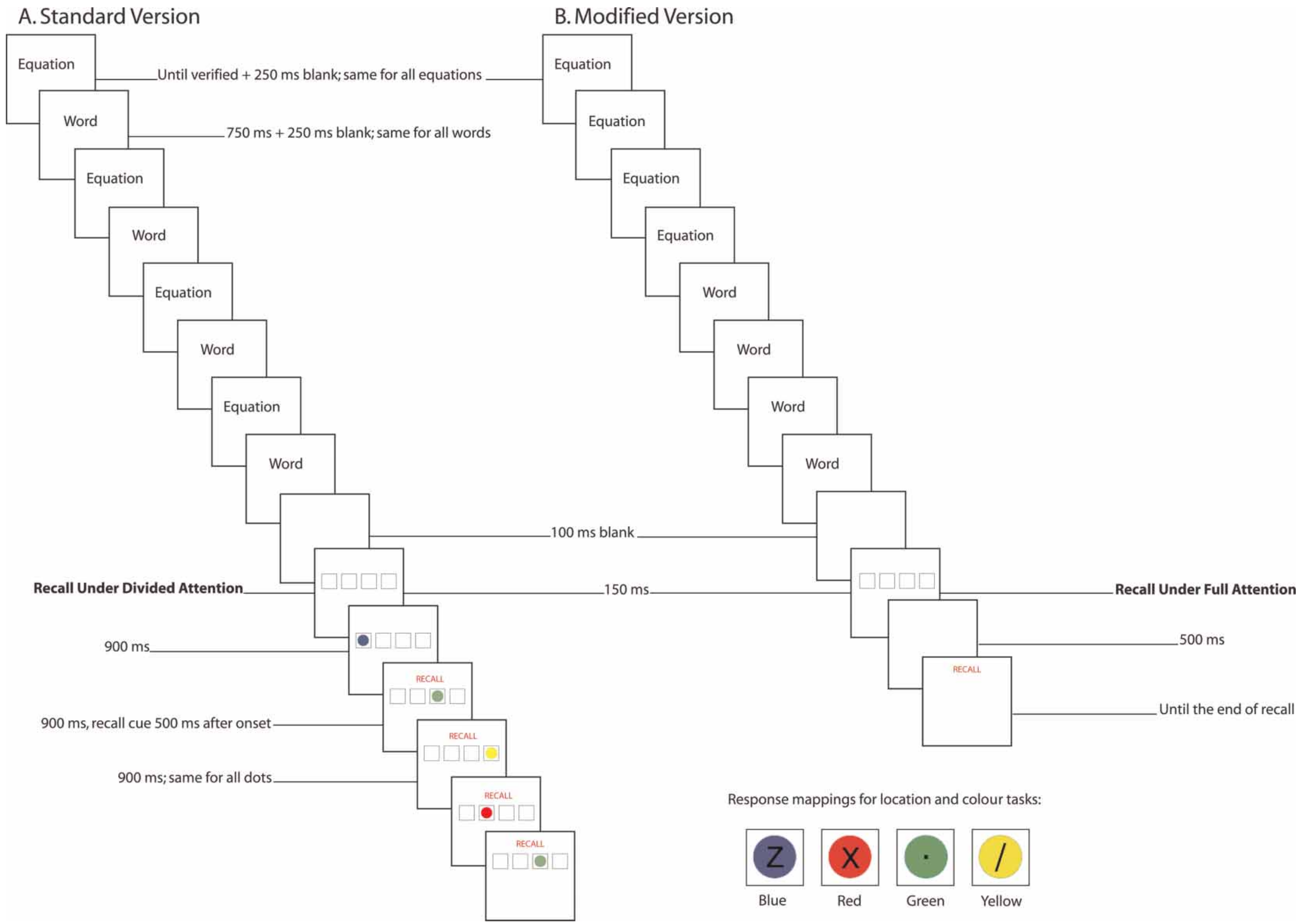

Figure 1. *Diagram illustrating the sequence of events in a trial of the standard span task with divided attention at recall (A), and a trial of the modified span task with full attention at recall (B). The insert at the bottom right shows the key mapping for the location (easy) and colour (hard) secondary tasks. In the print issue shadings and textures are used to represent dot colours. To view a colour version of this figure, please see the online issue of the Journal.*

both the standard and modified versions, all trials consisted of four equations and four target words. There were 9 trials in each of the six dual-task blocks (a total of 54 trials at Set Size 4).

