Journal of Memory and Language 96 (2017) 155–167

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Time to process information in working memory improves episodic memory ☆

Alessandra S. Souza *, Klaus Oberauer

Department of Psychology, University of Zurich, Switzerland

ARTICLE INFO

Article history:
Received 16 December 2016
Revision received 19 June 2017

Keywords:
Complex span
Simple span
Working memory
Presentation rate
Focus of attention

ABSTRACT

In simple-span tasks, participants encode items sequentially for immediate serial recall. Complex-span tasks are similar, except that items are interleaved with a distraction task. Whereas immediate memory is higher in simple than complex span, in tests of episodic long-term memory, better recall for words studied in complex than simple span has been observed (McCabe, 2008). This *McCabe effect* has been explained by assuming that distraction displace items from working memory, forcing people to covertly retrieve items after each distraction, thereby generating better episodic retrieval-cues than during simple span. Our experiments support an alternative hypothesis: individual words are attended to and processed longer in working memory in complex-span than in simple-span trials. We reduced the presentation rate of words in simple span, creating a *"slow span"* condition. Across four experiments, slow span improved episodic memory compared to simple span, and this benefit was larger than the McCabe effect.

We often need to maintain information readily accessible in mind to perform ongoing cognitive tasks – e.g., remember a list of assignments for your boss that you received over the phone. The same information may become relevant again minutes or hours later. For instance, your boss may ask you later to pass the list to another colleague. The memory systems supporting performance in these two scenarios are working memory (WM) and episodic long-term memory, respectively.

WM is a limited capacity system which retains only a handful of representations available for ongoing processing (see Oberauer, Farrell, Jarrold, & Lewandowsky, 2016 for a recent review). This limited capacity constrains how well one can retain information to perform quotidian, yet complex tasks. Taking the example above, a very long list, or the occurrence of distractions (e.g., other phone calls), severely increases the chance of list items getting lost or being corrupted in memory.

Episodic long-term memory, in contrast, is not limited in capacity. We are constantly storing new events, and we can retrieve them over periods that vary between hours, days, or even years. This is not to say that episodic long-term memory does not fail us: retrieval from this system is slow and error-prone, and it is assumed to strongly depend on the ability of so-called *retrieval cues* to activate the appropriate memory trace (Craik & Tulving, 1975; Rugg & Wilding, 2000; Tulving, 1985). One question of interest in research on episodic memory is therefore which processes foster creation of effective retrieval cues.

In the present paper, our focus is on understanding how processing of information in WM (i.e., for an immediate task goal) affects the creation of episodic retrieval cues for recall over the long-term. One approach to investigating this question has been to compare episodic memory performance for information studied in the context of different WM tasks, such as simple span and complex span (Loaiza, Duperreault, Rhodes, & McCabe, 2014; Loaiza & McCabe, 2012a, 2012b; Loaiza, Rhodes, & Anglin, 2013; McCabe, 2008). In a typical simple-span task (aka word span), words are presented sequentially for study, and participants have to retain these words in their correct order of presentation for an immediate recall test. Complex-span tasks also require memory of a list in serial order, and in addition, in between presentation of the words participants have to complete a distractor task (e.g., judge the correctness of a multiplication equation; aka operation span) (Turner & Engle, 1989). It is well known that immediate recall is better for simple span than for complex span.

McCabe (2008) was the first to assess how the study opportunities offered in simple and complex span affected episodic memory. In his experiments, participants initially studied words in simple-span and complex-span trials for immediate recall. As typical, immediate recall was higher for simple span than complex span.

☆ The data and analysis scripts for all the experiments are available at the Open Science Framework at: https://osf.io/ctgr3/.

* Corresponding author at: Department of Psychology, Cognitive Psychology Unit, University of Zürich, Binzmühlestrasse 14/22, 8050 Zurich, Switzerland.

E-mail address: a.souza@psychologie.uzh.ch (A.S. Souza).

http://dx.doi.org/10.1016/j.jml.2017.07.002
0749-596X/

When participants were however confronted with a delayed memory test for all words they had encountered during the immediate-memory tests, the opposite pattern emerged: recall was higher for words studied in complex span than simple span. This finding was observed regardless of the delayed test being a surprise or fully anticipated. Hereafter, we will refer to this observation as the *McCabe effect*.

McCabe explained his findings with a covert retrieval model in which the distractor task is assumed to displace items from WM[1]. During the distraction period, the learned words are maintained only in episodic long-term memory. After each distraction, participants try to covertly retrieve the words back into WM. McCabe assumed that these covert retrieval attempts strengthen the retrieval cues associated with the item, and more so the more often those items were retrieved. He also argued that items in early list positions are retrieved more often because after each distraction participants retrieve items in forward serial order. Accordingly, serial position curves for the delayed recall showed a primacy gradient in complex span, but not in simple span.

Loaiza and McCabe (2012a) extended these findings by showing that complex span yields better episodic memory compared to simple-span trials of different lengths (4 or 8 words). Lists with 8 words (aka supra-span lists) exceed the presumed capacity of WM, hence immediate recall of these lists may also require the formation of effective retrieval cues. If participants could form these cues during the study phase, and they did so whenever the memory demand exceeded WM capacity, then supra-span lists, just like complex-span lists, should yield better delayed recall than short simple-span lists. This was however not the case: Only complex span yielded better delayed recall. This finding suggests that the distraction period taking place during the time in-between words in complex span is critical to yield better episodic memory.

