

# Short-term retention of a single word relies on retrieval from long-term memory when both rehearsal and refreshing are disrupted

Nathan S. Rose · Bradley R. Buchsbaum ·  
Fergus I. M. Craik

© Psychonomic Society, Inc. 2014

**Abstract** Many working memory (WM) models propose that the focus of attention (or primary memory) has a capacity limit of one to four items, and therefore, that performance on WM tasks involves retrieving some items from long-term (or secondary) memory (LTM). In the present study, we present evidence suggesting that recall of even one item on a WM task can involve retrieving it from LTM. The WM task required participants to make a deep (living/nonliving) or shallow (“e”/no “e”) level-of-processing (LOP) judgment on one word and to recall the word after a 10-s delay on each trial. During the delay, participants either rehearsed the word or performed an easy or a hard math task. When the to-be-remembered item could be rehearsed, recall was fast and accurate. When it was followed by a math task, recall was slower, error-prone, and benefited from a deeper LOP at encoding, especially for the hard math condition. The authors suggest that a covert-retrieval mechanism may have refreshed the item during easy math, and that the hard math condition shows that even a single item cannot be reliably held in WM during a sufficiently distracting task—therefore, recalling the item involved retrieving it from LTM. Additionally, performance on a final free recall (LTM) test was better for items recalled following math than following rehearsal, suggesting that initial recall following math involved elaborative retrieval

from LTM, whereas rehearsal did not. The authors suggest that the extent to which performance on WM tasks involves retrieval from LTM depends on the amounts of disruption to both rehearsal and covert-retrieval/refreshing maintenance mechanisms.

**Keywords** Short-term memory · Working memory · Long-term memory

## Introduction

Cognitive psychologists and neuroscientists have increasingly acknowledged the close relation between short-term or working memory (WM) and long-term memory (LTM; see Jonides, Lewis, Nee, Lustig, Berman & Moore, 2008, for a review). For example, views on the capacity of WM have systematically shrunk from Miller’s (1956) magical number  $7 \pm 2$ , to Cowan’s (2001) magical number 4, to others who now place the limit at one item (McElree, 2006). Thus, given that WM tests typically require recalling more than one item, recall on tests ostensibly measuring WM may actually involve LTM processes to a substantial degree.

*Store* models of WM suggest that a small number of items (e.g., four) may be maintained in a temporary store (Atkinson & Shiffrin, 1968; Waugh & Norman, 1965)—for example, domain-specific buffers (Baddeley, 1986) or primary memory (Unsworth & Engle, 2007); accessing further items must involve retrieving them from LTM via the episodic buffer (Baddeley, 2000) or secondary memory (Atkinson & Shiffrin, 1968; Unsworth & Engle, 2007; Waugh & Norman, 1965). *State* models suggest that information “in WM” consists of representations that vary in their level of activation. Cowan (2001) suggested that four chunks of information may be within the focus of attention; accessing other recently processed information requires retrieval from the activated portion of LTM. McElree (2006) suggested that only

---

**Electronic supplementary material** The online version of this article (doi:10.3758/s13421-014-0398-x) contains supplementary material, which is available to authorized users.

---

N. S. Rose · B. R. Buchsbaum · F. I. M. Craik  
Rotman Research Institute at Baycrest, University of Toronto,  
Toronto, Ontario, Canada

N. S. Rose (✉)  
Department of Psychiatry, University of Wisconsin, 6001 Research  
Park Blvd, Madison, WI 53715, USA  
e-mail: nsrose@wisc.edu

one item may be in the focus of attention at any given time and that accessing other items requires retrieval from LTM. Oberauer (2002) also suggested that only one item can be in the focus of attention, but that a small number (e.g., three) of recently processed items may remain in a “region of direct access” or “broad focus” (Oberauer & Hein, 2012); accessing items outside this region involves retrieving them from the activated portion of LTM.<sup>1</sup>

In sum, the models reviewed above all claim that at least one item can be maintained in WM over short delays. Recently, we presented evidence that patient H.C., an amnesic with LTM impairment due to damage to her hippocampus, could not reliably maintain a single item in WM if rehearsal was disrupted and/or the stimulus was novel (Rose, Olsen, Craik & Rosenbaum, 2012). This finding is consistent with recent views that counter the traditional suggestion of a neuropsychological double dissociation between WM and LTM (e.g., Jonides et al., 2008; Ranganath & Blumenfeld, 2005). Here we present further evidence showing that performance on a WM task requiring the recall of just one word after 10 s can involve retrieving it from LTM.

One of the ways in which we assessed the involvement of LTM in performance on WM tasks was by examining whether deeper levels of processing (LOP) at encoding benefit a WM task, as they do for LTM tests. Examining LOP effects on an initial WM task and a subsequent LTM test offered an indicator of the involvement of LTM retrieval on the initial WM task, because manipulating LOP at encoding has one of the strongest and most reliable effects on (explicit) LTM tests (Craik & Tulving, 1975). In contrast, as we will discuss in the next section, a deeper LOP often does not benefit WM.

### LOP effects on WM

WM and LTM have long been thought to dissociate in terms of LOP effects (e.g., Craik & Jacoby, 1975; Mazuryk & Lockhart, 1974; for reviews, see Rose, 2010; Shivde & Anderson, 2011). However, recent research has shown both similarities and differences in LOP effects on WM and LTM tests. Rose, Myerson, Roediger and Hale (2010) developed a complex WM span task that manipulated LOP at encoding, but they did not find a benefit of deeper processing on WM recall (see also Flegel & Reuter-Lorenz, 2014; Loaiza &

Camos, 2013; Rose, 2013). In contrast, Loaiza, McCabe, Youngblood, Rose and Myerson (2011) found a benefit of deeper LOP on more traditional complex WM span tasks (i.e., reading span and operation span). Rose (2010) used Craik and Tulving’s (1975) original LOP materials in a modified reading span task, in which participants either did or did not expect to receive immediate recall tests, and found a benefit of deeper LOP, but only for longer (eight-item) list lengths when the tests were unexpected. To account for the variable pattern of LOP effects on WM tests in the literature, Rose and Craik (2012) calculated the correlation between the size of the LOP effect on immediate recall and the difference between deep and shallow LOP decision times in eight independent experiments. The difference in LOP decision times between deep and shallow LOP tasks was taken as a proxy for a difference in secondary task difficulty—that is, a difference in the amount of time that attention was drawn away from maintaining to-be-remembered items between the two conditions. Rose and Craik found a significant correlation, which led them to hypothesize that the extent to which deeper LOP benefits performance on WM tasks depends on the amount of disruption to active maintenance processes, and therefore on the amount of retrieval from LTM. However, this hypothesis was based on correlational data and was susceptible to methodological differences across experiments.

The primary goal of the present study was to directly test the hypothesis that a deeper LOP at encoding benefits WM performance to the extent that active maintenance processes are disrupted. We used a novel paradigm in which the amount of disruption to active maintenance processes was parametrically manipulated while the processing duration, set size, and retention interval were held constant. In prior research with WM span tasks, it was unclear how conditions differed with respect to the amount of attention that could be devoted to either maintaining the memoranda or performing the secondary distractor task. A clearer manipulation of LTM involvement was accomplished by comparing conditions that differed only in the amount of disruption to active maintenance processes. What might these active maintenance processes be?