The target words for both versions were common nouns, consisting of four to seven letters and one to two syllables. Each equation—for example, "$(2 \times 2) + 3 = 7$"—was composed of two numbers that were multiplied or divided, a third number that was added to or subtracted from the result, and a final answer. Half of the equations were correct; incorrect equations were within ± 2 of the correct answer. Six lists of words and equations were constructed, and the administration of the lists was counterbalanced across participants such that each list appeared with equal frequency in the six blocks.

The administration of the OSpan task was experimenter paced for both versions to minimize the possibility of idiosyncratic strategy use (Friedman & Miyake, 2004). Specifically, participants were required to read aloud both the equations and targets and, for each equation, to orally indicate whether it was correct or not by saying "yes" or "no" as quickly and as accurately as possible. For both versions, target words were presented for 750 ms. To avoid abrupt transitions between task elements, there was a 250 ms interstimulus interval following each equation and target.

Secondary tasks. There were two versions of the secondary task: the easier location task and the harder colour task. For both versions, four box outlines were first displayed in a horizontal arrangement, as shown in Figure 1. On each trial of both task versions, a coloured dot was displayed in one of the boxes for 900 ms. The location and colour of the dot were determined randomly with the restriction that a dot of the same colour did not appear in the same location on successive trials. To make the task as continuous as possible, no interstimulus interval separated trials. Dots were blue, green, red, or yellow. For the location task participants were required to press a key indicating the location of the dot; for the colour task they indicated the colour of the dot. Figure 1 shows the response mapping for both tasks (e.g., the "z" key indicated far left in the location task and blue in the colour task). To minimize the difficulty of the location task, only location/colour combinations that suggested the same response were used (e.g., every dot presented on the far left was blue). Thus, for the location task both location and colour cued the same response. To maximize the difficulty of the colour task, only incongruent location/colour pairs were used (e.g., a blue dot was never presented on the far left). Thus, for the colour task participants had to ignore the response suggested by the dot's location. The fact that location intuitively maps onto the key layout but colour does not made ignoring location cues especially difficult. For both tasks, participants had to make their response for a trial before the offset of the dot. A harsh beep served as feedback after an incorrect response.

Procedure

Participants first practiced the location secondary task by itself for 33 trials and then provided a baseline (single-task) measure of their performance on the task across 99 trials. Participants were then introduced to the hard colour task. Due to the difficulty of the colour task, participants first completed 60 practice trials that were made easier by slowing the presentation rate to 1,100 ms per dot and using only dots with congruent location/colour parings for the first 16 trials. After this supplemental practice, participants completed 33 practice trials and 99 baseline trials with a 900-ms presentation rate and only incongruent location/colour pairings.

After this secondary-task practice and baseline measurement, participants completed the six dual-task blocks. Each block began with instructions on which span task and secondary task were required for that block. To ensure that participants did not forget the response requirements for the secondary tasks, they performed 33 single-task trials of the relevant secondary task (if any) before beginning a block. Participants were then given 1 practice dual-task trial combining the relevant secondary task and span task, followed by 9 experimental trials for that block.

As illustrated in Figure 1, a blank screen was displayed for 100 ms following the termination of the last target word in the span task, after which the boxes for the secondary task were then displayed for 150 ms as a cue to prepare for recall. After this point the sequence of events was different for the full- and divided-attention conditions. In the divided-attention conditions (Figure 1A), the secondary-task boxes remained on screen, dots started to appear, and the recall signal (a beep and presentation of the word "recall" in red above and below the location of the secondary-task boxes) was presented 500 ms after the onset of the second dot. In the full-attention condition (Figure 1B), the box outlines disappeared, and the recall signal was given 500 ms later. As can be seen in Figure 1, this procedure resulted in a slightly longer retention interval in the divided-attention conditions (1,650 ms) than in the full-attention conditions (750 ms); however, this extra time was necessary to allow participants to initiate the secondary task before beginning recall, as requiring participants to initiate both tasks simultaneously may have disrupted both processes and contaminated the results. Extending the retention interval in the full-attention condition to match the divided-attention condition would also have been undesirable, as participants would probably use this unfilled period to begin retrieval even though the recall cue was not yet presented.