Loaiza and McCabe (2012a) also provided evidence to support the covert retrieval model of the McCabe effect over an alternative temporal distinctiveness explanation (Brown, Neath, & Chater, 2007). In complex span, the words are presented in a temporally distributed fashion (i.e., separated from one another by the intervening distractor episodes). This may render the individual list words more temporally distinctive, thereby facilitating episodic retrieval. In contrast, the covert retrieval model assumed that the covert retrieval opportunities after the distraction task were the cause of the improved episodic memory. To test for these possibilities, Loaiza and McCabe (2012a) compared trials in which they varied the position of the distractor task (arithmetic problems): 4 problems followed by 4 memoranda, which is essentially a simple span task; 4 memoranda followed by 4 problems (i.e., a Brown-Peterson task); or memoranda and problems in alternation (complex-span trials). The latter two tasks require participants to retain information in mind over a period of distraction, which arguably requires the covert retrieval of the memoranda back into WM. At the same time, they differ in the temporal distribution of the memoranda: In the complex-span condition the memory items were temporally more separated than in the simple-span condition, but in the Brown-Peterson task they were not. Delayed recall was higher for both complex span and the Brown-Peterson task compared to simple span. These results favored the covert retrieval interpretation (see also Camos & Portrat, 2015; Loaiza & McCabe, 2012b; Loaiza et al., 2013 for further data consistent with this interpretation).

[1] McCabe (2008) framed his theory in terms of Cowan's embedded-processing theory of WM, and assumed that the distractor task displaced items from the "focus of attention" in that theory, referring to a capacity-limited device for holding up to about four items. Because the term "focus of attention" refers to different constructs in different theories, we use the more generic term "working memory".

One question that has not been addressed in these studies concerns the role of the distractor task in yielding better episodic memory. In the original model, McCabe (2008) assumed that it was the displacement from, and subsequent retrieval of information back into WM that yielded better episodic cues. An alternative explanation starts from the assumption that list items are not displaced from WM during distractor processing – although they are likely to suffer interference from the concurrent processing of distractor material (Oberauer et al., 2016) or they may suffer from time-based forgetting (Barrouillet & Camos, 2012). In this alternative view, the memoranda remain in WM throughout the trial. Given that complex-span trials take longer to complete than simple-span trials, the memoranda – in particular the words early in the list – are maintained in WM for a longer period of time in complex span than in simple span. If longer maintenance of information in WM is beneficial for episodic memory, then there would be more opportunities for encoding of information into episodic memory in complex span than simple span. If it is the amount of time words remain in WM that matters for the creation of strong episodic retrieval cues, then replacing the distractor processing interval by an equally long unfilled time interval after each word in simple span should also have a beneficial effect on episodic memory. Because this condition is effectively a simple-span task with reduced rate of presentation of the words, we refer to it as *slow span*.

We may distinguish between two possible ways in which maintenance time may be relevant. First, it may be that the total time words remain in WM is relevant irrespectively of whether there is distraction or not. If this is the case, complex-span trials and slow-span trials should yield comparable performance in a delayed memory test. Alternatively, it may be that what is important is the total amount of free time (i.e., time in which attention is not engaged in distracting activities) that matters. According to this hypothesis, focusing attention on information held in WM promotes the creation of strong episodic retrieval cues. There are several ways in which free time may improve episodic memory. One possibility is that participants use this free time to cycle their attention sequentially through all items stored in WM, thereby refreshing them (Barrouillet, Portrat, & Camos, 2011; Johnson, 2012; Souza, Rerko, & Oberauer, 2015; Vergauwe & Cowan, 2014). Refreshing is assumed to be a domain-general mechanism used for maintenance of all types of information in WM, which depends on the availability of central attention capacity (Barrouillet & Camos, 2012; Souza & Oberauer, 2017; Vergauwe, Barrouillet, & Camos, 2010). Loaiza and McCabe have suggested that the McCabe effect may reflect the use of attentional refreshing to strengthen episodic retrieval cues (Loaiza, Duperreault, Rhodes, & McCabe, 2015; Loaiza & McCabe, 2012a, 2012b; Loaiza et al., 2013).

An alternative possibility is that free time is used to consolidate only the just encoded item into WM (Bayliss, Bogdanovs, & Jarrold, 2015; Ricker, 2015; Ricker & Cowan, 2014; Schrijver & Barrouillet, 2017). According to this view, free time is used to strengthen only the last presented item. This stands in contrast with the refreshing hypothesis, which assumes that free time is used to cycle attention over all items in memory. A last possibility is that participants elaborate on the memoranda, thereby improving their episodic memory. If participants have free time, they might focus on "deep" aspects of the memoranda (e.g., the semantics of the words) which have been found to be beneficial to long-term memory (Craik, 2002; Craik & Tulving, 1975) and, to a smaller extent, WM (Loaiza, McCabe, Youngblood, Rose, & Myerson, 2011; Rose, Buchsbaum, & Craik, 2014). In sum, if free time is important for creating helpful episodic retrieval cues, then complex-span trials should yield better delayed recall than simple span, but still worse recall than a slow simple-span condition.

One last possibility is a combination of two alternative hypotheses discussed so far: Free time to attend to and process information is important, but also the practice with covertly retrieving information after distraction. In this case, a slow-span condition should yield better delayed recall than simple span, because it provides more free time, but still lower delayed recall than complex span, because it provides no practice with covert retrieval.

