Available online at www.sciencedirect.com
ScienceDirect
Journal of Memory and Language 58 (2008) 480–494
Journal of Memory and Language
www.elsevier.com/locate/jml

The role of covert retrieval in working memory span tasks: Evidence from delayed recall tests ☆

David P. McCabe *

Department of Psychology, Colorado State University, Fort Collins, CO 80525, USA

Received 31 August 2006; revision received 17 April 2007
Available online 31 May 2007

Abstract

The current study examined delayed recall of items that had been processed during simple and complex span tasks. Three experiments were reported showing that despite more items being recalled initially from a simple span task (i.e., word span) than a complex span task (i.e., operation span), on a delayed recall test more items were recalled that had initially been processed during the complex span task. This delayed recall advantage for items processed during complex span tasks persisted when subjects were encouraged to process the items deeply during the initial span task (Experiment 2), and when initial recall during the span task was precluded (Experiment 3). The *covert retrieval* model explains these data as being the result of subjects maintaining items in working memory during complex span tasks by covertly retrieving them during the processing phases of the tasks. These covert retrieval attempts provide distributed practice retrieving items from long-term memory, creating effective retrieval cues for later delayed recall that are not created during simple span tasks. The covert retrieval model was supported by data showing that words presented in earlier serial positions during operation span, which had the greatest likelihood of being covertly retrieved, were most likely to be recalled on the delayed recall test. These data support the notion that complex working memory span tasks share considerable processing overlap with episodic memory tasks.

Keywords: Working memory; Long-term memory; Episodic memory; Forgetting; Simple and complex span tasks

Introduction

Working memory has become a core construct in theories of cognitive control (see Miyake & Shah, 1999 for a review of several theories). The notion that a distinct working memory system is responsible for concurrent storage and processing during on-line cognition was originally introduced by Baddeley and Hitch (1974), and their multiple component model is arguably the most influential model of working memory to date. According to this multiple component model, the working memory system is composed of modality specific rehearsal buffers that are responsible for maintenance of information, and a central executive component that is responsible for attention demanding processing, such as language comprehension, goal maintenance, and decision making. One approach to understanding the functioning of the working memory system is to examine

☆ I thank Jeff Heitman for assistance in running the subjects in this study. I also thank Lisa Geraci and Alan Castel for helpful comments on a previous revision of the manuscript. The author was supported by National Institute on Aging Grant 5T32AG00030 during implementation of this research project.
* Fax: +1 970 491 1032.
E-mail address: david.mccabe@colostate.edu

0749-596X/$ - see front matter
doi:10.1016/j.jml.2007.04.004

performance on memory span tasks. In fact, these tasks have been central to experimentation in the domain of working memory, and thus, understanding the encoding and retrieval dynamics of these tasks has been a topic of considerable interest of late (Barrouillet, Bernardin, & Camos, 2004; Bayliss, Jarrold, Gunn, & Baddeley, 2003; Conlin, Gathercole, & Adams, 2005; Hitch, Towse, & Hutton, 2001; Saito & Miyake, 2004; Towse, Hitch, & Hutton, 2000; Towse, Hitch, & Hutton, 2002). Moreover, complex span tasks are of interest because they are often strongly related to measures of higher-order cognition, such as general fluid intelligence (Engle, Tuholski, Laughlin, & Conway, 1999; Süß, Oberauer, Wittman, Wilhelm, & Shulze, 2002), reading comprehension (Daneman & Carpenter, 1980; Turley-Ames & Whitfield, 2003), and episodic memory (McCabe & Smith, 2002; McCabe, Smith, & Parks, in press; Park et al., 1996; Rosen & Engle, 1998).

Span tasks can been categorized as simple or complex, based on the structure and demands of the task. Simple span tasks typically involve presenting several to-be-remembered items successively, each for a brief period of time (usually one second), followed by a recall attempt. These span tasks use various stimuli as the to-be-remembered items, typically digits, letters or words, and the average number of items that subjects can recall typically ranges from 5 to 8 items. Complex span tasks are structurally identical to simple span tasks, but typically include an additional processing component prior to the presentation of each to-be-remembered stimulus. For example, on the operation span task subjects are asked to remember words that are presented for brief periods but prior to the presentation of each to-be-remembered word, subjects must complete an arithmetic problem (i.e., an operation). The addition of the processing component, the arithmetic task, reduces performance considerably, with many fewer items recalled from complex span tasks as compared to simple span tasks. One of the primary purposes of the present study was to examine different explanations for why there are fewer items recalled from complex span tasks, as compared to simple span tasks.

Conceptualizing complex span tasks as long-term memory tasks

Although simple and complex span tasks only differ structurally in that a processing requirement is inserted prior to the presentation of each to-be-remembered item, the tasks differ considerably in terms of their cognitive demands, and the extent to which they tap different cognitive processes. For example, in terms of the multiple component model, simple span tasks like word span only require storage in the phonological loop, whereas complex span tasks require this same storage plus recruitment of central executive resources. For simple span tasks subjects can devote their attention solely to maintaining to-be-remembered items, but on complex span tasks subjects must allocate attentional resources flexibly between processing and storage. Another difference is that simple span tasks unfold over a very short time period, whereas complex span tasks take much longer to complete. As such recall of items from simple span tasks is less likely to require episodic retrieval than are complex span tasks (Unsworth & Engle, 2007). This is not an exhaustive list of differences between simple and complex span tasks, and any one of these or other factors can influence the differences in the level of recall found for simple and complex span tasks.

The covert retrieval model

The present paper focuses on one potent factor affecting performance on simple and complex span tasks, namely, the amount of attentional resources devoted to maintaining to-be-remembered items. Although the purported intent of the interleaved processing component of complex span tasks (e.g., the arithmetic problems in operation span) is to preclude processing of to-be-remembered items, subjects nonetheless appear to use some of the time during the processing component of the task to rehearse, retrieve, or "refresh", to-be-remembered items. This role of maintenance operations during the processing portion of complex span tasks has recently been the subject of considerable research interest (Barrouillet et al., 2004; Friedman & Miyake, 2004; Lépine, Barrouillet, & Camos, 2005; Turley-Ames & Whitfield, 2003), and the results of these studies indicate that recall on complex span tasks increases when there is more opportunity for maintenance of to-be-remembered items during the processing portion of the task, e.g., by allowing subjects extra time for the processing portion of the tasks (for similar results, see Engle, Cantor, & Carullo, 1992; Waters & Caplan, 1996).

In the current paper a *covert retrieval* model of simple and complex span task performance is proposed. Drawing on models that conceptualize working memory as an activated portion of long-term memory, like the embedded-process model (Cowan, 1999), when items are attended to during a span task they are retrieved from long-term memory and become activated in the focus of attention. Because up to four items can be maintained in the focus of attention under normal presentation conditions, covert retrieval processes are unlikely to be employed when maintaining items on simple span tasks unless the length of the trial exceeds four items. On trials with four or fewer to-be-remembered items, the items can be maintained in and reported directly from the focus of attention (cf., Unsworth & Engle, 2006a). By contrast, on complex span tasks the processing phase of the task displaces items from the focus of attention. In an effort to maintain the accessibility of to-be-remembered items to working memory, subjects spend part of the time in which they are supposed to be completing the processing phases of the task covertly retrieving the to-be-remembered items. A comparison of the covert retrieval demands of simple and complex span task trials of length four is presented in Fig. 1. The simple span task, word span, shows no covert retrieval attempts because the three items can be maintained in the focus of attention without covert retrieval. By contrast, for the operation span task, after the first processing phase of the task, each subsequent processing phase provides an opportunity for covert retrieval of the to-be-remembered items. This covert rehearsal can take the form of subvocal rehearsal (Baddeley, 1986), a simple mental search (Cowan, 1992), or a "refreshing" process (Barrouillet et al., 2004). Regardless of the specific nature of the mechanism that maintains the accessibility of the to-be-remembered items, the most important point is that some portion of time during the processing phases of complex span tasks is spent covertly retrieving to-be-remembered items.