In all conditions, participants were told to begin recalling the words in serial order immediately upon recall cue presentation; recall continued until all words were recalled or the participant indicated that they could not recall any more, up to maximum of approximately 30 s. Instructions stressed the importance of continuing to perform the secondary task until recall was finished.

Data scoring

Equation verification accuracy. A total of 2 participants who failed to accurately verify a minimum of 75% of the equations from experimental trials were eliminated from all analyses on the grounds that they did not conscientiously complete the task. Including these participants did not alter the statistical significance of any of the main effects or interactions reported below. The mean equation verification accuracy of the remaining participants was 92.2% (range: 77.8–99.1%; $SD = 5.0\%$).

Recall accuracy. We computed a recall accuracy score for each condition by first calculating the proportion of words presented in a given trial that were accurately recalled in the correct serial position and then averaging across all trials in the condition. This proportion-based score has been shown to be an effective measure with good psychometric characteristics (Friedman & Miyake, 2005).

Secondary-task performance. Because the instructions emphasized accuracy of responding over speed and because the secondary tasks (both the easy location task and the hard colour task) had a fairly tight response deadline for each dot presented (i.e., 900 ms), error rate was used as the main dependent measure of secondary-task performance. An error was defined as either making no response before the deadline or making an incorrect response. Because the duration of the secondary-task performance was not constant across individual trials, we calculated the error rate for each secondary-task condition as a proportion by dividing the total number of errors made in that block by the total number of dots presented in the same block.[3]

Retrieval time. Waveforms from digital recordings of the recall periods were examined; for each trial the distinctive recall cue tone was identified and was used as a reference point. The lengths of two intervals were measured for each trial: the preparatory pause and the first interword pause. The preparatory pause was defined as the interval between

[3] This analysis pools together all the trials in each block and hence does not weigh the nine individual trials within a block equally. However, calculating the error proportion for each trial and then averaging them across trials yielded the same conclusions.

the onset of the recall cue and the onset of the utterance of the first recalled word. For the preparatory pause to be measured, the participant must have correctly recalled the target word in the first serial position. The first interword pause was defined as the interval between the offset of the utterance of the first recalled word and the onset of the utterance of the second recalled word. For this pause to be measured a word must have been recalled in both the first and second serial positions. For both intervals, the mean interval length for each condition was obtained by averaging across all valid trials within the condition.

If a participant did not have any valid interval to measure in one or more of the six dual-task conditions due to poor recall performance, that participant's data were excluded from all analyses involving the respective intervals. For the preparatory pause analyses, data from 5 participants were excluded based on this criterion. For the interword pause analyses, the data from the same 5 participants and 8 additional participants were excluded. Data from 2 additional participants were also excluded from all retrieval time analyses due to technical errors with the recorder.[4]

Results

Recall accuracy

Figure 2 shows the proportion of words recalled in each condition. As expected, recall was highly accurate in the modified word-span-like version and significantly higher than in the standard version of the OSpan task, $F(1, 39) = 143.78$, $p < .001$, $\eta_p^2 = .787$, consistent with most words being actively maintained in the modified but not the standard version. The effect of secondary task was also significant, $F(2, 78) = 16.03$, $p < .001$, $\eta_p^2 = .291$. Although a visual inspection of Figure 2 suggests a slight trend, the interaction between span version and secondary task did not reach statistical significance, $F(2, 78) = 0.17$, $p = .843$, $\eta_p^2 = .004$. Planned comparisons revealed that with the exception of the hard and easy tasks in the standard condition, all conditions were significantly different from each other at the $\alpha = .05$ level. Given that the effect of dividing attention during retrieval in LTM tasks generally manifests itself in secondary-task performance but not in recall accuracy (Craik et al., 1996, 2000; Fernandes & Moscovitch, 2000; Guez & Naveh-Benjamin, 2006; Naveh-Benjamin & Guez, 2000), the finding that divided attention had any impact on recall performance is noteworthy and suggests that recall in WM span tasks is indeed attention demanding.