The goal of the present study was to examine the hypothesis that the amount of time information remains in WM is critical for promoting episodic memory. We tested this possibility by investigating the delayed retention of words studied in a condition in which the rate of presentation of words in simple span was reduced (by inserting a blank screen after each word) to match the inter-word spacing in complex span. We termed this condition "*slow span*". Our goal was to compare delayed recall from slow-span trials to recall from the regular simple-span trials, which served as our baseline. We also include complex-span trials in our experiments with the aims of (a) replicating the McCabe effect, and (b) comparing episodic memory for trials with uninterrupted maintenance (slow span) and maintenance interrupted by distractor processing (complex span). Given that the blank interval introduced in the slow-span condition does not displace items from WM, this manipulation allowed us to test whether it is covert retrieval after the loss of information from WM that produces strong episodic retrieval cues, or whether it is the time during which information is held in WM, and potentially further attended to (e.g., thereby refreshing, consolidation, or elaborating on the memoranda) that matters.

Experiments 1 – 3

The design of Experiments 1–3 was modelled after the one used by Loaiza and McCabe (2012a) in their Experiment 2. Participants were exposed to a first phase in which three types of trials were randomly intermixed: complex span, simple span, and slow span. In complex-span trials, each word was followed by the evaluation of a multiplication problem as the distractor task. Simple-span trials consisted of the presentation of the memoranda with no intervening activity at a relative fast pace, as in previous studies demonstrating the McCabe effect. In the slow-span trials, the presentation of each word was followed by an unfilled interval that took as long as the distractor task in the complex-span trials.

Each block of immediate-memory tests comprised three trials (one of each span type). This was followed by a brief filler activity lasting about 3 min which served to largely erase remaining memory traces from WM such that any attempt to recall the words later would have to rely on episodic memory. After the filler task, a free recall test of all the words studied in the previous set of three immediate-memory trials followed. There were six repetitions of this cycle (hereafter referred to as a block), across which the order of the span trials was fully counterbalanced. Using a similar procedure, Loaiza and McCabe (2012a, 2012b) reported that delayed recall of words studied in complex span was higher than in simple span. We expected to replicate this finding, and asked whether words from slow-span trials are also recalled better than words from simple-span trials. We were also interested in testing whether slow-span trials yield a smaller, similar, or larger delayed recall than complex-span trials.

The three experiments reported here vary in terms of the parameters of the simple-span and complex-span trials. We started off with a generic design which we thought would replicate the study of Loaiza and McCabe (2012a). Unexpectedly, we did not replicate the MCabe effect, that is, the better delayed recall of words studied in complex span than simple span. We then gradually attempted to match our simple-span and complex-span trials more closely to the ones of Loaiza and McCabe by varying task parameters in the subsequent experiments. Although these manipulations had a significant impact on immediate recall performance (with simple span becoming increasing better than complex span), they did not change the main pattern of results for delayed recall, namely that we failed to observe a McCabe effect. In contrast, inserting unfilled intervals between memoranda had a substantial and robust effect on delayed retention: across all three experiments, we consistently found better delayed recall for words from slow-span trials compared to words from simple-span and from complex-span trials.

General method

Participants

Seventy-two students from the University of Zurich participated in one 1-h session in exchange for 15 Swiss francs or course credit. All participants were native German speakers. They took part in one of three versions of the experiment (each with a sample of $n = 24$). Participants signed an informed consent form prior to the start of the experiment, and were debriefed in the end of the experimental session.

Materials and procedure

The stimuli in the main task consisted of a pool of 640 common German nouns[2] that were 4 to 5 letters long. Seventy-two words were randomly sampled from this pool without replacement for each participant for the use in the immediate recall tasks. In addition to the verbal memory span tasks, participants also completed a spatial memory test that served as a filler task in between the immediate and delayed memory tests. The experimental session was divided into six blocks, with each block consisting of a cycle through three immediate serial-recall trials, followed by a period of filler trials, and a delayed memory task. We describe the tasks below.

1. Immediate Serial Recall Task. Participants studied words for an immediate serial recall test under three different span conditions: complex span, simple span, and slow span. For all conditions, the trial started with the presentation of "*****" in the middle of the screen for 1 s. Thereafter, words (the memoranda) were presented in the middle of the screen one by one (see presentation duration in Table 1). In *complex-span* trials, presentation of two words was separate by an inter-stimulus interval (ISI) of several seconds (see Table 1), which was filled with the processing of a distractor task. In the distractor task, participants were required to judge the accuracy of a multiplication problem (e.g., $3 \times 7 = 20$?) that replaced the word in the center of the screen. The multiplications were randomly constructed with each operand sampled from a number between 3 and 9. Half of the equations had the correct result, and half the incorrect result (i.e., correct value ±1 or 2). The equation was presented for the duration of a fixed ISI (see Table 1), during which participants had to give their response. In Experiments 1 and 2, participants read the equation silently and pushed a button (left or right arrow key) to indicate that the displayed result was correct or incorrect, respectively. In Experiment 3, participants were instructed to read the equation aloud and to say aloud whether the result was correct or not.

In *simple-span* trials, words were separated by a short blank ISI. Lastly, in *slow-span* trials, words were separated by an unfilled ISI which was of the same duration as the interval for processing the distractor task in complex span trials. Table 1 presents the duration of the ISI implemented in each trial type across Experiments 1–3,

[2] For 597 words, we could obtain frequency per million estimates from the CLEARPOND database (Marian, Bartolotti, Chabal, & Shook, 2012). The average frequency was 65.5 (min = 0.75, max = 2772).

Table 1
Task Parameters Across Experiments 1 to 5.