Delayed recall of items processed during simple and complex span tasks

In the series of experiments reported in this paper delayed recall for items processed during simple and complex span tasks was compared. The interest was in the "downstream" consequences of having processed items during simple and complex span tasks, and how this could inform theoretical accounts of processes involved in these tasks. The general methodology was fairly simple. Subjects completed both the word span and operation span tasks, a simple and complex span task, respectively. On both tasks subjects completed trials of two, three, and four words, with the only difference being that operation span also included an arithmetic task that had to be completed prior to the presentation of each word. Note that typically longer trials are also included on span tasks, but the purpose of only including trials of length four or less was to ensure that subjects were reporting the items from within the focus of attention for the simple span task. After a delay, subjects were asked to recall the words that they had been exposed to during both the word and operation span tasks.

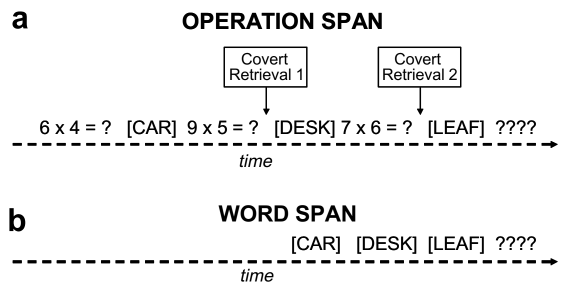

Fig. 1. The covert retrieval model of complex and simple span task performance. At short lists lengths (i.e., less than four items), only complex span tasks afford the opportunity for covert rehearsal.

Because covert retrieval is encouraged by the task demands of the operation span task, but not the word span task, the covert retrieval model makes the counterintuitive prediction that delayed recall should be greater for items that had been processed during the operation span task, as compared to the word span task. This prediction follows from the model because items processed during the operation span task are subject to repeated retrieval from long-term memory, whereas items processed during word span are not. This idea is consistent with Cowan's embedded-process model of working memory (1999), which views items in the focus of attention as the currently available information in working memory, which can be thought of as an activated portion of long-term memory. During the processing phases of the operation span task attention is repeatedly devoted to retrieving to-be-remembered items from long-term memory in to the focus of attention, and as such, they should have a relatively high activation level compared to items processed during word span, which are not repeatedly retrieved. In this model using attentional resources to retrieve items in to the focus of attention creates effective long-term memory retrieval cues. Because items on word span are maintained in the focus of attention during the initial span task, and do not undergo additional retrieval attempts, they are not activated to the same degree in long-term memory, and fewer long-term retrieval cues are created. Under the conditions in the current experiment, i.e., using subspan recall for word span, these differences in retrieval practice should lead to better recall for items processed during complex span tasks. The primary purpose of the present study was to demonstrate this prediction of the model and consider the theoretical implications of this result.

The covert retrieval model not only predicts that the multiple covert retrieval attempts that occur during complex span task performance should increase the availability of individual items for retrieval (cf., Tan & Ward, 2000), but also predicts *which* items will be made more available. As mentioned previously, according to the covert retrieval model, items from early serial positions of a complex span task will be more likely to be retrieved more often during the processing phases of the task. Each covert retrieval attempt, assuming it is successful at maintaining accessibility of the item, creates an association between the encoding context and the to-be-remembered item(s) that can later be accessed at retrieval. Thus, because items from earlier serial positions have the opportunity to be covertly retrieved more often, they will have a greater likelihood of being sampled at retrieval as compared to items from later serial positions. This characteristic of the model makes definitive predictions about how covert retrieval should influence delayed recall—predictions that are not made by other models. Specifically, the number of covert retrieval opportunities should be directly related to the probability of delayed recall, with items from earlier serial positions having a greater likelihood of being recalled after a delay. Using the operation span example presented in Fig. 1, the covert retrieval model predicts that CAR, which is in the first serial position, should be better recalled after a delay than DESK, which is in the second serial position, which should be better recalled than LEAF, which is in the final serial position. For word span the model predicts that there is no need (or very little need) for covert retrieval, so the probability of recalling CAR, DESK, or LEAF should not differ as a function of serial position because the items are maintained in the focus of attention and never undergo any retrieval practice.

Other accounts of complex span task performance have emphasized the role of time-based forgetting in determining recall from complex span tasks, and do not predict a delayed recall advantage for items from complex, as compared to simple, span tasks (Baddeley, 1986; Bayliss et al., 2003; Salthouse, 1996; Towse & Hitch, 1995). One model that has been applied to the issue of differences in recall during simple and complex span tasks is the task-switching model proposed by Towse, Hitch, and colleagues (Hitch et al., 2001; Towse & Hitch, 1995; Towse, Hitch, & Hutton, 1998). According to this model while subjects are engaged in processing activities during complex span task performance, attention is "switched away" from maintenance of the to-be-remembered items, and the to-be-remembered items are forgotten due to the passage of time. The model proposes that processing and storage operations are carried out in a serially alternating fashion (Hitch et al., 2001; Towse et al., 1998, 2002), in both children and adults (Towse et al., 2000), such that time spent processing is time spent away from maintaining items. In discussing the role of the interpolated processing task in the counting span task, a complex span task, Hitch et al. (2001) argued that, "there is no active maintenance…that competes with the execution of [processing] operations. The observation that spans were lower when [processing] took longer was interpreted in terms of rapid loss of activation of [to-be-remembered items] during time switched "away" from retention" (p. 185). Thus, the covert retrieval and task-switching models suggest very different roles of the interpolated processing phases of complex span tasks in influencing recall. The covert retrieval model views the interpolated processing phases as an opportunity for covert retrieval, whereas the task switching model views them as an opportunity for time-based forgetting.

Based on the notion that time leads to forgetting, it would appear that the task switching model predicts less recall for items processed during the operation span task, as compared to the word span task, both immediately and after a delay. However, proponents of the task switching model have acknowledged that other factors in addition to the passage of time (e.g., interference) constrain recall performance as well, and as such, it could also be argued that the task switching model does not make strong predictions in the present context. Nonetheless, one important issue that must be considered when comparing models of span task performance, or models of any cognitive phenomenon for that matter, is whether they make clear predictions about task performance. The covert retrieval model makes the counterintuitive prediction that the passage of time will lead to a crossover interaction between recall from simple and complex span tasks, whereas the task switching model does not.

Experiment 1

The primary purpose of Experiment 1 was to test the prediction of the covert retrieval model that there would be greater recall for items from complex span tasks than simple span tasks on a delayed recall test. The span tasks were administered using typical procedures, and the delayed recall test was given without prior forewarning.

Methods

Subjects

Twenty-four Washington University in St. Louis undergraduates participated for course credit.