WM maintenance mechanisms: Rehearsal, refreshing, and covert retrieval

Many models propose that items can be actively maintained in WM by at least two distinct types of maintenance mechanisms (Burgess & Hitch, 1999; Camos, Lagner & Barrouillet, 2009; Cowan, 1992). Articulatory *rehearsal* involves continuous (overt or covert) repetition of articulatory/phonological codes and is effective for maintaining verbal information for WM recall, provided that rehearsal is not disrupted by articulatory suppression (Baddeley, 1986) or interference from overlapping phonological codes (e.g., Camos, Mora & Oberauer, 2011). Some models propose that when to-be-remembered

<sup>1</sup> The terminology used in this field is quite variable, even when researchers refer to seemingly similar concepts. Here, we use the term “WM” to refer to both WM and STM concepts. The terms “primary” and “secondary memory” refer to theoretical memory stores tapped principally by WM and LTM tests, respectively. Because maintaining information in primary memory is conceptually similar to maintaining information in the focus of attention in performance on (Cowan, personal communication), here we use the term “focus of attention” to refer to both concepts, and the terms “WM” and “LTM” as shorthands to refer to the collection of processes involved in performance in WM and LTM tests, respectively.

items cannot be continuously rehearsed in a phonological loop (Burgess & Hitch, 1999) or maintained in the focus of attention (Cowan, 1992), participants may try to periodically return the items to the focus of attention by *refreshing* them—that is, by “thinking briefly of a just-activated representation” (Camos et al., 2009, p. 458; see also Johnson, 1992).<sup>2</sup>

Is “thinking” of a recently processed representation the same as rehearsing it? If participants refresh to-be-remembered items by rehearsing them, then refreshing and rehearsal would affect performance equally. However, because it is hypothesized that participants refresh to-be-remembered items when it is difficult to rehearse them, refreshing and rehearsal should affect performance differently. Indeed, by orthogonally manipulating the nature and pace of distraction on complex WM span tasks, Camos et al. (2009) showed that rehearsal and refreshing had independent effects on performance (see also Hudjetz & Oberauer, 2007). As opposed to rehearsal, refreshing to-be-remembered items may involve *retrieving* them from activated LTM (Rose & Craik, 2012; see also Cowan, 1995). For example, Camos et al. (2011) showed that refreshing eliminated a phonological similarity effect that was observed under rehearsal conditions, suggesting that refreshing involved the retrieval of more semantic features, similar to LTM retrieval. Thus, it is unclear whether “refreshing” an item is conceptually different from covertly retrieving the item from LTM. Clarifying the distinction between rehearsal, refreshing, and covert retrieval was a second goal of the present study.

#### Differential forgetting following rehearsal versus retrieval

One way to reveal the involvement of different maintenance mechanisms in WM processing would be to examine the consequences of using such maintenance mechanisms to subsequent LTM (e.g., final free recall of the items; McCabe, 2008; Rose et al., 2010). Rehearsal of articulatory/phonological codes can support WM recall, but such shallow codes are typically suboptimal for retrieval on LTM tests (Craik & Tulving, 1975). Consider some classic examples that demonstrated the ineffectiveness of rehearsal for promoting long-term retention. Craik and Watkins (1973) had participants maintain one word for initial recall, varied the amount of time that participants spent rehearsing the word, and then examined the impact on final free recall. They found no relation between the amount of time spent rehearsing the word and final free recall. Jacoby and Bartz (1972) had participants

initially recall lists of words after either 15 s of rehearsal or 15 s of a rehearsal-preventing task, and then compared final free recall of the words. Final free recall was best when initial recall was preceded by 15 s of distraction; 15 s of rehearsal contributed little to final recall.

Although it was initially suggested that “control processes such as ‘rehearsal’ are essential to the transfer of information from the short-term store to the long-term one” (Atkinson & Shiffrin, 1971, p. 82), rote rehearsal is widely considered to be suboptimal for long-term retention (e.g., Craik & Watkins, 1973; Jacoby & Bartz, 1972; for a review, see Delaney, Verhoeven & Spigler, 2010). In contrast, recall following a distracting task involves retrieving the item from LTM—a process that consists of a more reconstructive or elaborative form of recall (Craik & Jacoby, 1975). The advantage to subsequent LTM that retrieval following distraction has over rehearsal is analogous to the advantage of retrieval practice or testing over restudying (i.e., the testing effect) (for reviews, see Delaney et al., 2010; Roediger & Butler, 2011).

A leading account of the testing effect is the elaborative-retrieval account. According to this account, effortful retrieval from LTM involves deeper, more elaborative retrieval operations that activate related concepts and generate effective retrieval cues or “routes” to a target memory item on later retrieval attempts (Carpenter, 2009; Delaney et al., 2010; Pyc & Rawson, 2010; Roediger & Butler, 2011). In contrast, similar to restudying, rehearsing and reporting an item directly from the focus of attention does not generate cues, because the item is directly available (McCabe, 2008; Rose et al., 2010). Therefore, if “refreshing” is similar to rehearsal, it should affect subsequent LTM in a way similar to rehearsal. In contrast, if the two maintenance mechanisms differ, as is hypothesized (Camos et al., 2009; Camos et al., 2011), they should affect subsequent LTM differently. To test this hypothesis, we compared subsequent memory for items that had initially been recalled on a WM task in conditions that varied in the amount of rehearsal or refreshing opportunities on a final free recall (LTM) test administered 10 min after completion of the WM task.