Exp.	Condition	List-length	Word Dur. (s)	ISI (s)	Immediate Test	Distractor task
1	a. Complex span	5	1	4	Typed	Keypress
	b. Simple span			1		–
	c. Slow span			4		–
2	a. Complex span	4	0.9	3.5	Typed	Keypress
	b. Simple span			0.1		–
	c. Slow span			3.5		–
3	a. Complex span	4	0.9	3.5	Oral	Oral
	b. Simple span			0.1		–
	c. Slow span			3.5		–
Pilot	a1. Complex span	4	0.9	3.5	Oral	Oral
	b1. Simple span			0.1		–
	a2. Slow span	4	0.9	3.5	Oral	–
	b2. Simple span			0.1		–
4	a1. Complex span	4	0.9	3.5	Oral	Oral
	b1. Distractor first			3.5/0.1		Oral
	c1. Distractor last			0.1/3.5		Oral
	a2. Slow span	4	0.9	3.5	Oral	–
	b2. Blank first			3.5/0.1		–
	c2. Blank last			0.1/3.5		–
5	a. Complex span	4	0.9	3.5	Oral	Oral
	b. Simple span			0.1		–

Note. ISI = Inter-Stimulus Interval, Dur = Duration. In the Pilot Experiment and in Experiment 4, there were two types of blocks which consisted of the mixing of different conditions.

alongside all task parameters that differed between these experiments. In Experiment 1 participants studied lists of 5 words, whereas in Experiments 2–3, list length was always 4. Immediately after the ISI following the last word (with or without distractor task processing), there was a prompt for the immediate serial recall test. In Experiments 1 and 2, participants were asked to type the words using the computer keyboard. Each list serial-position was cued by a visual prompt (e.g., "Word 1:"). Participants were instructed that upper and lower case was irrelevant, and that they could correct their input using the backspace. When they were satisfied with the entered word, they were to push the Enter key to confirm their response, and move on to recalling the next word. In Experiment 3 participants were instructed to recall the items orally. If they forgot a word in a certain serial position they were to say "*blank*" and move on. The immediate recall test was signaled with the presentation of "?????" in the middle of the screen together with a beep. Participants were allowed 10 s to attempt recall of the words. In Experiment 3, the immediate recall test and the distractor task in complex span were recorded for offline accuracy check. Participants wore a headset throughout the experiment and were instructed to speak into the microphone.

Each experimental block consisted of three span trials, one of each type (i.e., complex span, simple span, and slow span). The order of the three trial types within a block was fully counterbalanced across the six experimental blocks.

2. Spatial Working Memory Test. After completion of the third span trial in a block, participants worked on a spatial working memory task for 21 trials (approx. 3 min). In the beginning of each trial, a grid with 30 randomly scattered black hollow dots were displayed against a white background for 500 ms. Next, six randomly chosen locations were filled with blue color for 1 s. Participants were instructed to memorize the blue dot locations. Next, all dot locations were masked for 200 ms, and the screen went blank for either 1 s or 3 s. Finally, the empty grid (in grey) was again displayed together with the mouse cursor. Participants had to click on the locations previously occupied by the blue dots. Clicking on the hollow dots turned them black. After selecting the six dots, correctly recalled dots turned blue and incorrectly recalled ones turned red for 1 s serving as accuracy feedback.

3. Delayed Free Recall Test. Next, participants were instructed to recall all of the words they have previously studied during the immediate recall phase of the current block. They were instructed that they should type as many words as they remembered and that recall order was not relevant. For this test, a table grid (in grey) with three columns and n rows, with n standing for the number of words in each list (e.g., 3×5 in Experiment 1; 3×4 in Experiments 2 and 3) was presented against a white background. The top, left-most cell was highlighted by showing it in black, and the first typed word appeared therein. When participants pressed the Enter key, the next cell in that column was highlighted, and the second typed word was presented therein, and so on until all cells were highlighted. When participants could not recall more words, they were instructed to simply press Enter to leave the subsequently highlighted cells empty. When the delayed free recall test was finished, participants were instructed that a new block would start with a new set of words for immediate recall.

Before the start of the experimenter proper, participants underwent a practice block with the distractor task used in complex span trials. They had to judge the accuracy of 30 equations within the same time interval used during the complex-span trials in each experiment. Participants were fully instructed prior to the start of the experiment about the three tasks (1. immediate serial recall, 2. spatial memory task, and 3. delayed recall) that would occur in the course in each block, and about the occurrence of multiple experimental blocks (i.e., multiple rounds of immediate and delayed tests).

Data analysis

We submitted our data to Bayesian t-tests (Rouder, Speckman, Sun, Morey, & Iverson, 2009) using the BayesFactor package (Morey & Rouder, 2014) implemented in R (R core team, 2014). This test computes the strength of the evidence for the presence or absence of an effect by comparing a model including the effect (H_1) to a model omitting it (H_0). The relative likelihood of the two models under comparison is the Bayes factor (BF). Interpretation of the BF is straightforward: it indicates the factor by which we should multiply our prior odds ratio for the two models to obtain the posterior odds ratio. The odds ratio is the ratio of our degree of belief that the data was generated from one model (e.g., H_1) relative to our degree of belief that it was generated from the other model (H_0). For instance, if we think that both models are equally probable before seeing the data (i.e., our prior odds ratio is 1), and the BF in favor of H_1 is 5, then our posterior odds ratio should be 5, meaning that we now regard H_1 as 5 times more probable than H_0. In general, BFs can be used as a continuous index of the relative strength of evidence for one model over an alternative model. The BF can express the evidence in favor of H_1 over H_0 (BF_{10}) or the evidence of H_0 over H_1 (BF_{01}). Here we will report BF_{10}. To arrive at BF_{01}, the reader only needs to compute $1/BF_{10}$. A BF_{10} smaller than 3 or larger than .33 is usually considered as "weak evidence" for H_1 or H_0, respectively; a BF_{10} larger than 10, or smaller than 0.1, is considered "strong" evidence.