Materials and procedure

Each subject completed a word span task and an operation span task. The order of the tests was counterbalanced for each group. For the word span task subjects were asked to view the to-be-remembered words on the screen, and recall them out loud when they saw the recall cue (i.e., ????). One practice trial was completed by the experimenter to demonstrate the procedure, and subjects completed one practice trial of length two before beginning the scored trials. Each to-be-remembered word was presented in 72-point black Arial font on a 17″ computer monitor for one second, with a 250 ms ISI. The recall cue followed the last word on each trial. The test included trials with set sizes of two, three, or four to-be-remembered words, with three trials of each length. The trials were presented in three continuous blocks (no breaks), with one trial of each length in each block, presented in a random order.

The to-be-remembered stimuli were high-frequency concrete nouns, with a frequency of greater than 60 per million (Kucera & Francis, 1967) and concreteness ratings greater than 500 (on a scale of 100-700; Coltheart, 1981). For the word and operation span tasks each of two sets of 27 words served as the to-be-remembered stimuli an equal number of times, and the order of the span tasks was counterbalanced.

The operation span task was identical to the word span task except that before each to-be-remembered was presented, an arithmetic problem (e.g., $8 \times 6 =$) was presented. The problem stayed on the screen until the subject said the answer to the problem out loud, at which time the experimenter advanced the screen to expose the to-be-remembered word on a new screen for one second, followed by the next arithmetic problem (or recall cue if it was the final word in a set). The arithmetic problems all involved multiplication of single digits, and were drawn from a pool of problems that included all of the problems in which one digit was multiplied by another, with the exception of the digits, 1 and 0. Also, the larger of the two numbers in a given problem always came first, and no digit was ever multiplied by itself. The arithmetic problems were sampled without replacement from this set.

After completing the span tasks, subjects completed a distracter task, the Letter Sets task (Ekstrom, French, Harman, & Dermen, 1976), which requires subjects to decide which of four sets of letters does not follow a particular rule in terms of the their sequencing (note that none of these letter sets forms a word). Subjects completed as many of the problems as they could in eight minutes, followed by instructions for the delayed recall test.

For the delayed recall test subjects were told to recall as many words as they could from both of the span tasks that they had completed. They were told to recall the words in any order, and not to guess. Subjects were given five minutes for recall.

Results and discussion

Initial recall during span task performance

Recall of to-be-remembered words during the span tasks will be referred to as *immediate recall*, whereas recall on the delayed recall test will be referred to as *delayed recall*. Although subjects were told to recall the items in their correct serial order, immediate recall was scored without regard to order to make the scoring procedure comparable to that used for delayed recall data (which required free recall). Of primary importance is the interaction of immediate and delayed recall as a function of whether words were presented as part of the word or operation span tasks. The recall data for Experiment 1 are presented in Table 1. Immediate recall was nearly perfect for word span (.99) but substantially lower for operation span (.92), but for delayed recall the opposite was true, with delayed recall being substantially higher for operation span (.18) as compared to word span (.11). Thus, although forgetting was greater for operation span initially, forgetting was greater for word span after a delay. This crossover interaction between immediate and delayed recall is consistent with the predictions of the covert retrieval model, but inconsistent with models assuming time-based forgetting.

Table 1
Recall results for Experiment 1

Task type	Immediate recall	Delayed recall
Word span	.99 (.01)	.11 (.01)
Operation span	.92 (.02)	.18 (.02)

Standard errors are in parentheses.

The aforementioned observations were tested statistically using a 2 (Time of Test: Immediate, Delayed) $\times$ 2 (Task Type: Word Span, Operation Span) repeated measures ANOVA. There was a main effect of Time of Test, $F(1, 23) = 1993.36$, $MSE = 15.57$, $\eta_p^2 = 0.99$, but no main effect of Task Type ($F < 1$). Critically, there was a Time of Test $\times$ Task Type interaction, $F(1, 23) = 20.27$, $MSE = 0.13$, $\eta_p^2 = 0.47$. Planned comparisons confirmed that immediate recall was greater for the word span task, $F(1, 23) = 19.79$, $MSE = 0.06$, $\eta_p^2 = 0.46$, but delayed recall was greater for the operation span task, $F(1, 23) = 11.57$, $MSE = 0.06$, $\eta_p^2 = 0.34$.

Experiment 2

Experiment 1 was important in establishing that there was greater delayed recall for items that had initially been processed during a complex span task, as compared to a simple span task. But there remain questions as to why this counterintuitive result occurred. One possible reason for this finding, as suggested by the covert retrieval model, is that subjects engage in long-term memory retrieval practice during complex span task performance, which later provides effective retrieval cues for a delayed recall test. A tacit assumption here is that the type of rehearsal strategies used in both tasks involves simple rote rehearsal, rather than more elaborative strategies. However, an alternative possibility is that the type of rehearsal strategies subjects engage in for simple and complex span tasks differs qualitatively (Craik & Lockhart, 1972), due to the affordances of the tasks. During a complex span task like operation span, in which information is displaced from the focus of attention and must be retrieved from long-term memory, it may be beneficial for subjects to engage in deeper, more elaborative semantic rehearsal. If subjects were aware that more elaborative encoding strategies would be effective, and had the opportunity to engage those strategies, they might do so. This latter possibility may seem unlikely given that subjects have very little time to rehearse items on either span task (i.e., one second per item), which may discourage elaborative rehearsal, but nonetheless it is an empirical question as to whether subjects are engaging in qualitatively distinct types of rehearsal, which could provide an alternative explanation for the pattern of data from the first experiment.

In order to address the issue of whether the results of Experiment 1 were due to differences in the type of rehearsal strategies engaged for simple and complex span tasks, in Experiment 2 the exact design of the first experiment was replicated, but in addition, subjects were informed that there would be a delayed recall test and that they should pay attention to the meaning of the words so that they would be able to recall them later. Thus, the type of rehearsal strategy employed was equated using an instructional manipulation.

Methods

Subjects

Thirty-two Washington University in St. Louis undergraduates participated for course credit.

Materials and procedure

The materials and procedure were identical to Experiment 1 with the exception that during the instructions for both the word span and operation span tasks subjects were told they would receive a delayed recall test 10 min after they were done with the initial span task. Subjects were told to focus on the meanings of the words in order to be able to effectively recall them later.

Results and discussion

Initial recall during span task performance

The recall data for Experiment 2 are presented in Table 2. As in Experiment 1 initial recall was nearly perfect for word span (.99) but substantially lower for operation span (.88), but for delayed recall the opposite was true, with delayed recall being substantially higher for operation span (.19) as compared to word span (.14). Thus, encouraging semantic processing of to-be-remembered items during the span task did not affect the pattern of results, with greater forgetting initially for the operation span task compared to the word span task, but less forgetting after a delay. In fact, the results changed very little from Experiment 1 to Experiment 2, which provides tentative support for the notion that these tasks afford little opportunity for elaborative encoding.