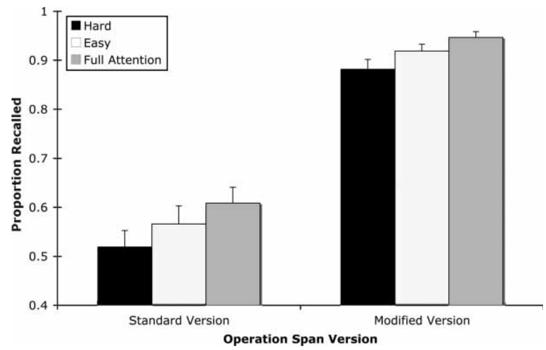

Figure 2. *Mean proportion of words recalled under each secondary task and under full attention for the two span tasks. Error bars represent 1* SEM.

Secondary-task performance

Figure 3 shows the proportion of errors made on the secondary tasks in the baseline control conditions and the divided-attention conditions. The proportion of errors differed both as a function of secondary task (easy vs. hard), $F(1, 39) = 153.65$, $p < .001$, $\eta_p^2 = .798$, and span version (standard, modified, and baseline), $F(2, 78) = 72.52$, $p < .001$, $\eta_p^2 = .650$. Moreover, the effects of secondary task and span version interacted, $F(2, 78) = 32.42$, $p < .001$, $\eta_p^2 = .454$, indicating that, as predicted, performance declined more steeply with secondary-task difficulty for the

[4] Analyses including all trials in which any word was recalled, regardless of accuracy, and all participants (except for the 2 participants for whom we did not have retrieval time data due to recording errors) yielded identical statistical conclusions.

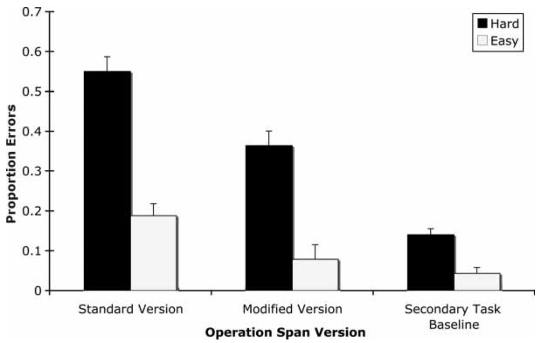

Figure 3. *Mean proportion of errors under each secondary task for the two span tasks and the baseline test. Error bars represent 1* SEM.

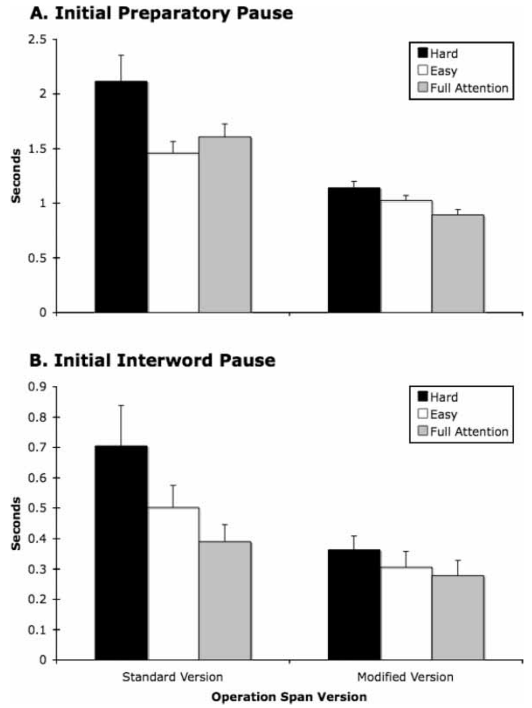

Figure 4. *Mean initial preparatory pause (Panel A) and mean initial interword pause (Panel B) under each secondary task and under full attention for the two span tasks. Error bars represent 1* SEM.