#### The present study

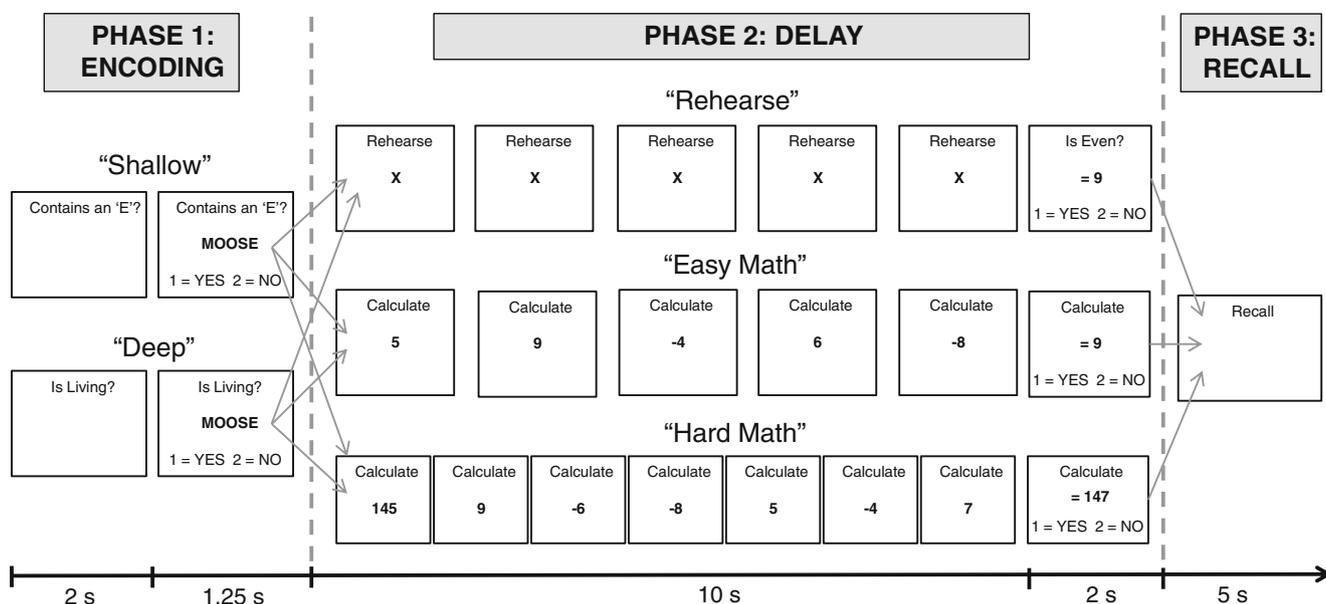
For the present study, we employed a novel design that tested what might be considered a boundary condition for the distinction between WM and LTM. We administered a WM task that required recalling just one word after a 10-s delay on each trial. Moreover, the word was a short, high-frequency, concrete noun, so one might expect that the word could be easily recalled on each trial in this task. After participants had encoded the word in a deep or shallow manner on each trial, they either rehearsed the word or performed an easy math task—in which it was possible to retrieve/refresh the to-be-remembered word between math operations—or a hard math

<sup>2</sup> Note that some models assume no role for active maintenance mechanisms to prevent forgetting (e.g., G. D. A. Brown, Preece & Hulme, 2000; Lewandowsky & Oberauer, 2009). For now, we remain agnostic as to whether the source of forgetting is interference (Nairne, 1990; Saito & Miyake, 2004) or decay (Barrouillet et al., 2004; Towse, Hitch & Hutton, 2000), and focus instead on testing models that hypothesize roles of rehearsal, refreshing, and retrieval in supporting WM.

task—in which it was difficult to retrieve/refresh the to-be-remembered word before attempting to recall the word (see Fig. 1 for details).

If one WM task (e.g., the hard math condition) were more disruptive to active maintenance processes, thereby placing greater demands on LTM search and retrieval processes, than another WM task (e.g., the easy math condition), the former condition should demonstrate a larger benefit of deep encoding over shallow encoding than would the latter condition. Surprisingly little research has examined LOP effects using this type of approach; however, some research on the Brown–Peterson task (J. Brown, 1958; Peterson & Peterson, 1959) has provided a basis on which to make some predictions. Elmes and Bjork (1975) instructed participants to encode five words by either semantically associating the words or rehearsing the words, and then to count back by 3 s from a number for 0, 4, 12, or 18 s before attempting to recall the words. Deep encoding had no benefit to recall for the 0-s delay, a small benefit for the 4-s delay, and a large benefit for the 12- and 18-s delays. Marsh, Sebrechts, Hicks and Landau (1997) instructed participants to read or make semantic or acoustic judgments on three words and then to count back by 3 s from a number for 0, 2, or 4 s. Participants were led to believe that they would not have to recall the words; however, surprise recall tests revealed a benefit of semantic over acoustic encoding. These experiments demonstrated that disrupting active maintenance processes by having participants do mental arithmetic and either increasing the delay or reducing test expectancy produced a larger benefit of a deep LOP at encoding to WM recall.

In the present study, we varied the extent to which maintenance of one to-be-remembered word was possible by manipulating delay activity, while keeping delay and test expectancy constant. Initial recall after a rehearsal-filled delay should involve reporting the word directly from the focus of attention, and should therefore be rapid, accurate, and unaffected by the LOP at encoding. For the math-filled delay conditions, technically only two “items” needed to be maintained on each trial—the to-be-remembered word and the current arithmetic sum. Therefore, in the strictest sense, models that assume that up to four items may be maintained in the focus of attention (e.g., Cowan, 2001; Unsworth & Engle, 2007) should predict that recalling the word following a math-filled delay in this task should also involve reporting it directly from the focus of attention, and should therefore be rapid, accurate, and unaffected by the LOP at encoding. In contrast, according to McElree (2006), recall in both easy and hard math conditions should involve retrieval from LTM, and therefore both conditions should benefit equally from a deeper LOP at encoding. If, however, participants were able to covertly retrieve the item to refresh its activation during easy-math delays more so than during the hard-math delays, then recall should involve retrieving a refreshed representation of the item (perhaps from the “broad focus of attention”; Oberauer & Hein, 2012), and may therefore show a negligible LOP effect in the easy relative to the hard math condition. That is, if refreshing is used to restore the item’s accessibility, a deeper LOP at encoding should have minimal effect on recall in the easy math condition, similar to what was observed in prior studies (Loaiza & Camos, 2013; Mazuryk & Lockhart, 1974; Rose, 2013; Rose & Craik, 2012, Exp. 1; Rose et al., 2010).



**Fig. 1** Procedure of the working memory task. On each trial, participants made a shallow (“Does the word contain an ‘e’?”) or deep (“Is the word something living?”) level-of-processing judgment on a single word and

tried to recall the word after a 10-s delay. During the delay, participants either rehearsed the word or performed an easy or a hard math task. Note that the to-be-remembered words were counterbalanced across conditions

In contrast to recall on the WM task, final free recall of the words from the initial WM task after a delay of 10 min is a purer test of recall from LTM, and all models would predict a benefit of deeper LOP at encoding. The amount of forgetting of words initially recalled following a rehearsal-, easy-math-, or hard-math-filled delay can shed additional light on the nature of recall from WM. If initial recall for all conditions involved reporting the item from the focus of attention, the amounts of forgetting should be similar across the conditions. If, however, initial recall following easy or hard math necessitates retrieval from LTM, this should provide a *benefit* to subsequent memory relative to final recall of items initially recalled following rehearsal. Moreover, if refreshing is conceptually similar to rehearsal, then final recall should be similar for items from the easy math condition and the rehearsal condition.

To clarify, our hypotheses and predictions were as follows: In the rehearsal condition, we hypothesized that participants would maintain items via articulatory rehearsal and, thus, that initial recall would involve reporting the item directly from the focus of attention. As a result, initial recall should be rapid and accurate, with no effect of LOP; final free recall should be relatively poor, because reporting an item from focal attention does not involve elaborative retrieval from LTM. In the easy math condition, we hypothesized that doing math would disrupt articulatory rehearsal, but that most participants would be able to maintain the to-be-remembered item by refreshing the item between (at least the first few) computations during the delay. Thus, initial recall would involve reporting a refreshed representation on most trials, and so should be less dependent on having encoded deeper cues for retrieval. Therefore, initial recall should show a reduced LOP effect relative to the hard math condition. Nonetheless, because recall following easy math would involve more effortful search and retrieval (i.e., a more elaborative form of retrieval) than the rehearsal condition, final free recall should be better for items from the easy math than from the rehearsal condition. In the hard math condition, we hypothesized that doing math would disrupt both articulatory rehearsal and the ability to refresh the to-be-remembered item. Thus, initial recall would involve elaborative retrieval from LTM, and so a large benefit of deeper LOP at encoding should appear, and final free recall should be better for items from the hard math than from the rehearsal condition. To summarize, Table 1 depicts the hypothesized active maintenance processes involved during the delay and the hypothesized retrieval processes involved during recall in the three delay conditions, as well as their effects on both the LOP effect on the initial WM test and overall performance on the final free recall test.