The aforementioned observations were tested statistically using a 2 (Time of Test: Immediate, Delayed) × 2 (Task Type: Word Span, Operation Span) repeated measures ANOVA. There was a main effect of Time of Test, $F(1,23) = 2126.93$, $MSE = 18.76$, $\eta_p^2 = 0.99$, a main effect of Task Type, $F(1,23) = 5.08$, $MSE = 0.03$, $\eta_p^2 = 0.14$, and critically, there was a Time of Test × Task Type interaction, $F(1,23) = 32.68$, $MSE = 0.20$, $\eta_p^2 = 0.51$. Planned comparisons confirmed that immediate recall was greater for the word span task, $F(1,23) = 35.43$, $MSE = 0.20$, $\eta_p^2 = 0.53$, but delayed recall was greater for the operation span task, $F(1,23) = 5.18$, $MSE = 0.04$, $\eta_p^2 = 0.14$. Thus, the pattern of recall data observed in Experiment 1 was replicated in Experiment 2.

These data suggest that the pattern of recall data found in Experiment 1 was not simply an artifact of different types of rehearsal strategies for simple and complex span tasks. Indeed, comparing across the two experiments, there appear to be no differences in the pattern of data. If the delayed recall advantage for complex span tasks had been the result of much more cursory encoding of items for the word span task, as compared to the operation span task, then we would have expected increases in delayed recall from the word span task in Experiment 2, when subjects were encouraged to engage in elaborative rehearsal. There did appear to be a small increase in delayed recall for the word span task from Experiment 1 to Experiment 2 (i.e., subjects recalled 14% of words in Experiment 2 as compared to 11% in Experiment 1). However, this increase in recall was not statistically significant, $F(1,54) = 2.73$, $MSE = 0.05$, $\eta_p^2 = 0.05$, $p > .10$, nor was there any difference in the overall percentage of words recalled from Experiments 1 and 2, $F < 1$.

Table 2
Recall results for Experiment 2

Task type	Immediate recall	Delayed recall
Word span	.99 (.01)	.14 (.01)
Operation span	.88 (.02)	.19 (.02)

Standard errors are in parentheses.

Experiment 3

Experiments 1 and 2 firmly established a delayed recall advantage for items processed during complex span tasks, relative to simple span tasks. Although the data presented so far are consistent with the covert retrieval model it would be useful to have more direct evidence indicating that retrieval practice during initial span task performance drives this effect. This was the purpose of Experiment 3 and subsequent analyses.

According to the covert retrieval model, during presentation of the to-be-remembered items on operation span subjects are forced to covertly retrieve to-be-remembered items to keep them available for recall. It is this covert retrieval that is believed to drive the delayed recall advantage for operation span. However, it is also possible that differences in overt retrieval of items during the *recall phase* of the initial span tasks could be driving the delayed recall advantage for operation span, rather than differences in the covert retrieval of items during the *processing phases*, as the covert retrieval model predicts. The crux of this alternative explanation for the complex span advantage in delayed recall is that if recall during word span simply requires reporting of items directly from the focus of attention, but recall during operation span requires retrieval from long-term memory, then the recall phase of operation span may be providing long-term memory retrieval practice, but the recall phase of word span is not. This idea would be consistent with the way that retrieval demands have been conceptualized in the present study, but the question of interest is whether this is the reason for the delayed recall advantage for items processed during complex span tasks.

In order to examine whether the initial recall phase of the span tasks caused the delayed recall advantage for operation span, in Experiment 3 the word and operation span tasks were modified such that for half the trials recall was required during the initial span task, as in the first two experiments, but the other half of the trials recall was precluded by a distracter task. Prior to a given trial subjects did not know whether they would have to recall the items immediately after seeing the final word, or if they would have to immediately switch to the distracter task. This aspect of the design required that subjects process all the to-be-remembered words in the same fashion, and thus equating the amount for covert retrieval that might occur, but allowed an examination of delayed recall for words that underwent initial recall practice and those that did not. According to the covert retrieval model, eliminating recall during the initial operation span task should not eliminate the delayed recall advantage for operation span.

Methods

Subjects

Twenty-four Washington University in St. Louis undergraduates participated for course credit.

Materials and procedure

The materials and procedure were identical to Experiment 1 with exceptions noted here. Because there were now two within-subjects conditions (Immediate Recall, No Immediate Recall) for each span task, for counterbalancing purposes there were two trials at each length instead of three, with one trial at each length completed in each condition. Thus, there were nine to-be-remembered words in each condition, with these to-be-remembered words comprising a random subset of the word pool used in the first two experiments. The Immediate Recall trials were identical to the trials administered in the first two experiments. Prior to instructions for the No Immediate Recall trials, subjects were given practice on the matching task, the distracter task that would be performed instead of recall after words were presented for this condition. For the matching task subjects had to decide whether the digits in a two digit number matched in terms of being odd or even (e.g., "yes" for 73 or 64), or did not match (e.g., "no" for 74 or 63). The first two digit number appeared in red 72-point font immediately following the last word on the No Immediate Recall trials, and responses were given verbally and recorded by the experimenter. Subjects were given 1500 ms to respond to each match trial, and there were six matching task trials completed following presentation of the words for each No Immediate Recall trial. Subjects did not know whether a trial was an Immediate Recall or No Immediate Recall trial until they saw either the recall cue or first two digit number of the match task.

For the delayed recall test subjects were told to recall all of the words that had been presented during the initial span tasks, regardless of whether they had been asked to recall them initially. The delayed recall test was given after a two minute delay instead of an eight minute delay used in the first two experiments because pilot testing showed that the no recall group was at floor levels of recall with the longer delay.

Results and discussion

Initial recall during span task performance

There was no immediate recall in the No Immediate Recall condition, but the Immediate Recall condition showed the same pattern as previous experiments. Initial recall was greater for word span (mean = .95, $SD = .07$), but substantially lower for operation span (mean = .77, $SD = .20$), $F(1, 23) = 18.31$, $MSE = 0.39$, $\eta_p^2 = 0.44$.

Delayed recall

On the delayed test, there were more words recalled from the operation span task than the word span task, replicating the results of the first two experiments. Moreover, the delayed recall advantage for the operation span task occurred regardless of whether there had been an initial test. These delayed recall data are presented in Fig. 2. Completing an initial recall trial was helpful to later recall, as evidenced by greater levels of delayed recall for trials on which there had been an initial recall trial compared to those for which there had not. However, initial test condition did not interact with the type of span task. For the Immediate Test condition recall was greater for the operation span task (mean = .28, $SD = .16$) than for the word span task (mean = .17,

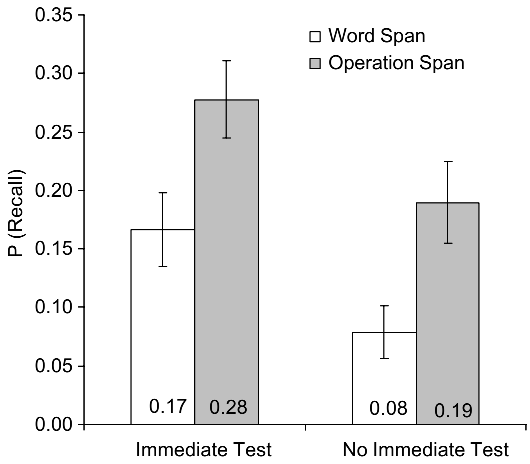

Fig. 2. Experiment 3: Delayed recall results. Note that the categories on the x-axis indicate whether recall occurred on the initial span task (Immediate Recall) or did not (No Immediate Recall).