standard version than for the modified word-span-like version. Supporting this interpretation, the two-way interaction remained significant even when the baseline condition was not included in the analysis, $F(1, 39) = 5.59$, $p = .023$, $\eta_p^2 = .125$. The finding that dividing attention had a disproportionately larger negative impact on secondary-task performance for the standard version than for the modified version suggests that retrieval during complex WM span tasks is more attention demanding than retrieval during simple span tasks.

Comparisons were conducted to determine how secondary-task performance differed between the span tasks and baseline. For the hard task, performing the standard span task resulted in more errors than performing the modified span task, $F(1, 39) = 55.66$, $p < .001$, $\eta_p^2 = .588$, which resulted in more errors than the no secondary-task control condition, $F(1, 39) = 43.67$, $p < .001$, $\eta_p^2 = .528$. The pattern was similar for the easy task; participants were more error prone while performing the standard span task than the modified task, $F(1, 39) = 18.42$, $p < .001$, $\eta_p^2 = .321$, and performing the modified span task produced significantly more errors than the control condition, $F(1, 39) = 6.93$, $p = .012$, $\eta_p^2 = .151$.

Retrieval time

Figure 4 presents the preparatory pause data. Participants ($n = 33$) required more time to begin outputting the targets when performing the standard OSpan task than when performing the modified word-span-like version, $F(1, 32) = 40.28$, $p < .001$, $\eta_p^2 = .557$. The main effect of secondary task was also significant, $F(2, 64) = 10.31$, $p < .001$, $\eta_p^2 = .244$. More importantly, the two-way interaction was significant, $F(2, 64) = 3.40$, $p = .040$, $\eta_p^2 = .096$, indicating that the hard secondary task led to more slowing for the standard than the word-span-like span task. This finding is consistent with our contention that the retrieval processes in the standard version are more attention demanding than those in the modified word-span-like version. Comparisons revealed that for the standard operation span task, the preparatory pause was longer under the hard task than under the easy task, $F(1, 34) = 11.91$, $p < .002$, $\eta_p^2 = .259$, but was not significantly different between the easy task and the no secondary-task control condition, $F(1, 34) = 2.40$, $p = .131$, $\eta_p^2 = .066$. For the modified span task, the hard task was marginally different from the easy task, $F(1, 36) = 3.29$, $p = .078$, $\eta_p^2 = .084$, which differed from control, $F(1, 36) = 4.75$, $p = .036$, $\eta_p^2 = .117$.

The overall pattern of results for the first interword pause was similar (see Figure 4B). Specifically, participants ($n = 25$) paused longer between the first and second words that they recalled on the standard version of the operation task than on the modified word-span-like version, $F(1, 24) = 20.69$, $p < .001$, $\eta_p^2 = .463$. The first interword pause was also sensitive to the attentional demand of the secondary task, $F(2, 48) = 7.34$, $p = .002$, $\eta_p^2 = .234$. Unlike the initial preparatory pause, however, the two-way interaction did not approach significance, $F(2, 48) = 1.82$, $p = .173$, $\eta_p^2 = .071$, perhaps because of reduced power due to a smaller number of participants included in the analysis. But it is also possible that, at least on some trials, participants retrieved more than one word during the initial preparatory pause. Specific comparisons revealed that for the standard span task, all the secondary-task conditions were significantly different from each other, but for the modified span task, none of the conditions were significantly different.

Because the overall patterns of results for the initial preparatory pause and for the first interword pause were similar, we conducted a multivariate analysis of variance (MANOVA), including the two retrieval time measures simultaneously as the dependent measure ($n = 25$). In this analysis, not only the main effects of span version, $F(2, 23) = 16.72$, $p < .001$, $\eta_p^2 = .592$, and secondary task, $F(4, 21) = 4.47$, $p = .009$, $\eta_p^2 = .460$, but also the two-way interaction between these factors were significant, $F(4, 21) = 7.51$, $p = .001$, $\eta_p^2 = .589$. Thus, the MANOVA results suggest that performing a highly attention-demanding secondary task (i.e., the hard colour task) disproportionately slowed successful retrieval of target words during the performance of the standard version of the OSpan task relative to the modified word-span-like version.