**Table 1** Hypothesized active maintenance process involved during the delay and retrieval process during recall, and their effects on the level-of-processing (LOP) effects on working memory and final free recall (FFR) for the three experimental conditions

Condition	Process Involved During		Effect On	
	Delay	Recall	LOP	FFR
Rehearsal	rehearsal	report from focal attention	–	–
Easy math	refreshing	elaborative retrieval	–	+
Hard math	neither	elaborative retrieval	+	+

Negative and positive signs indicate the direction of the predicted effects.

## Method

### Participants

A group of 47 University of Toronto undergraduate students participated in the experiment. The participants were fluent English speakers.<sup>3</sup>

### Design and procedure

The design was a 2 (LOP: deep, shallow) × 3 (delay condition: rehearse, easy math, hard math) × 2 (test: WM, LTM) within-subjects design. See Fig. 1 for details of the WM task procedure. The LOP task at encoding required indicating whether the to-be-remembered word represented something living (deep) or contained an “e” (shallow) by pressing the left (“no”) or right (“yes”) mouse button. Half of the words represented something living, and half of the words contained an “e”; the words were counterbalanced across conditions. Trials for the different conditions were mixed randomly, with equal numbers per block. The words were relatively short (mean length=5.9 letters), high-frequency (mean log-HAL frequency=8.52), concrete English nouns. The math task involved adding a series of five (easy math condition) or seven (hard math condition) numbers presented individually at a rate of 2,000 (easy math condition) or 1,430 ms/number (hard math condition). The first number for hard-math delays was a number randomly selected between 100 and 200; the other numbers for hard-math delays, and all numbers for the easy-math delays, consisted of numbers randomly selected between 9 and –9 (excluding those from 3 to –3). Pilot testing indicated that adding or subtracting 0, 1, 2, or 3 from the current sum was considerably less distracting than adding or subtracting 4, 5, 6, 7, 8, or 9. To ensure equal amounts of distraction on each computation, the numbers between 3 and –3 were therefore excluded. After the numbers were presented, either the correct

<sup>3</sup> Of the 47 participants, one participant was excluded from the analysis because of LOP decision accuracy below 70 %, and five were excluded because of overall math accuracy below 60 %.

sum or  $\pm 1$  the correct sum was presented; participants were to indicate whether or not the sum was correct by pressing the right or the left mouse button, respectively. On rehearsal trials, participants were told to “repeat the word over and over in your head during the delay.” At the end of rehearsal-filled delays, participants were asked to press either the right or the left mouse button at random. Following the delay, they were given 5 s to recall the word by saying it aloud. An experimenter recorded accuracy and reaction times via a buttonpress time-locked to response termination.

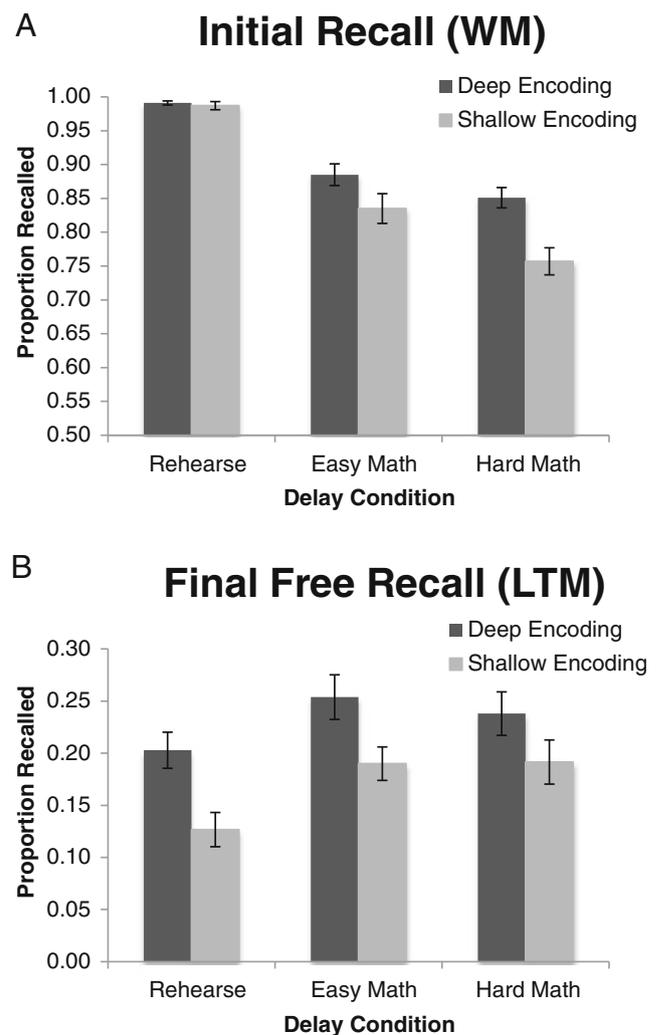
Participants performed five blocks of WM recall trials, with 24 trials per block, thereby providing four observations for each of the six (LOP $\times$ delay) conditions per block. Feedback was provided on the LOP decision and math accuracies between blocks. Participants were explicitly told not to sacrifice performance on the math task so that they could recall the word—that is, both tasks were equally important. After performing all 120 WM trials, participants performed a distractor task (Tetris) for 10 min. Then participants were administered a surprise final free recall (LTM) test, in which they were asked to recall as many words from the experiment as possible and write them down on a sheet of paper over 5 min. At the end of the experiment, participants were administered a questionnaire about their use of maintenance strategies during the WM task (e.g., “Could you think of the word during the easy or the hard math task?”).

## Results

We first verified that the LOP decisions at encoding did not differ between the deep and shallow conditions in terms of accuracy (93 % vs. 92 %),  $t(40)=1.12$ ,  $p=.27$ , or reaction time (891 vs. 883 ms),  $t(40)=0.82$ ,  $p=.42$ . In a further manipulation check, we confirmed that math performance was indeed faster and more accurate on the easy math than on the hard math task [reaction time: 1,002 vs. 1,205 ms,  $t(40)=-12.4$ ,  $p<.001$ ; accuracy: 90 % vs. 63 %,  $t(40)=15.9$ ,  $p<.001$ ].