$SD = .15$), and the same pattern held for the No Immediate Test condition, with greater recall for the operation span task (mean = .19, $SD = .17$) than for the word span task (mean = .08, $SD = .11$). Thus, these data rule out the notion that the overt retrieval occurring during the recall phase of the span tasks drives the delayed recall advantage for operation span. In fact, overt retrieval had little effect at all on the delayed recall advantage for items processed during complex span tasks. Instead the data are consistent with the notion that covert retrieval during presentation of to-be-remembered items during the operation span task drives the effect.

The aforementioned observations for the delayed recall test were examined statistically using a 2 (Immediate Test Condition: Test, No Test) × 2 (Task Type: Word Span, Operation Span) repeated measures ANOVA. There was a main effect of Immediate Test Condition on delayed recall, $F(1,23) = 9.00$, $MSE = 0.19$, $\eta_p^2 = 0.28$, and a main effect of Task Type, $F(1,23) = 12.13$, $MSE = 0.30$, $\eta_p^2 = 0.35$, but there was no Immediate Test Condition × Task Type interaction, ($F = 0$). Planned comparisons confirmed that delayed recall was greater for the operation span task than for the word span task in the Immediate Test condition, $F(1,23) = 4.45$, $MSE = 0.15$, $\eta_p^2 = 0.16$, and in the No Immediate Test condition, $F(1,23) = 5.31$, $MSE = 0.15$, $\eta_p^2 = 0.19$.

Initial serial position of items recalled on the delayed recall Test

I have argued that the present data are most consistent with the covert retrieval model of span task performance. The data presented thus far cast doubt on alternative explanations of the delayed recall advantage for the operation span task, but do not provide strong evidence in support of the covert retrieval model. In order to provide stronger support for the covert retrieval model then, recall of items on the delayed recall task was examined as a function of their serial position on the initial span task. As outlined in the Introduction, the covert retrieval model not only predicts that covert retrieval practice should enhance recall after a delay for operation span as compared to word span, but makes the specific prediction that items from earlier serial positions should have more opportunities for retrieval practice, and should therefore be more likely to be recalled. This is consistent with similar research examining serial position effects suggesting that items from early serial positions undergo more covert retrieval than items from later serial positions (Tan & Ward, 2000). That is, the covert retrieval model predicts that recall will decrease as a function of initial serial position for operation span. By contrast, because the model suggests that there was no covert retrieval practice involved in the word span task because attention was never switched away from maintaining these items, recall of items from the word span task should not have differed as a function of serial position.

Data from all three experiments were combined for the analysis of serial position in order to increase power (the data for the Immediate Test and No Immediate test conditions were combined in to one score for each subject in Experiment 3), and the same general pattern of results occurred for each experiment (i.e., there were no interactions between experiments, conditions, and serial position). Thus, there were 80 subjects contributing to the serial position analysis.

The serial position data, shown in Fig. 3, provide compelling support for the notion that retrieval practice during the operation span task leads to greater recall of items from earlier than latter serial positions on the delayed recall test. Panel one shows the probability that words would be recalled from trials of length two, with recall being greater for words in the first serial position than in the second for the operation span task, but no difference for the word span task. For trails of span length three, shown in the second panel, again recall from operation span decreases linearly from the first to third serial positions, but there was no decrease in recall as a function of serial position for the word span task. Finally, for span length four shown in the third panel, on the operation span task recall was much greater for words from the first two, as compared to the last two, serial positions. There was no decrease in recall across serial positions for word span. Thus, these data support the notion that words from earlier serial positions of the operation span task were rehearsed more effectively, presumably because they were more likely to be covertly retrieved during the processing portion of the task.

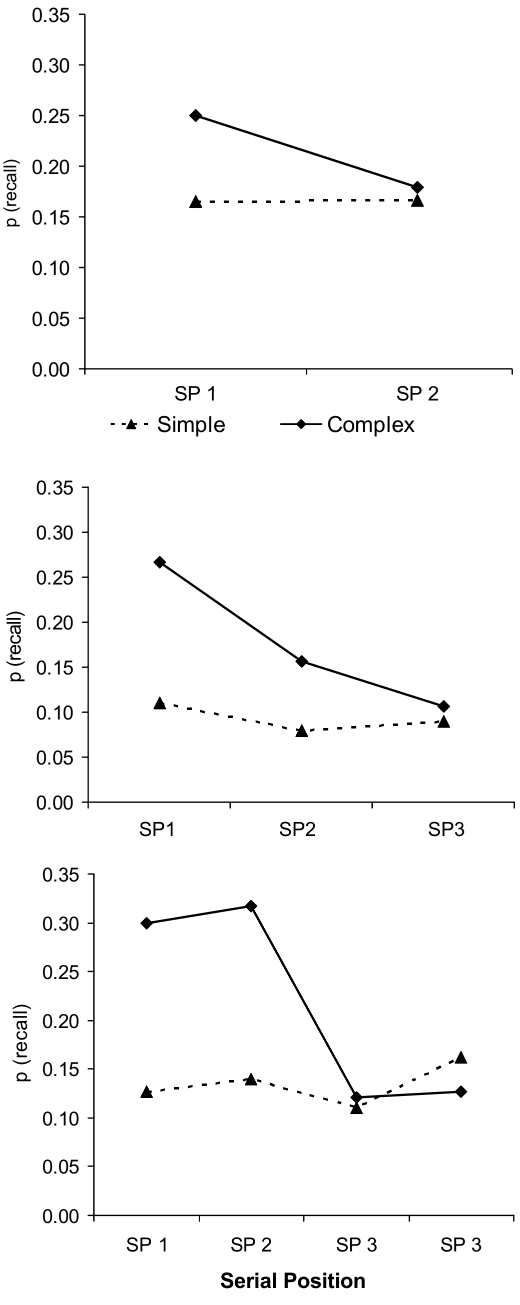

Fig. 3. Initial serial position of items recalled on the delayed recall test. Panel 1, Span task length 1; panel 2, span task length 3; panel 3, span task length 4.