In the present experiments, we were interested in the pairwise comparisons between conditions. For immediate recall we performed two-tailed t-tests for all pairwise comparisons between conditions. Our main interest lay, however, with the delayed recall data. Our first aim was to test whether we could replicate the McCabe effect, as reflected by better delayed recall of words studied in complex-span trials than words studied in simple-span trials. Our second aim was to assess for the contribution of the distraction period following each word in yielding better delayed recall. Hence our second test contrasted delayed recall of words studied in slow-span trials to recall of words from simple-span trials. These two hypotheses were directional (worse performance in simple span), hence we performed one-tailed t-tests. We also compared complex span to slow span in terms of delayed recall using a two-tailed test, because complex span could be more or less effective than slow span depending on the hypothesis under consideration. Furthermore, we also used the Bayesian estimation software (BEST) developed by Kruschke (2013) to get a Bayesian estimate of the effect size for the pairwise comparison of conditions, and of its 95% highest-density interval (HDI) which reflects the range of credible values of the effect size.

Results

Two participants in Experiment 1 were excluded because they did not perform the task as instructed (one participant never responded to the distractor task in complex span; the other participant always responded using the incorrect key and only wrote mocking words in the delayed test). One participant had to be excluded from Experiment 2 due to performance in the distractor task at chance level (50%). Two participants had to be excluded from Experiment 3 because of misunderstanding of the instructions for the oral recall task (e.g., attempted to recall the equations as well as the words). Another two participants were excluded because no immediate recall output was recorded, and two additional participants were excluded for failing to read/respond to the equations in a substantial proportion of trials (E3 final sample n = 18). Inclusion of the delayed recall data of those participants does not change the overall pattern of results. Descriptive statistics of performance of the distractor task in complex span and the filler spatial WM task are provided in the Online Supplementary Materials.

Immediate recall

We computed a serial recall score (proportion of words recalled in the correct list position) and a free recall score (proportion of list words that were recalled irrespectively of position) for complex-span, simple-span, and slow-span trials in Experiments 1–3 (see Fig. 1a). Table 2 shows the evidence for the pairwise comparison of conditions. We ran separate analyses having as dependent measure the serial recall score and the free recall score.

The three experiments differ in the ranking of conditions at immediate recall. In Experiment 1, complex span and simple span yielded similar performance, whereas slow span recall was better than simple span. The similar level of performance yielded by complex and simple span was unexpected given that processing of a distractor task usually impairs immediate memory. We reasoned that this may have been due to participants having too much time to process the equations in complex span trials or due to the use of rehearsal. Hence in Experiments 2 and 3 we modified the procedure to be more comparable to that of Loaiza and McCabe (2012a). In Experiment 2, we made the presentation rate in simple span faster, and also slightly reduced the time to process the equations in complex span. Both of these changes were made to more fully match the timing used by Loaiza and McCabe (2012a). This translated into better immediate serial recall performance in simple span trials compared to complex span. In Experiment 3, we increased the degree of interference by the distractor task in complex span by requiring participants to read the equations aloud and provide their decision aloud. This limits the use of articulatory rehearsal in complex span, which may be an inefficient way to create strong episodic retrieval cues. Moreover, as in the experiments of Loaiza and McCabe, in Experiment 3, immediate recall was performed orally. The change in the processing task in complex span yielded a massive reduction in immediate recall. In both Experiments 2 and 3, simple span and slow span yielded similar levels of performance. For free recall scoring (see Table 2), the results tended to mirror the ones obtained for serial recall scoring.

Delayed free recall

Fig.1b shows the proportion of words recalled in the delayed test that were studied in each span condition. We again compared the conditions in a pairwise fashion using Bayesian t-tests (see Table 2). These analyses revealed a similar pattern across all three experiments: For the comparison of complex span and simple span, the evidence favored the Null hypothesis, whereas for the comparison of slow span and simple span, the evidence favored the alternative hypothesis. There was also evidence in the direction that delayed recall was larger for words studied in slow-span trials than words studied in complex-span trials.

We also computed an overall analysis across all experiments in which we entered two conditions (complex span vs. simple span; or slow span vs. simple span) and experiment as factors in a Bayesian Analysis of Variance, ANOVA (Rouder, Morey, Speckman, & Province, 2012) using the BayesFactor package with its default settings and the model comparisons set to "top". This setting provides the comparison of a full model including all predictors of interest with models omitting one of the terms at a time (reduced model). The BF for the reduced model over the full model indicates whether removing this term improved the model ability to account for the data. For the model comparing complex span to simple span, the evidence was against including a condition x experiment interaction ($BF_{10} = 0.14$), against including a main effect of experiment ($BF_{10} = 0.28$), and against including an effect of span condition ($BF_{10} = 0.29$). Hence none of the predictors explained variance in delayed recall performance when contrasting these two conditions. For the model comparing slow span to simple span, the evidence was against including the condition x experiment interaction ($BF_{10} = 0.13$) and against including the main effect of experiment ($BF_{10} = 0.29$). The main effect of condition was however strongly supported ($BF_{10} = 9.6 \times 10^6$), showing that slow span substantially improved delayed recall compared to simple span.