### WM performance

First, performance on the initial WM task was considered. As predicted, recall on the WM task was better following a rehearsal-filled delay than following a math-filled delay, and deep processing at encoding benefited WM recall most in the hard-math delay condition (see Fig. 2A). These data were analyzed with a 2 (LOP: deep, shallow) $\times$ 3 (delay condition: rehearsal, easy math, hard math) repeated measures analysis of variance (ANOVA). We found a significant LOP $\times$ Delay Condition interaction,  $F(2, 80)=9.25$ ,  $p<.001$ ,  $\eta_p^2=.188$ . The LOP effect (i.e., the difference between deep and shallow processing) was not significant for the rehearsal-filled delay condition (0 %),  $t(40)=0.70$ ,  $p=.49$ ; was small for the easy-math delay condition (4 %),  $t(40)=2.81$ ,  $p<.01$ ,  $d=0.89$ ; and



**Fig. 2** (A) Mean proportions of words recalled on the initial (WM) task for deep versus shallow processing at encoding, following a rehearsal-, easy-math-, or hard-math-filled delay. (B) Mean proportions of words recalled on the initial (WM) task from the different conditions that were recalled again on the subsequent, final free recall (LTM) test, which was administered after participants had performed the WM task and played Tetris for 10 min. Error bars are  $\pm 1$  SEM

was comparatively large for the hard-math delay condition (9 %),  $t(40)=5.03$ ,  $p<.001$ ,  $d=1.59$ . To see whether the LOP effect was larger for the hard-math than the easy-math delay condition, a separate 2 (deep, shallow) $\times$ 2 (easy math, hard math) repeated measures ANOVA was conducted. The interaction was significant,  $F(1, 40)=4.4$ ,  $p=.04$ ,  $\eta_p^2=.10$ , confirming that the LOP effect was larger for the hard-math delay condition than for the easy-math delay condition.

The average recall response times were submitted to the same 2 (LOP) $\times$ 3 (delay condition) ANOVA. Recall response times differed across the three delay conditions,  $F(2, 80)=43.91$ ,  $p<.001$ , because recall was faster following a rehearsal-filled delay (1,364 ms) than following an easy-math delay (1,588 ms),  $t(40)=3.28$ ,  $p=.002$ , which in turn was faster than recall following a hard-math delay (1,957 ms),

$t(40)=4.45, p<.001$ . LOP at encoding did not affect recall response times and did not interact with delay condition,  $F_s<1$ .

Taken together, the initial recall accuracy and response time data suggest that recalling the to-be-remembered word following hard math involved slower, cue-driven search and retrieval from LTM, whereas recalling the word following a rehearsal-filled delay involved reporting it directly from focal attention (or readout from primary memory; Unsworth & Engle, 2007). Next, we considered the consequences of the differences in the nature of initial recall between the three conditions of the WM task for subsequent final free recall of the items on the LTM test.

### LTM performance

We analyzed the proportions of words recalled on the final free recall test, conditionalized on initial recall in order to control for baseline differences in the levels of initial recall across conditions. Nonetheless, the patterns of results were similar for the nonconditionalized data (see the [supplemental materials](#)).

In contrast to performance on the initial WM task, subsequent final free recall on the LTM test was worst for items initially recalled following a rehearsal-filled delay, and was best for deeply processed items, regardless of the initial delay condition (see Fig. 2B). The data were analyzed with a 2 (LOP: deep, shallow)  $\times$  3 (delay condition: rehearse, easy math, hard math) repeated measures ANOVA. The main effect of LOP was significant, because there was an overall benefit to LTM recall of deep (23.2 %) over shallow (16.9 %) encoding,  $F(1, 40)=47.66, p<.001, \eta_p^2=.544$ . The main effect of delay condition was also significant,  $F(2, 80)=6.44, p=.003, \eta_p^2=.139$ , because LTM recall was better for words initially recalled on the WM task in both the easy-math (22.2 %) and hard-math (21.5 %) delay conditions than for those in the rehearsal-filled delay condition (16.5 %),  $t(40)>2.57, ps\leq.014$ . This finding suggests that initially recalling information after either an easy or hard math task had a beneficial impact on LTM for that information, relative to immediately recalling information being rehearsed in focal attention. That is, LTM recall was significantly better for words from the easy math condition than for those from the rehearsal condition, which suggests that refreshing and rehearsal did not have similar consequences for subsequent LTM. The interaction between LOP and delay condition was not significant,  $F<1$ .

Before discussing theoretical interpretations of these findings, it is important to address a potential criticism, which concerns whether recall differed for trials on which math performance was incorrect. Such a result might suggest that participants sacrificed math performance in order to rehearse the word and support recall. However, recall was not different on correct and incorrect math trials. The mean numbers of trials and mean (*SD*) proportions of items recalled on the initial recall

and final free recall tests on trials in which the math task was correct or incorrect are presented in the [supplemental materials](#).

## Discussion

LOP effects on WM depend on the amount of disruption to active maintenance

The results of the present experiment showed that recalling one word on a WM task can benefit from deeper LOP at encoding, which is one piece of evidence that suggests that recall involved retrieving the word from LTM. Yet, the amount of benefit (i.e., the extent to which LTM was involved in WM) depended on the amount of disruption to active maintenance processes. When the to-be-remembered word could be rehearsed in mind, accessing it was rapid, and accuracy was nearly perfect. The differences between the rehearsal condition and the math conditions in the speed and accuracy of initial recall on the WM test, and the rate of forgetting on the LTM test, all suggest that in the rehearsal condition of the WM test the word was reported directly from the focus of attention, and accessing it did not involve retrieval processes per se. In contrast, performing a hard math task during the delay disrupted active maintenance processes, so WM recall was slower and error-prone, and there was a large LOP effect which suggests that recall involved cue-dependent search and retrieval from LTM.

Performing an easy math task during the delay also disrupted rehearsal, but participants were likely able to covertly retrieve the word between math operations in order to refresh the word's accessibility. To further test the idea that attentional refreshing/covert retrieval supported WM maintenance during the easy math condition more than during the hard math condition, we considered participants' reported strategies on the postexperiment questionnaire. Participants were asked whether they were able to "think of the word" during the math-filled delays on the initial WM task.<sup>4</sup> In total, 59 % of the participants responded "yes" and 41 % responded "no," which was interpreted to indicate whether the participant did or did not attempt to use a covert-retrieval/refreshing maintenance strategy on either the easy math or hard math trials.

We reanalyzed the WM recall data to see whether performance differed between those who did report thinking of the word during the math-filled delays and those who did not. The interaction between reported maintenance strategy (covert-retrieval/refreshing, no covert-retrieval/refreshing) and delay condition (easy, hard) was significant,  $F(1, 20)=5.56, p=.03$ . WM recall in the easy math condition was higher for those

<sup>4</sup> The reported strategies were recorded for only 22 of the 41 participants, due to experimenter error.

who reported covert-retrieval/refreshing (90.4 %) than for those who did not (79.7 %),  $t(20)=2.80$ ,  $p<.05$ , and WM recall in the hard math condition did not differ between these groups (82 % and 80 %, respectively),  $t(20)=0.41$ ,  $p=.69$ . Notably, WM recall was not better for the easy than for the hard math condition for those who did not report refreshing (79.7 % vs. 80 %),  $t(8)=0.18$ ,  $p=.87$ , but it was significantly better for those who did (90.4 % vs. 82 %),  $t(12)=3.73$ ,  $p<.01$ . This pattern supports our hypothesis that most participants were able to covertly retrieve/refresh the to-be-remembered word during a math-filled delay and that doing so supported WM recall, particularly for the easy math condition.