In order to confirm to observations outlined in the previous paragraph, trend analyses were performed for recall at each trial length (2, 3, or 4 to-be-remembered words) separately, beginning with a 2 (Task Type: Word Span, Operation Span) × 2 (Serial Position: 1, 2) repeated measures ANOVA, and followed by ANOVAs with 3 and 4 repeated measures on the latter variable for span task trials of those lengths. For trials of length 2 the main effect of Task Type did not reach the conventional level of significance, $F(1,79) = 3.58$, $MSE = 0.19$, $\eta_p^2 = 0.04$, $p = .062$, nor did the effect of Serial Position, $F(1,79) = 2.56$, $MSE = 0.09$, $\eta_p^2 = 0.03$, but the Task Type × Serial Position interaction was significant, $F(1,79) = 4.01$, $MSE = 0.11$, $\eta_p^2 = 0.05$. This interaction was driven by greater recall of items from the first as compared to the second serial position for operation span, $F(1,79) = 5.90$, $MSE = 0.10$, $\eta_p^2 = 0.07$, but there was no difference in delayed recall as a function of serial position for word span, ($F < 1$). The linear trend analysis at span length 3 revealed a main effect of task type, $F(1,79) = 15.59$, $MSE = 0.83$, $\eta_p^2 = 0.17$, and a main effect of Serial Position, $F(1,79) = 13.29$, $MSE = 0.66$, $\eta_p^2 = 0.14$, as well as a significant interaction, $F(1,79) = 9.30$, $MSE = 0.39$, $\eta_p^2 = 0.11$. There were no quadratic main effects or interactions (all $p > .10$). The linear interaction was driven by a decrease in recall for operation span as a function of serial position, $F(1,79) = 16.78$, $MSE = 1.03$, $\eta_p^2 = 0.18$, but no linear effect on word span, ($F < 1$). Finally, for span length 4, the linear trend analysis revealed a main effect of task type, $F(1,79) = 15.75$, $MSE = 1.06$, $\eta_p^2 = 0.17$, a main effect of Serial Position, $F(1,79) = 12.64$, $MSE = 0.81$, $\eta_p^2 = 0.14$, and a significant interaction, $F(1,79) = 17.30$, $MSE = 1.25$, $\eta_p^2 = 0.18$. Although there were no quadratic main effects or interactions (F's < 1), there was a significant cubic effect of serial position, $F(1,79) = 16.10$, $MSE = 0.58$, $\eta_p^2 = 0.17$, and a significant cubic interaction, $F(1,79) = 5.59$, $MSE = 0.17$, $\eta_p^2 = 0.07$. Decomposing the linear effect revealed a linear decrease in delayed recall as a function of serial position for operation span, $F(1,79) = 34.82$, $MSE = 2.04$, $\eta_p^2 = 0.31$, but no linear effect for word span ($F < 1$). A similar pattern emerged when the cubic effects were examined, with a cubic effect for operation span, $F(1,79) = 18.55$, $MSE = 0.69$, $\eta_p^2 = 0.19$, but no cubic effect for word span, $F(1,79) = 2.06$, $MSE = 0.06$, $\eta_p^2 = 0.03$. Thus, recall dropped off dramatically from serial position 2 to serial position 3, but not from 1 to 2 or 3 to 4, creating a cubic trend in the data, in addition to the linear trend noted above.

General discussion

Three experiments were reported examining delayed recall of items that were processed during a simple span task, i.e., word span, and a complex span task, i.e., operation span. In all three experiments delayed recall was greater for words that had been processed during operation span than during word span. This result occurred despite the typical finding that initial recall of words during the span tasks showing the opposite pattern (superior recall of items from word span). This result did not appear to be due to differences in the quality, or type of processing completed during the span tasks, as semantic rehearsal instructions had little effect on performance, and did not affect the interaction of span task type, i.e., simple or complex, and time of test, i.e., immediate versus delayed. The delayed recall advantage for words processed during the complex span task persisted even when immediate recall during the span task was precluded, ruling out the possibility that the effect was driven by retrieval practice during the initial recall phase of the operation span task. Instead a covert retrieval model was proposed to explain the data. According to the covert retrieval model subjects spent part of the time they were supposed to be completing the arithmetic phases of the operation span task covertly retrieving, or refreshing, to-be-remembered items, which drove the delayed recall advantage for items processed during operation span. This hypothesis was supported by the finding that words from earlier serial positions of operation span trials were more likely to be recalled after a delay than words from latter serial positions, but this effect did not occur for word span.

The covert retrieval model of complex span task performance

The covert retrieval model proposed here suggests that covert retrieval occurred during the processing portion of the operation span task, and that this had important consequences for later recall. Cowan's embedded-process model provides a general framework for understanding how the working memory demands of simple and complex span tasks are related to long-term memory retrieval (Cowan, 1999). Repeatedly retrieving items in to the focus of attention enhances their level of activation in long-term memory, making them more available than items that are retrieved less often (or less recently). Because the processing demands of operation span required this sort of repeated retrieval, but those of word span did not, items from operation span were more available for delayed recall than items from word span. Thus, the embedded process model provides a way of understanding how activation might make items more available, in the context of a theory of how working memory is related to cognition more generally.

Because remembering is determined by the overlap between encoding and retrieval cues (e.g., Tulving, 1983), the specific mechanism that enhances retrieval from items processed during operation span relative to word span must address why the encoding-test match is superior for operation span. As alluded to in the Introduction, repeated retrieval opportunities create more encoding cues during study. Although an increase in the number of retrieval cues would be enough to explain the present delayed recall results, the repeated retrievals were spaced out over time as well, which could also have influenced the results. Because elements of the encoding context, e.g., a subject's mood, the other items on the list, etc., change over time, encoding variability is enhanced for items processed during operation span as compared to word span. Because encoding variability increases long-term retention, this mechanism could be contributing to the delayed recall advantage for items processed during operation span.

Taking the encoding variability argument a step further, one can draw an analogy between the present paradigm and studies examining distributed versus massed practice (see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006 for a review). It is well established that distributing practice over time enhances long-term retention relative to massed practice, and this finding is most often explained as the result of encoding variability (Greene, 1989). Indeed, a common finding in the literature of spaced versus massed practice is that initial recall is greater for massed practice conditions, whereas delayed recall is greater for distributed practice conditions (see Roediger & Karpicke, 2006 for a discussion). Indeed, some models, like the temporal context model (Glenberg, 1987), have explained this crossover interaction as the result of the match between the encoding and retrieval contexts changing over time. The argument is that the initial recall context is more similar to the encoding context in the massed condition, resulting in better recall initially, whereas the delayed recall context is more similar to the encoding context in the distributed practice condition, resulting in better recall after a delay. Thus, to the extent that simple span task performance resembles massed practice conditions, and complex span task performance resembles distributed practice conditions, this analogy is apt. It should be pointed out though that the covert retrieval model proposes that there is no retrieval practice during simple span task performance, at least for blocks of subspan trials like in the present experiments, and thus, this analogy is limited in some respects.

The strongest evidence supporting the covert retrieval model was provided by the analysis of delayed recall as a function of serial position (see Fig. 3). According to the model the number of covert retrieval opportunities is greater for items from earlier serial positions than from latter ones because each processing phase of the span task provides an opportunity to retrieve previously presented items (see Fig. 1). Because items from the earliest serial positions were most likely to be covertly retrieved, and each retrieval led to an association with particular contextual cues, items from earlier serial positions were most likely to be recovered on the delayed recall task. The data presented in Fig. 3 confirmed that delayed recall was greater for words from earlier, as compared to later, serial positions on trials at each length.

However, the influence of covert retrieval opportunities was not perfectly monotonic. Although linear effects were found for all trial lengths, on trials of length four the likelihood of recall was roughly equivalent for serial positions one and two, dropped precipitously from serial position two to three, but was again roughly equivalent for serial positions three and four (see Panel 3 of Fig. 3), leading to a quadratic trend in the data. It is possible that the benefit of these covert retrievals diminishes for latter serial positions, after two or more items are already being maintained in working memory. This could be due to the attentional load imposed by already having to maintain items in memory, which could reduce attentional resources available to devote to rehearsing more than two items (see Raaijmakers & Shiffrin, 1981 for a model incorporating this idea). This would lead to a diminution of the benefit of covert retrieval attempts for items from latter serial positions. Indeed, there is evidence to suggest that maintenance of items in working memory, i.e., a memory load, disrupts the effectiveness of encoding of to-be-remembered items (Bayliss et al., 2003; McCabe, Robertson, & Smith, 2005). The idea that memory load reduces the amount of covert retrieval that occurs is also consistent with models that suggests that processing and maintenance functions of working memory depend on the same pool of limited attentional resources (Barrouillet et al., 2004; Case, Kurland, & Goldberg, 1982; Daneman & Carpenter, 1980). The finding that there was no difference for recall of items from the first two serial positions for trials of length four might be the result of retrieval practice providing diminishing returns, with two retrieval attempts leading to a benefit in delayed recall, but three retrieval attempts being no more effective than two. Again, this may have to do with the dynamics of resource trade-offs during span task performance but these speculations would need to be tested by examining delayed recall following trials of longer lengths than four items. Nonetheless, the covert retrieval model provides a fairly impressive account of the delayed recall data, and makes testable predictions about delayed recall following span task processing.