Discussion

Our first three experiments showed little evidence for a McCabe effect, despite our efforts to make the implementation of the simple span and complex span trials similar to the one used by Loaiza and McCabe (2012a). The lack of a difference between delayed recall for these two span conditions occurred even when their

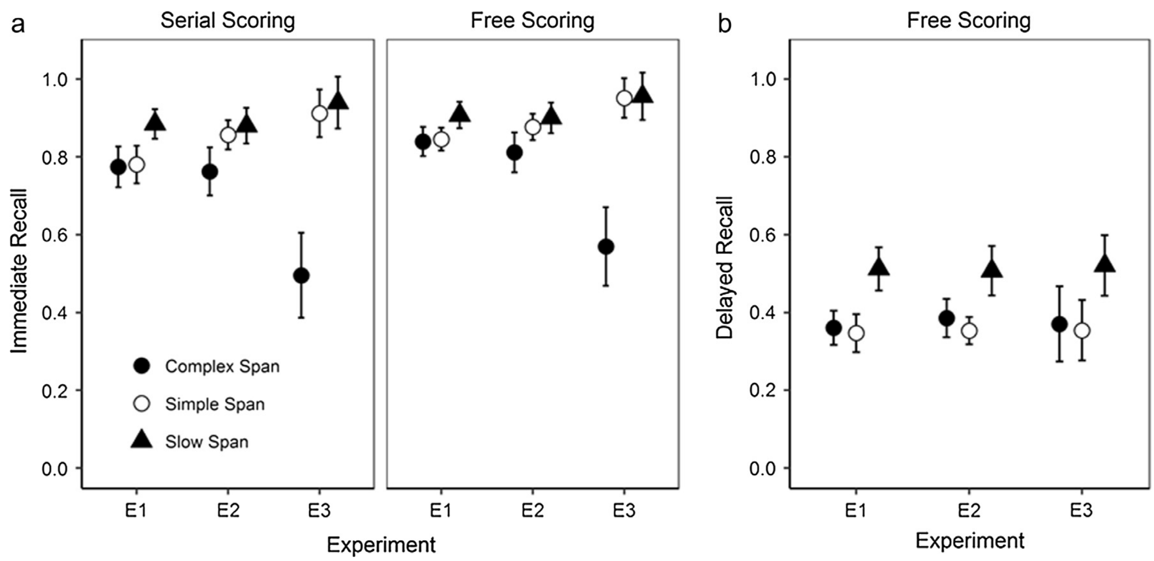

Fig. 1. Data of Experiments 1–3. (a) Proportion of words recalled in the correct position (serial scoring) and regardless of list position (free scoring) in the immediate serial-recall test. (b) Proportion of words recalled in the delayed free-recall test that were studied in each span task. Error-bars show 95% within-subjects confidence intervals.

Table 2
Bayes Factor (BF for the Alternative Hypothesis over the Null Hypothesis), Bayesian Effect Size (d), and Highest-Density Interval (HDI) of the Effect size for the Pairwise Comparison Between Span Conditions in Experiments 1–3 in terms of Immediate and Delayed Recall Performance.

Score	*Comparison*	E1			E2			E3		
		BF_{10}	d	95% HDI	BF_{10}	d	95% HDI	BF_{10}	d	95% HDI
Immediate Test										
Serial	*Complex vs. Simple*	0.22	−0.05	[−0.55, .43]	3.35	−0.50	[−0.99,−.02]	4349.9	−1.51	[−2.26, −.78]
	Slow vs. Simple	38.16	0.80	[0.30, 1.31]	0.36	0.01	[−0.33, 0.36]	0.40	0.24	[−0.26, 0.74]
	Complex vs. Slow	24.49	0.75	[0.26, 1.26]	6.13	0.59	[0.12, 1.05]	5874.6	1.54	[0.81, 2.29]
Free	*Complex vs. Simple*	0.23	−0.01	[−0.47, .44]	1.48	−0.39	[−0.86, .09]	5095.5	−1.54	[−2.31, −.80]
	Slow vs. Simple	8.09	0.66	[0.17, 1.16]	0.39	0.21	[−0.23, 0.65]	0.25	0.00	[−0.32, 0.33]
	Complex vs. Slow	3.63	0.53	[0.07, 1.00]	3.53	0.52	[0.07, 0.99]	2709.1	1.44	[0.73, 2.15]
Delayed Test										
Free	*Complex vs. Simple*[a]	0.26	0.01	[−0.44, .46]	0.40	0.13	[−0.30, .57]	0.30	0.10	[−0.40, 0.61]
	Slow vs. Simple[a]	4.52	0.52	[0.02, 1.00]	224.34	0.91	[0.39, 1.44]	32.8	0.79	[0.24, 1.35]
	Complex vs. Slow	125.7	0.92	[0.39, 1.45]	1.57	0.50	[0.10, 0.99]	2.53	0.57	[0.04, 1.10]

[a] One-tailed *t*-test.

performance differed in the expected direction in immediate recall (Exp. 2 and 3). In contrast, slow span consistently improved recall across all three experiments, with effect sizes that ranging from medium (0.52) to large (0.91; see Table 2).

So far, our data shows that prolonging the time for which items are maintained in WM improves episodic memory. Covert retrieval of words from episodic memory, assumed to occur in complex span after processing the distractor task, did not translate into better delayed recall in our experiments. One possible reason why we have not been able to reproduce the McCabe effect may be related to the inclusion of the slow-span trials. We have inter-mixed the presentation of all three span trials in each experimental block. Following the brief filler task, participants attempted recall of all words studied in that block. Given that slow-span trials improved memory substantially, and words from different span trials competed for recall at the delayed test, recall of words studied in complex span and simple span may have suffered from the strong competition for recall by the slow-span words.