Because the strategy questionnaire did not distinguish between easy-math and hard-math trials and because of the reduced sample size, interpretation of this result should be treated with caution. It is nonetheless reassuring that this result is compatible with the main finding of the experiment that, as compared to the hard math condition, WM recall following easy math was faster and more accurate, and the LOP effect was substantially smaller. Future studies should consider a more direct probe for use of a refreshing maintenance strategy on a trial-by-trial basis.

Taken together, these findings are consistent with the hypothesis proposed by Rose and Craik (2012) that the extent to which deeper LOP benefits performance on WM tasks—that is, the amount of LTM involvement on WM performance—depends on the amount of disruption to active maintenance processes. Addressing this hypothesis was a primary goal of the present study.

#### Two maintenance mechanisms: Rehearsal and retrieval

Another goal of the present study was to examine the existence of a putative attentional maintenance mechanism—“refreshing.” The reason that the LOP effect was larger for the hard than for the easy math condition was likely because performing the hard math task generally required more WM resources. The hard math task involved larger numbers and more computations than did the easy math task. Therefore, during the delay, more information had to be maintained, and the arithmetic computations had to be conducted more quickly in the hard than in the easy math condition. These features likely resulted in the hard math task requiring more WM resources than did the easy math task. What might these “resources” be? We propose that the capacity-limited focus of attention is necessary to engage in attentional refreshing/covert retrieval (Camos et al., 2011; Johnson, 1992), and the reason that recall following hard math involved more retrieval from LTM was that the hard math task more fully occupied the focus of attention than did the easy math task, so that less of this resource was available for attentional refreshing/covert retrieval.

The pattern of subsequent memory on the final free recall (LTM) test also argues for the existence of two maintenance mechanisms that were involved in the initial WM task. If the nature of processing on the initial WM task was similar following rehearsal and math-filled delays, then similar rates of forgetting would be expected. However, LTM recall was better for items that were initially recalled after a math-filled delay than after a rehearsal-filled delay. This “subsequent memory effect” replicates observations comparing the final free recall or recognition of items from simple versus complex span tasks (McCabe, 2008; Rose, 2013) and subspan versus supraspan lists (Rose et al., 2010), and extends it to trials with single words as the memoranda and with LOP decision times and retention intervals equated for all conditions. Thus, recalling a word just rehearsed impacts subsequent memory differently than recalling a word recently refreshed.

The different rates of forgetting reflect the consequences of rehearsing an item in the focus of attention to maintain it in WM, relative to retrieving it from LTM (e.g., Craik & Watkins, 1973; Jacoby & Bartz, 1972). Relative to retrieval from LTM, rehearsing and reporting an item directly from the focus of attention provides less benefit to subsequent memory (a phenomenon also illustrated by the negative recency effect; Craik, 1970). That is, the act of retrieval from LTM benefits subsequent LTM more than does reporting a recently rehearsed item (a phenomenon analogous to the retrieval practice or testing effect; Roediger & Butler, 2011), because effortful retrieval from LTM involves deeper, more elaborative retrieval operations, which generate more effective cues for later retrieval than does rehearsing and reporting items from the focus of attention (Carpenter, 2009; Craik & Jacoby, 1975; Delaney et al., 2010). For example, Whitten and Bjork (1977) noted that “the effect of a test on subsequent memory may depend on the level-of-processing or quality of information used during the test” (p. 473). That final recall was better for items from the easy math condition than for those from the rehearsal condition is consistent with the notion that recall of a recently refreshed item is a deeper, more elaborative form of retrieval than is recalling an item recently rehearsed.

To further test the hypothesis that effortful retrieval following a math task might involve elaborative processing that would generate effective retrieval cues for recall on a subsequent LTM test, we conducted follow-up analyses. Follow-up  $t$  tests were conducted to see whether deep encoding “rescued” the final recall of items that were initially recalled after a rehearsal-filled delay (i.e., shallow retrieval) and, analogously, whether initial recall following a math-filled delay (i.e., deep retrieval) “rescued” shallowly encoded items. Indeed, final recall of deeply encoded items was not significantly worse for items initially recalled following a rehearsal-filled delay than following either an easy- or a hard-math delay,  $t_s<1.84$ ,  $p_s\geq.073$ , and final recall of shallowly encoded items was

significantly better when they were initially recalled after either an easy- or a hard-math delay than after a rehearsal-filled delay,  $t_s > 2.87$ ,  $p_s \leq .0065$  (see Fig. 2). Put another way, the combination of shallow processing at encoding and shallow processing at retrieval (i.e., reporting a shallowly encoded word from focal attention following rehearsal) resulted in the worst final recall of all of the conditions. This pattern is particularly striking when considered alongside the initial recall results, in which shallow processing at encoding and shallow processing at retrieval produced near-perfect WM.

Notably, the amount of benefit to subsequent memory relative to the rehearsal condition was similar for the easy and hard math conditions. If refreshing processes are the same as rehearsal processes, then final free recall of words from the easy math condition should have been similar to final free recall of words from the rehearsal condition, but it was not. That final free recall of words from the easy math condition was similar to final free recall of words from the hard math condition suggests that retrieval of recently refreshed items has the same consequences for subsequent memory as retrieval from LTM. Therefore, elaborative retrieval appears to have been involved in both the easy and hard math conditions, suggesting that the concept of refreshing may be closer to the concept of retrieval from LTM than to the concept of rehearsal in the focus of attention. Indeed, Barrouillet, Bernardin and Camos (2004) initially hypothesized, “as soon as attention is switched away from the memory traces of the items to be recalled, their activation suffers from a time-related decay. Refreshing these decaying memory traces requires their *retrieval from memory* . . .” (p. 84, italics added; see also Cowan, 1992). The present results provide some of the first evidence to support the notion that refreshing to-be-remembered representations involves their covert retrieval from (long-term) memory.

However, it is important to note a number of limitations with the present study. For example, if easy math involved more refreshing opportunities than did hard math, and if refreshing is indeed conceptually similar to retrieval from LTM, then why was final free recall not better for easy than for hard math? A limitation of the present design is that it is difficult to know exactly when participants refreshed the to-be-remembered words. Future studies would do well to have participants overtly rehearse items or to collect strategy reports on every trial in order to gauge whether items were refreshed during a delay and precisely when they were refreshed. That said, it is likely that requiring overt rehearsal or strategy reports could induce interference or the use of strategies that could affect performance in other ways. Another limitation of the present design is that we only included one final free recall test, after participants had processed all 120 words. As a result, overall recall performance was rather poor. It is possible that the test simply was not sensitive enough to detect a difference between final free recall of items from the easy and hard math conditions. Future research should include a more sensitive

test, such as free recall after each block of trials, for example. Another possibility is that the benefit of retrieval from LTM at recall dominated any effect that a difference between refreshing during easy math versus hard math could have conferred to long-term retention. Future research should compare subsequent memory of items as a function of refreshing by comparing, for example, subsequent memory on “catch” trials on which initial recall was not required, so that the effects of refreshing on long-term retention could be separated from the effects of actually recalling items.