The covert retrieval model also predicts a positive correlation between the probability of delayed recall and the number of possible covert retrieval attempts, regardless of the length of a trial (assuming all other influences were equivalent for these trials). Thus, for example, items from the first serial position of trials of length three have the opportunity to be retrieved twice, and items from the second serial position of trials of length four have the opportunity to be retrieved twice. In both cases similar relatively high levels of delayed recall are predicted by the model. Conversely, items from the third serial position of trials of length three would not be retrieved, nor would items from the fourth serial position of trials of length four, or items from the second serial position of trials of length two, and these items should therefore show relatively poor delayed recall. Thus, the number of opportunities for retrieval attempts is dependent on the number of arithmetic problems that follow presentation of a given item, assuming that each processing episode provides an opportunity for covert retrieval. Although the number of covert retrieval opportunities is correlated with trial length, the two are somewhat independent of each other. In order to test this prediction of the covert retrieval model the number of covert retrieval opportunities (0, 1, 2, or 3) was calculated for each serial position across all three trial lengths (for all three experiments combined), and the correlation between this estimate and delayed recall was examined. Consistent with model predictions, the number of possible retrieval attempts showed a strong positive correlation of +.83 with delayed recall performance for the operation span task (see Fig. 4). Controlling for trial length in a partial correlation had little effect on the correlation (+.88 when trial length was controlled). These data do not appear to be a case of Type 1 error, as the correlation between number of covert retrieval opportunities and delayed recall was present in all three experiments, with correlations of +.79, +.69, and +.87 in Experiments 1, 2, and 3, respectively, and in each case the correlation increased when trial length was controlled in a partial correlation. These data indicate that delayed recall increased with each covert retrieval opportunity, regardless of the length of the trial, again providing strong support for the covert retrieval model.

It is also worth noting that the covert retrieval opportunities that occur during complex span tasks do not

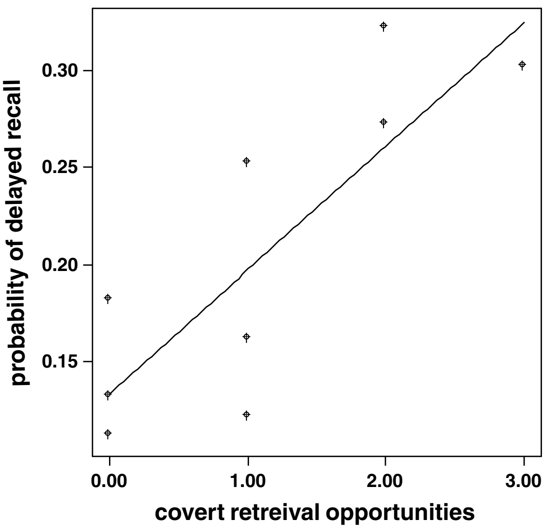

Fig. 4. Delayed recall as a function of the number of covert retrieval opportunities. Note: Data were collapsed across conditions in all three Experiments.

necessarily involve subvocal rehearsal. It is possible that what is retrieved is a marker or pointer for the set of to-be-remembered items, rather than the specific items themselves (Cowan, 1992; Cowan et al., 1998). This possibility was well expressed by Barrouillet et al. (2004), who suggested that in complex span tasks, ''participants may keep short pauses free for the brief reactivation of the memory items by means of a simple mental search without engaging in time-consuming activities such as rehearsal'' (p. 85). Indeed mental search processes that occur too quickly to allow subvocal rehearsal can have dramatic effects on performance (Cowan et al., 1998), suggesting that covert retrieval can occur relatively quickly and without articulation of the to-be-remembered items. And recent findings by Hudjetz and Oberauer (in press) indicate benefits of covert retrieval on recall from complex span tasks under conditions of articulatory suppression, which suggests a role of mental search or refreshing in producing these benefits. It should be noted that the covert retrieval model predicts a benefit of covert retrieval on delayed recall performance regardless of the nature of the covert retrieval attempts. Thus, regardless of whether the items were actually rehearsed subvocally or a pointer to a set of items was retrieved by means of a quick mental search, the same pattern of data would have been predicted.

Another important issue relevant to the present study is whether covert retrieval influences the correlation between complex span tasks and measures of higher-level cognition. Data have recently been reported showing that reducing covert retrieval opportunities during complex span tasks, by minimizing the amount of time subjects have available to covertly retrieve the to-be-remembered items, increases the correlation between those tasks and higher-level cognition (Lépine et al., 2005). This result has been interpreted as indicating that covert retrieval attempts add noise to the correlation between complex span tasks and higher-level cognition. However, another way to interpret these findings is that reducing the opportunity for covert retrieval leads to disproportionately greater benefits to recall for those who are more efficient at covertly retrieving to-be-remembered items, which would be expected to lead to the same result. The issue of the role of covert retrieval in the complex span task/higher-level cognition relationship will need to be addressed in future research.

Implications for time-based forgetting models

The present data are inconsistent with the idea that time-based forgetting constrains recall. For example, according to the task switching model (Towse & Hitch, 1995; Towse et al., 1998, 2000), on a complex span task like operation span, the level of activation for an item begins to decay as soon as attention is removed from maintaining the item in working memory, and switched to the processing component of the task. Because the model proposes that the passage of time is associated with forgetting, and posits that the role of the processing phases of complex span tasks is to cause time-based forgetting, the model cannot readily account for the delayed recall data presented here. However, as noted by proponents of the task switching model (Hitch et al., 2001; Towse et al., 2002), there are other mechanisms that influence recall performance on span tasks, which could include covert retrieval. Thus, the task switching model could be augmented to include a covert retrieval mechanism, and consequently, explain the present data. However, this sort of change reduces the explanatory power of the model, and I would argue, fundamentally changes the nature of the model. In fact, adding a covert retrieval component to the task switching model would make it quite similar to the time-based resource sharing model of Barrouillet and colleagues (Barrouillet et al., 2004; Lépine et al., 2005). This model incorporates time-based forgetting (i.e., decay) and covert retrieval, but as the authors of the time-based resource sharing model point out, the model departs from the task switching model on several core issues, including the notion of resource sharing between processing and storage. The covert retrieval model proposed here is similar to the time-based resource sharing model in many respects, with one of the primary difference being that the covert retrieval model views the notion of time-based decay as superfluous, given that cue-based retrieval mechanisms are necessary and sufficient to explain the data. Indeed, the idea that forgetting is due to decay is also incorporated in the embedded process model (Cowan, 1999), and is one of the primary differences between the covert retrieval model and that model.