To assess the viability of this interpretation, we ran an additional experiment in which we created two types of experimental blocks: one block mixing simple-span and complex-span trials, and another block mixing simple-span and slow-span trials. This way, participants never had to recall, at the same time, words studied in complex span and slow span. Participants completed 8 blocks of trials (4 blocks of each type). Each block consisted of 6 trials (3 of each type of span). The general task design of this experiment was the same as for Experiment 3: participants read and judged the operations in complex span aloud, immediate recall was performed orally, and the delayed test involved typing the remembered words in any order using the computer keyboard (see Pilot experiment in Table 1). We ran 22 students in this study. Unfortunately, due to a malfunctioning apparatus, we lost the data of the immediate study phase for half of the participants (processing of the equations in complex span and immediate recall). Given that we had no means of checking that complex span trials were indeed processed as they were supposed to, the results of the delayed test provided only tentative evidence about the viability of assessing the McCabe effect and the slow-span effect across different blocks of trials. Delayed recall of words studied in simple-span trials did not differ between blocks types (block with complex span, $M = 23.7\%$; block with slow span, $M = 24.1\%$; $BF_{10} = 0.22$). The evidence for a McCabe effect was ambiguous (complex span, $M = 31.9\%$; $BF_{10} = 1.8$). There was overwhelming evidence for a slow-span benefit (slow span, M = 47%, $BF_{10} = 12{,}620$), and for a difference between slow span and complex span ($BF_{10} = 36.9$).

Although this experiment did not provide unambiguous evidence for a McCabe effect, it was the first time that we actually saw some evidence favoring better delayed recall in complex span compared to simple span. This encouraged us to set up our next experiment with two types of blocks: one block type for assessing evidence for the McCabe effect, and another block type for assessing evidence for the slow-span effect. In addition, in our next experiment we decided to revisit the evidence against the temporal-distinctiveness explanation of the McCabe effect (Loaiza & McCabe, 2012a), and consequently, the viability of this explanation for the slow-span effect.

Experiment 4

In this experiment we implemented a design similar to the one reported by Loaiza and McCabe (2012a, Experiment 3) in which they varied the position of the distractor task across three span tasks. Fig. 2 presents the general design of the conditions. As in the previous experiments, complex span trials involved the interleaved presentation of the memoranda (words) and distraction task (multiplication problems). In contrast, the *Distraction First* condition started with the presentation of the 4 multiplication problems, followed by four memoranda. On the assumption that the preceding multiplication task has little impact on the subsequent immediate-memory task, this condition is approximately equivalent to the simple span trials in the preceding experiments, and therefore we use it as the baseline condition. The *Distraction Last* condition started with the uninterrupted presentation of the four memoranda followed by the four problems (aka Brown-Peterson task). The distraction-first and distraction-last conditions have temporally crowded presentation of the memoranda, whereas the complex-span condition presents the memoranda spread out in time. If delayed recall is improved by temporal distinctiveness, it should be better for words from complex span than for words from the other two conditions. A consolidation account could also explain better delayed recall in complex span: as long as participants take some of the time available for processing the distractor to continue consolidating the last presented item in WM (Ricker, 2015), they should show better memory for words studied in complex span than in the distraction-first and distraction-last conditions. The distractor-first and the distractor-last conditions differ in the opportunities for covert retrieval of the memoranda. Only in the distraction-last condition, words are displaced from WM during distractor processing, so that, according to McCabe (2008), they need to be covertly retrieved from episodic memory for successful immediate recall. Hence, if delayed recall is improved by practice with covert retrieval, delayed recall should be better for the distraction-last condition – and the complex-span condition – than for the distraction-first condition. These three trial types were implemented by Loaiza and McCabe (2012a) concurrently, and they observed that delayed recall was as good in complex span as in the distraction-last condition, with both being better than the distraction-first condition.

Here we were also interested in testing for the possibility that unfilled ISIs between the memoranda, or an unfilled retention interval following presentation of all memoranda, would improve delayed recall. Therefore, we implemented similar control conditions for the slow-span trials: a *Blank First* condition in which a long blank period was inserted prior to the presentation of the memoranda (which are essentially a form of simple span), and a *Blank Last* condition in which the long blank period followed the memoranda. A temporal-distinctiveness account predicts a beneficial effect of unfilled ISIs separating individual words, but not of an unfilled RI following presentation of the entire list. A consolidation account makes the same prediction. The temporal distinctiveness and the consolidation account differ in that the latter also predicts that slow-span trials should yield better episodic memory than complex-span trials. This is because in the slow span there is more free time to consolidate the last presented item than in complex span. The covert-retrieval account predicts no difference between all three "blank" conditions for delayed recall, because none of them requires covert retrieval during the immediate-memory trials. Finally, if the time for which words are held in WM determines delayed recall success, delayed recall should be best in the blank-last condition, because all list words are maintained in WM for an extended duration, providing ample time for refreshing or elaboration. Delayed recall should be intermediate for the slow-span condition, in which the earlier words are maintained in WM about as long as in the blank-last condition, but later list words are maintained more briefly. Delayed recall should be worst for the blank-first condition, in which all words are held in WM only briefly.