Discussion

We sought to test the hypotheses that the retrieval processes in span tasks are attention demanding and that the magnitude of this demand can distinguish between complex and simple span tasks. The data from all three dependent measures support these hypotheses. Recall accuracy suffered when attention was divided, indicating that directing attention elsewhere disrupted retrieval processes. This finding is particularly noteworthy given that even with traditional LTM tasks, divided-attention effects most often manifest in secondary-task performance (Craik et al., 1996, 2000; Fernandes & Moscovitch, 2000; Guez & Naveh-Benjamin, 2006; Naveh-Benjamin & Guez, 2000). Secondary-task performance was severely impaired when a span task was conducted concurrently, providing further evidence that retrieval processes were consuming attentional resources. Finally, divided attention led to longer pauses before beginning recall, indicating that retrieval processes became less efficient as attention was diverted. In sum, retrieval was slower and less accurate when attention was divided and still disrupted secondary-task performance. Importantly, and consistent with predictions, the impact of divided attention on secondary-task accuracy and recall time was disproportionately disruptive for the standard OSpan task. These interactions suggest that attentional demand during retrieval is indeed part of what distinguishes complex and simple span tasks.

We argue that the attentional demand of retrieval arises from the need to retrieve targets that have been displaced from WM into LTM (see Unsworth & Engle, 2006, 2007, for similar arguments). That is, although targets that are actively maintained in WM will be available for quick recall with little need for attention, the considerable processing and storage demands of complex span tasks exceed WM's limited capacity for active maintenance necessitating the retrieval of some targets from LTM. Retrieval of targets from LTM is likely to require many processes that are attention demanding, such as the generation and elaboration of retrieval cues, search of memory to determine which traces most closely match the cues, processes to prevent irrelevant traces from interfering with retrieval, source monitoring to ensure that recalled traces are from the current trial, and possibly others. Given this account, we argue that the greater attentional demand of complex relative to simple span tasks is due to the fact that simple span tasks do not require concurrent processing and storage, which allows more targets to be maintained actively in WM, reducing the need for retrieval from LTM.

We have distinguished between items that are highly accessible because they are maintained in WM and items that are inaccessible because they reside in LTM. We do not, however, view accessibility as a simple dichotomy between highly accessible maintained targets and highly inaccessible unmaintained targets. Rather, we assume that accessibility varies in a roughly continuous fashion, with the level of accessibility of a given trace being determined by multiple factors (e.g., the quality of original encoding, the quality of the retrieval cues being used, etc.). This position is consistent with the attention-based models discussed above, which explicitly propose multiple levels of accessibility. For example, Cowan (1999) and Oberauer (2002) both distinguish between traces that were recently attended and retain some residual activation making them relatively accessible and the majority of LTM traces, which are in an inactive, inaccessible state. The critical point for the present study, however, is that maximum accessibility can only be sustained through active maintenance, and once a trace is no longer being actively maintained its level of accessibility drops immediately and continues to drop with increases in the amount of time (Cowan, 1999) or intervening processing (Saito & Miyake, 2004) since it was last in the focus. That is, even if a LTM trace is relatively accessible, it should never be as accessible as a similar trace that is within the focus. Therefore, retrieving a target from LTM into the focus of attention for output (Cowan, 1999) should require more attention than outputting an item that is already maintained in the focus of attention.

Based on our interpretation of the findings, it is not immediately clear why any divided-attention effects should have been observed for the simple span task. That is, if all or most targets were in an active state, very little attention should have been needed to output them. A possible explanation is that even in the case of the simple span version, not all of the target words can be actively maintained. Four targets is near the maximum estimate of the capacity of active WM, thus if even some of that capacity is directed elsewhere (e.g., to the secondary task) targets may be displaced. Indeed, Unsworth and Engle (2006) found that for list lengths greater than three the correlation between simple span tasks and higher order abilities began to increase, reaching the level of complex span/ability correlations by list length six, suggesting that when list length increases beyond three, simple span tasks begin to measure the same abilities as complex span tasks.