Implications for unitary and dual component memory models

The covert retrieval model outlined in this paper has borrowed concepts from both unitary and dual component theories of memory. For example, as noted earlier, the notion that the delayed recall results reported here could be explained as the result of distributed practice is consistent with Glenberg's temporal context model, a unitary memory model. According to this model, rather than arguing that recall from simple and complex span tasks relies on different memory structures, one could argue that the time course involved in retrieving immediately and after a delay interacts with the processing demands of simple and complex span tasks. The central concept in the temporal context model is that all items are processed in a temporal context that changes over time, and the similarity of the encoding and retrieval contexts determines the success of recall. Because items recalled on the immediate span task for word span were retrieved from a temporal context that was very close in time, as compared to items retrieved from operation span, recall from word span was superior. By contrast, after a delay, the temporal distinctiveness created by the interpolated processing tasks on operation span benefited recall, as compared to temporal similarity created during processing of words in a massed fashion on the word span task. This account is consistent with the covert retrieval model if one assumes that one of the consequences of covert retrieval is to enhance the temporal distinctiveness of items, which seems to be an obvious consequence of distributed retrievals. Because covert retrieval only occurs in the operation span task, this account is consistent with the covert retrieval model. Indeed, none of the data presented here are inconsistent with this unitary account, and the notion of temporal distinctiveness provides a fitting explanation of how context is retrieved. Other models that have a unitary character, such as the embedded process model (Cowan, 1999) can account for the data in a similar manner (see also, Nairne, 2002).

Despite being consistent, in principle, with a unitary memory model, it is also possible that functionally distinct processing structures, i.e., the focus of attention and long-term memory, are involved in retrieving items during span tasks. Because up to four items can simply be reported directly from the focus of attention, there is no need for covert retrieval from simple span trials of just a few items. As longer trials are encountered, or when interpolated processing tasks are introduced, as in complex span tasks, covert retrieval is required to maintain the accessibility of to-be-remembered items. This dual component conceptualization of the covert retrieval model is similar to the cue-based retrieval model of Unsworth and Engle (2006a, 2006b), and suggests that separate systems are responsible for retrieval immediately and after a delay.

Although there are no data reported in the current paper that demands a preference for a unitary or dual component explanation, there are other data that suggest a dual component approach provides the most comprehensive explanation of all of the extant memory data. For example, factor analytic studies reveal separable, often weakly correlated, factors for immediate and delayed recall (Geiselman, Woodward, & Beatty, 1982; Unsworth & Engle, 2006a, 2006b). Furthermore, evidence indicates that brain damage can affect performance on short-term memory tasks, but leave long-term memory tasks unaffected, and vice versa (Baddeley, 2000). These data support distinguishing between working memory and long-term memory systems, though this issue will certainly continue to be debated.

Conclusions

The idea that retrieval from complex span tasks primarily involves cue-based retrieval from long-term memory brings up some important questions about the relationship between complex span tasks and other measures of cognition. The idea presented here, that complex span tasks are similar to long-term memory tasks (cf., Unsworth & Engle, 2006a, 2006b, 2007), begs the question of whether the controlled processing estimated using complex span task performance is identical to the controlled processing estimated in other tests of long-term memory, such as recollection estimated using the process dissociation procedure (Jacoby, 1991), recollection involved in source memory (Johnson, Hashtroudi, & Lindsay, 1993), or recollection contributing to remember responses using the remember-know procedure (Tulving, 1985). The present data certainly stress the usefulness of conceptualizing working memory capacity tasks (i.e., complex span tasks) as tests of long-term memory. More research will certainly be needed to understand the memory dynamics of complex span tasks, and to understand how similar, or how different, the processes involved in these tasks are from those involved in other measures of episodic memory.

References

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

Baddeley, A. D. (2000). Short-term and working memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 77–92). New York: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.). *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133*, 83–100.

Bayliss, D., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General, 132*, 71–92.

Case, R., Kurland, M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386–404.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*, 354–380.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A*, 497–505.

Conlin, J. A., Gathercole, S. E., & Adams, J. W. (2005). Children's working memory: Investigating performance limitations in complex span tasks. *Journal of Experimental Child Psychology, 90*, 303–317.

Cowan, N. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language, 31*, 668–684.

Cowan, N. (1999). An embedded process model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory* (pp. 62–101). New York: Cambridge University Press.

Cowan, N., Wood, N. L., Wood, P. K., Keller, T. A., Nugent, L. D., & Keller, C. V. (1998). Two separate verbal processing rates contributing to short-term memory span. *Journal of Experimental Psychology: General, 127*, 141–160.

Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684.

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

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

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

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language, 51*, 136–158.

Geiselman, R. E., Woodward, J. A., & Beatty, J. (1982). Individual differences in verbal memory performance: A test of alternative information-processing models. *Journal of Experimental Psychology, 111*, 109–134.

Glenberg, A. M. (1987). Temporal context and recency. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and learning: The Ebbinghaus centennial conference*. Hillsdale, NJ: Erlbaum.

Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 371–377.

Hitch, G. J., Towse, J. N., & Hutton, U. (2001). What limits children's working memory span? Theoretical accounts and application to scholastic development. *Journal of Experimental Psychology: General, 130*, 184–198.

Hudjetz, A., & Oberauer, K. (in press). The effects of processing time and processing rate on forgetting in working memory: Testing four models of the complex span paradigm. Memory & Cognition.

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

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3–28.

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

Lépine, R., Barrouillet, P., & Camos, V. (2005). What makes working memory spans so predictive of high level cognition? *Psychonomic Bulletin & Review, 12*, 165–170.

McCabe, D. P., Robertson, C. L., & Smith, A. D. (2005). Age differences in Stroop interference in working memory. *Journal of Clinical and Experimental Neuropsychology, 27*, 633–644.

McCabe, D. P., & Smith, A. D. (2002). The effect of warnings on false memories in younger and older adults. *Memory & Cognition, 30*, 1065–1077.

McCabe, D. P., Smith, A. D, & Parks, C. P. (in press). Inadvertent plagiarism in young and older adults: The role of working memory capacity in reducing memory errors. Memory & Cognition.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.

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

Park, D. C., Smith, A. D., Lautenschlager, G., Earles, J., Frieske, D., Zwahr, M., et al. (1996). Mediators of long-term memory performance across the life span. *Psychology and Aging, 11*, 621–637.

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

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181–210.

Rosen, V. M., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language, 39*, 418–436.

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.

Salthouse, T. A. (1996). The processing speed theory of adult age differences in cognition. *Psychological Review, 103*, 403–428.

Süß, H., Oberauer, K., Wittman, W., Wilhelm, O., & Shulze, R. (2002). Working memory capacity explains reasoning ability- and a little bit more. *Intelligence, 30*, 261–288.

Tan, L., & Ward, G. (2000). A recency-based account of primacy effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1589–1625.

Towse, J. N., & Hitch, G. J. (1995). Is there a relationship between task demand and storage space in tests of working memory capacity? *Quarterly Journal of Experimental Psychology, 48*, 108–124.

Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A re-evaluation of working memory capacity in children. *Journal of Memory and Language, 39*, 195–217.

Towse, J. N., Hitch, G. J., & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition, 28*, 341–348.

Towse, J. N., Hitch, G. J., & Hutton, U. (2002). On the nature of the relationship between processing activity and item retention in children. *Journal of Experimental Child Psychology, 82*, 156–184.

Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*, 1–12.

Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language, 49*, 446–468.

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

Unsworth, N., & Engle, R. W. (2006b). A temporal-contextual retrieval account of complex span: An analysis of errors. *Journal of Memory and Language, 54*, 346–362.

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.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 49A*, 51–79.