#### Implications for the theoretical distinction between WM and LTM

The results of the present study provide important data from something of a boundary condition for the distinction between WM and LTM. The WM task in this study technically only required maintaining at most two items on each trial—the to-be-remembered word and the current arithmetic sum for math-filled delays. Nevertheless, recalling a word after 10 s of hard math demonstrated slower, error-prone recall that benefited from deeper processing at encoding and subsequently enhanced long-term retention, all of which suggests that recalling the item involved search and retrieval from LTM. Although the results cannot definitively clarify the role of LTM in performance on WM tasks and distinguish between contemporary models of WM, the results do present a challenge to models that assume a strict capacity limitation of up to four items that can be maintained in and reported directly from the focus of attention (e.g., Cowan, 2001; Unsworth & Engle, 2007).

It has been suggested that in some situations the focus of attention may “shrink” down to only one item and, depending on the kind of incoming information, the item can be displaced from the focus of attention (Cowan, 2005; Unsworth & Engle, 2008). Indeed, Atkinson and Shiffrin (1968) acknowledged that the distractor task in the Brown–Peterson paradigm prevents attention from being paid to the current item, which, they hypothesized, results in the item being lost from the short-term store, and therefore, recalling it would require retrieving it from the long-term store. More recent instantiations of Atkinson and Shiffrin’s “buffer” model maintain this account of performance on the Brown–Peterson task (Davelaar & Usher, 2002; Davelaar et al., 2005). Moreover, although most computational modeling with versions of Atkinson and Shiffrin’s buffer model has simulated data with a fixed buffer capacity of four items, updated versions include a flexible buffer whose size can change as a function of task demands (Lehman & Malmberg, 2013; Raaijmakers, 1993). For example, to simulate the effect of distraction, Lehman and Malmberg reduced their buffer capacity parameter to two items. According to this more-flexible view, it is possible that shifting attention to performing either an easy or a hard math task and adding or subtracting each

digit to the current sum could have caused the focus of attention to shrink to less than two items (even though it would not be advantageous to do so, because it would make recall harder).

However, if the capacity limit can flexibly vary between one and four items, it is difficult to see how results could definitively rule out one class of model over another. For example, how could one know when performance on a task involved reporting information from a zoomed-in or zoomed-out focus of attention or involved retrieval from LTM? If the capacity can be anything between one and four, its value is essentially a free parameter that can be estimated from the data ad hoc, thereby rendering the notion of a focus of attention with a capacity around four items untestable. An aim for future research should be to specify the mechanisms by which the focus of attention could expand or contract with greater precision. Perhaps a more agreeable view would be to conclude that the commonalities among contemporary models of WM may supersede their subtle differences.

Testing subtle differences between these models may benefit from using more than just behavioral data. For example, one critical difference between “store” and “state” models concerns their conceptualization of the structural architecture of WM. An inherent property of store models is a structural distinction between items maintained in WM and items retrieved from LTM. For example, in at least one computational model of WM (e.g., Just & Carpenter, 1992), items are literally copied from their LTM representations, and these copies or “tokens” are moved to a capacity-limited storage site (i.e., WM) where they are temporarily maintained. Because it is difficult to envisage the psychological and neural processes that are responsible for selecting, copying, and transferring information from one site to the other, we find this proposition psychologically unnecessary and neurologically implausible.

Neuropsychological, neuroimaging, and neurostimulation studies may therefore be informative for distinguishing between models of WM (for a recent review, see LaRocque, Lewis-Peacock, & Postle, 2014). For example, we recently showed that patient H.C., a medial-temporal-lobe amnesic, was impaired at maintaining a single nonword or unfamiliar low-frequency word, but was unimpaired at maintaining a familiar, high-frequency word (Rose et al., 2012). Here we have shown that a familiar, high-frequency, concrete noun could not even be reliably maintained in WM by healthy young adults under distracting conditions that were very demanding of attention.

## Conclusion

In our view, which is similar to state models (Cowan, 2008; Oberauer, 2009), items “in WM” are not in a separate temporary store. Rather, we prefer to view WM as attention paid to the particular features of representations that are relevant to

the task at hand, with information in the current focus of attention existing in a highly accessible state, ready for cognitive action without the need for retrieval per se. We suggest that WM and LTM differ in that the rehearsal of shallow, articulatory/phonological codes is typically sufficient for retrieving to-be-remembered items on verbal WM tasks, but deeper, conceptual/semantic cues are more effective for retrieval from LTM (Craig & Lockhart, 1972), which can occur on WM tasks, depending on the conditions (Rose & Craik, 2012). The present results show that a WM task involving recall of just one word after a short delay can involve retrieving it from LTM, provided that the distraction during the delay was sufficiently demanding of attention. We (among others; e.g., Speer, Jacoby & Braver, 2003; Unsworth, 2010) prefer a processing approach (as opposed to a structural approach) to the distinction between WM and LTM, because the extent to which performance on the present WM and LTM tests evoked similar processes depended on the task conditions—namely, the extent to which both rehearsal and covert-retrieval/refreshing maintenance mechanisms were disrupted. Future work should pursue more direct observation of this covert-retrieval/refreshing process in action and address whether recently retrieved representations really do exist in a distinct state.

## References

- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.
- Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American*, 224, 82–90.
- Baddeley, A. (1986). *Working memory*. Oxford: Oxford University Press, Clarendon Press.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417–423. doi:10.1016/S1364-6613(00)01538-2
- Barrouillet, P., Bernardin, S., & Camos, C. (2004). Time constraints and resource sharing in adults’ working memory spans. *Journal of Experimental Psychology: General*, 133, 83–100. doi:10.1037/0096-3445.133.1.83
- Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, 107, 127–181. doi:10.1037/0033-295X.107.1.127
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 10, 12–21.
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, 106, 551–581. doi:10.1037/0033-295X.106.3.551
- Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language*, 61, 457–469.
- Camos, V., Mora, G., & Oberauer, K. (2011). Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working