The possibility that even the simple span version was not completely dependent on active maintenance may help explain why the effect of divided attention on retrieval accuracy did not interact with type of span task. That is, if retrieval is at least somewhat attention demanding in both span tasks, the lack of an interaction in recall accuracy may simply indicate that this particular measure of attention demand was not sensitive enough to differentiate between the two span tasks. Indeed, given that any effect of divided attention on recall accuracy was unexpected it is not surprising that of our three measures of attention demand, recall accuracy was the least sensitive.

Modern models of WM stress a close relationship between WM and LTM (Cowan, 1999; Oberauer, 2002), which may contribute to the emerging view that complex span task performance is not completely reliant on active maintenance (e.g., Cowan et al., 2003; Miyake & Friedman, 2004; Unsworth & Engle, 2006, 2007). This new view of span tasks leads to the important observation that processing and storage are not the only theoretically interesting components of span task performance: Retrieval processes are an important but poorly understood component. Here we take a step toward understanding retrieval processes in complex span tasks by showing that they are attention demanding and that the magnitude of the attentional demand distinguishes them from simple span tasks. Moreover, by showing that the retrieval processes in a WM task are empirically similar to those in LTM tasks, we hope to further stress the importance of the interplay between WM and LTM and forge a link between the WM and LTM literatures.

Original manuscript received 23 May 2007
Accepted revision received 23 April 2008
First published online 21 August 2008

REFERENCES

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*, 769–786.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (chap. 3, pp. 62–101). New York: Cambridge University Press.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–114.

Cowan, N., Towse, J. N., Hamilton, Z., Saults, J. S., Elliott, E. M., Lacey, J. F., et al. (2003). Children's working-memory processes: A response-timing analysis. *Journal of Experimental Psychology: General*, *132*, 113–132.

Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*, 159–180.

Craik, F. I. M., Naveh-Benjamin, M., Ishaik, G., & Anderson, N. D. (2000). Divided attention during encoding and retrieval: Differential control effects? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1744–1749.

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

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19–23.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999a). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (chap. 4, pp. 102–134). New York: Cambridge University Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999b). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309–331.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.

Fernandes, M. A., & Moscovitch, M. (2000). Divided attention and memory: Evidence of substantial interference effects at retrieval and encoding. *Journal of Experimental Psychology: General*, *129*, 155–176.

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, *51*, 136–168.

Friedman, N. P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods*, *37*, 581–590.

Gardiner, J. M., Craik, F. I. M., & Birtwistle, J. (1972). Retrieval cues and release from proactive inhibition. *Journal of Verbal Learning and Verbal Behavior*, *11*, 778–783.

Guez, J., & Naveh-Benjamin, M. (2006). Divided attention at encoding and retrieval for once- and thrice-presented items: A micro-level analysis of attentional costs. *European Journal of Cognitive Psychology*, *18*, 874–898.

Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in working memory* (chap. 9, pp. 227–249). New York: Oxford University Press.

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.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 336–358.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, *1*, 153–161.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, *130*, 199–207.

Miyake, A., & Friedman, N. P. (2004, November). *Individual differences in working memory: What do working memory span tasks really measure?* Paper presented at the 45th annual meeting of the Psychonomic Society, Minneapolis, MN, USA.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology*, *53*, 53–81.

Nairne, J. S., & Neath, I. (2000). Long-term memory span. *Behavioral and Brain Sciences*, *24*, 134–135.

Naveh-Benjamin, M., & Guez, J. (2000). Effects of divided attention on encoding and retrieval processes: Assessment of attentional costs and a componential analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1461–1482.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 411–421.

Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing–storage relationship in reading span performance. *Journal of Memory and Language*, *50*, 425–443.

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

Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, *54*, 68–80.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*, 104–132.