- memory. *Memory & Cognition*, 39, 231–244. doi:10.3758/s13421-010-0011-x
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1563–1569. doi:10.1037/a0017021
- Cowan, N. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language*, 31, 668–684.
- Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87–114. doi:10.1017/S0140525X01003922. disc. 114–185.
- Cowan, N. (2005). Understanding intelligence: A summary and an adjustable-attention hypothesis. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 469–488). Thousand Oaks: Sage.
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research* 169, 323–338. doi:10.1016/S0079-6123(07)00020-9
- Craik, F. I. M. (1970). The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*, 9(2), 143–148.
- Craik, F. I. M., & Jacoby, L. L. (1975). A process view of short-term retention. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. R. Lindman, & D. B. Pisoni (Eds.), *Cognitive theory* (pp. 173–192). Hillsdale: Erlbaum.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671–684. doi:10.1016/S0022-5371(72)80001-X
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268–294. doi:10.1037/0096-3445.104.3.268
- Craik, F. I. M., & Watkins, M. J. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 12, 599–607.
- Davelaar, E. J., & Usher, M. (2002). An activation-based theory of immediate item memory. In J. A. Bullinaria, & W. Lowe (Eds.), *Proceedings of the Seventh Neural Computation and Psychology Workshop: Connectionist Models of Cognition and Perception*. Singapore: World Scientific.
- Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review*, 112, 3–42. doi:10.1037/0033-295X.112.1.3
- Delaney, P. F., Verkoijen, P. J. L., & Spigel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 53, pp. 63–148). Orlando: Academic Press.
- Elmes, D. G., & Bjork, R. A. (1975). The interaction of encoding and rehearsal processes in the recall of repeated and nonrepeated items. *Journal of Verbal Learning and Verbal Behavior*, 14, 30–42.
- Flegal, K. E., & Reuter-Lorenz, P. A. (2014). Get the gist? The effects of processing depth on false recognition in short-term and long-term memory. *Memory & Cognition*. doi:10.3758/s13421-013-0391-9
- Hudjetz, A., & Oberauer, K. (2007). The effects of processing time and processing rate on forgetting in working memory: Testing four models of the complex span paradigm. *Memory & Cognition*, 35, 1675–1684.
- Jacoby, L. L., & Bartz, W. H. (1972). Encoding processes and the negative recency effect. *Journal of Verbal Learning and Verbal Behavior*, 11, 561–565.
- Johnson, M. K. (1992). MEM: Mechanisms of recollection. *Journal of Cognitive Neuroscience*, 4, 268–280.
- Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, 59, 193–224. doi:10.1146/annurev-psych.59.103006.093615
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122–149. doi:10.1037/0033-295X.99.1.122
- LaRocque J. J., Lewis-Peacock J. A., & Postle B. R. (2014). Multiple neural states of representation in short-term memory? It's a matter of attention. *Frontiers in Neuroscience*, 8(5), 1–14. doi:10.3389/fnhum.2014.00005
- Lehman, M., & Malmberg, K. J. (2013). A buffer model of encoding and temporal correlations in retrieval. *Psychological Review*, 120, 155–189. doi:10.1037/a0030851
- Lewandowsky, S., & Oberauer, K. (2009). No evidence for temporal decay in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1545–1551. doi:10.1037/a0017010
- Loaiza, V., & Camos, V. (2013). *Does controlling for temporal parameters change the levels-of-processing effect on working memory?* Washington: Poster presented at the annual meeting of the Association for Psychological Science.
- Loaiza, V. M., McCabe, D. P., Youngblood, J. L., Rose, N. S., & Myerson, J. (2011). The influence of levels of processing on recall from working memory and delayed recall tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5), 1258–1263. doi:10.1037/a0023923
- Marsh, R., Sebrechts, M., Hicks, J., & Landau, J. (1997). Processing strategies and secondary memory in very rapid forgetting. *Memory & Cognition*, 25(2), 173–181.
- Mazuryk, G. F., & Lockhart, R. S. (1974). Negative recency and levels of processing in free recall. *Canadian Journal of Psychology*, 23, 114–123.
- McCabe, D. P. (2008). The role of covert retrieval in working memory span tasks: Evidence from delayed recall tests. *Journal of Memory and Language*, 58, 480–494. doi:10.1016/j.jml.2007.04.004
- McElree, B. (2006). Accessing recent events. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 46, pp. 155–200). San Diego: Academic Press. doi:10.1016/S0079-7421(06)46005-9
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.
- Naime, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, 18, 251–269. doi:10.3758/BF03213879
- 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. doi:10.1037/0278-7393.28.3.411
- Oberauer, K., & Hein, L. (2012). Attention to information in working memory. *Current Directions in Psychological Science*, 21, 164–169.
- Oberauer, K. (2009). Design for a working memory. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 51, pp. 45–100). San Diego: Elsevier Academic Press. doi:10.1016/S0079-7421(09)51002-X
- Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58, 193–198.
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, 330, 335. doi:10.1126/science.1191465
- Raaijmakers, J. G. W. (1993). The story of the two-store model: Past criticisms, current status, and future directions. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 467–488). Cambridge: MIT Press.

- Ranganath, C., & Blumenfeld, R. S. (2005). Doubts about double dissociations between short- and long-term memory. *Trends in Cognitive Science*, 9, 374–380.
- Roediger, H. L., III, & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15, 20–27. doi:10.1016/j.tics.2010.09.003
- Rose, N. S. (2010). A processing approach to the working memory/long-term memory distinction: Evidence from a levels-of-processing span task. *Washington University Open Scholarship: Electronic Theses and Dissertations*, 300
- Rose, N. S. (2013). Individual differences in working memory, secondary memory, and fluid intelligence: Evidence from the levels-of-processing span task. *Canadian Journal of Experimental Psychology*, 67, 260–270.
- Rose, N. S., & Craik, F. I. M. (2012). A processing approach to the working memory/long-term memory distinction: Evidence from a levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(4), 1019–1029.
- Rose, N. S., Myerson, J., Roediger, H. L., III, & Hale, S. (2010). Similarities and differences between working memory and long-term memory: Evidence from the levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(2), 471–483.
- Rose, N. S., Olsen, R. K., Craik, F. I. M., & Rosenbaum, R. S. (2012). Working memory and amnesia: The role of stimulus novelty. *Neuropsychologia*, 50, 11–18.
- 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.
- Shivde, G., & Anderson, M. C. (2011). On the existence of semantic working memory: Evidence for direct semantic maintenance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1342–1370.
- Speer, N. K., Jacoby, L. L., & Braver, T. S. (2003). Strategy-dependent changes in memory: Effects on behavior and brain activity. *Cognitive, Affective, & Behavioral Neuroscience*, 3, 155–167. doi:10.3758/CABN.3.3.155
- Towse, J. N., Hitch, G. J., & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition*, 28, 341–348.
- Unsworth, N. (2010). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica*, 134, 16–28. doi:10.1016/j.actpsy.2009.11.010
- 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. doi:10.1037/0033-295X.114.1.104
- Unsworth, N., & Engle, R. W. (2008). Speed and accuracy of accessing information in working memory: An individual differences investigation of focus switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 616–630. doi:10.1037/0278-7393.34.3.616
- Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, 72, 89–104. doi:10.1037/h0021797
- Whitten, W. B., II, & Bjork, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning and Verbal Behavior*, 16, 465–478. doi:10.1016/S0022-5371(77)80040-6

#### Author note

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (Grant Nos. 8261 to F.I.M.C. and 386631 to B.R.B.). We thank Ashley Bondad for her help with data collection and scoring, and Nelson Cowan, Pierre Barrouillet, Klaus Oberauer, Tim Curran, Ian Dobbins, and seven anonymous reviewers for helpful comments on earlier versions of the manuscript.