Method

Participants

We set up to collect the data of a minimum of 30 participants. After collecting this minimum sample size, we started monitoring the BF_{10} for the comparison of the delayed recall of words studied in complex span vs. words studied in the distraction-first condition (i.e., our baseline condition akin to simple span). Our aim was to collect data until we reached a $BF_{10} = 0.1$ or $BF_{10} = 10$, which indicates strong evidence for the Null or the Alternative hypothesis, respectively, or until we collected data of 60 participants. We monitored the BF_{10} daily after data collection of the participants that signed up for the experiment on that day.[3]

Given that the BF never reached the criterion, we ended up collecting data of 61 participants. We collected 61 data sets because we had to replace a participant that did not complete the complex span trials as instructed (this person did not read and evaluate the distractor problems). After exclusion of this participant, we ended up with a sample of $n = 60$.

Materials and procedure

In Experiment 4, a single subset of the words used in the previous experiments was selected for use for all participants (set = 72 words). As before, the words were 5–6 letters long, and they were selected to have minimal overlap with each other (no two words with the same first two letters). Words were randomly distributed across the different trial types implemented in the experiment.

As shown in Fig. 2, the present experiment comprised the mixture of two block types: one block mixing complex-span, distraction-first, and distraction-last conditions (hereafter Distraction block), and another block mixing slow-span, blank-first, and blank-last conditions (No Distraction block). In every block, there was a single trial of each span task (i.e., 3 trials per block). There were 6 blocks in total, half of each type. The block types strictly alternated, and the order of the blocks was counterbalanced across participants.

Similar to the previous experiments, participants completed a spatial working memory task as a filler activity in between the immediate and delayed recall tests. In this task, four colored dots were presented in random locations on the screen. The task was to remember the precise location of each dot over a brief retention interval (1 or 3 s) for a cued recall task. Participants completed 10 trials of this task in each block.

The delayed recall test was similar to the one implemented in Experiments 1–3: participants were instructed to type the words

[3] It was impractical to check the BF after collecting data of each individual participant because this would require scheduling participants in a very sporadic fashion, thereby slowing data collection.

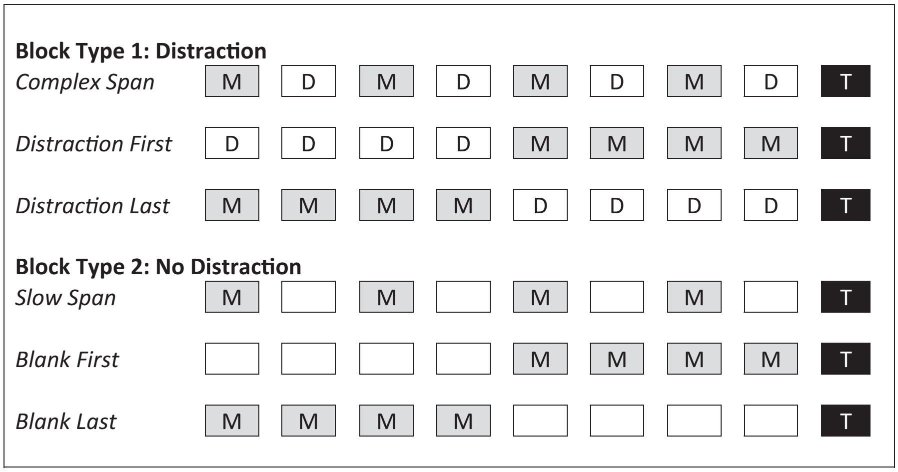

Fig. 2. Depiction of the span tasks implemented in Experiment 4 across the two block-types. Block type 1 replicates the conditions used by Loaiza and McCabe (2012a, Experiment 3). Block type 2 shows the new conditions included in present experiment to test for a temporal distinctiveness explanation of the slow-span effect. D = distraction task; M = memoranda. T = immediate serial recall test. Empty cells indicate that a blank screen was presented for the same duration as the time allowed for processing the distractor task.

they remembered having learned during the current block. The currently typed word appeared in a box in the top half of the screen. When participants were sure about the inserted word, they pressed the Enter key. The word then moved from the top box to a list in the bottom half of the screen (all recalled words remained visible throughout the test). When participants could not recall any more words, they were instructed to type “end” followed by an Enter-key press to finish the test.

Participants were not informed ahead about the delayed memory test, hence this test was a surprise in the first block. With the repetition of the blocks, however, the sequence of the phases (immediate test, spatial filler task, and delayed test) was predictable. In previous research, instructing participants or not about the delayed memory test did not influence the observation of a McCabe effect (McCabe, 2008).

Results

Accuracy in the performance of the distractor task in complex span was relatively high (above 80%). Descriptive of the performance in equation-verification task as well as in spatial filler task can be found in the Online Supplementary Materials.

Immediate recall

Fig.3a presents performance across the six types of span tasks in terms of serial and free-recall scoring. Recall was high and close to ceiling for the distraction-first condition (which is akin to simple span), and for all span tasks in the no-distraction block (blank span, blank first, and blank last) irrespectively of the type of scoring. Immediate recall was much reduced for span tasks requiring maintenance and concurrent processing, namely complex span and the distraction-last condition. Table 3 presents the evidence for the comparison between conditions that were presented within the same block. The pairwise comparison of conditions in the Distraction block support the conclusion that performance was best for the distraction-first condition, followed by complex span, with the worst recall being observed in the distraction-last condition. For the No-Distraction block, there was no evidence for differences between conditions, with the BF being either ambiguous or strongly favouring the Null hypothesis.

Delayed free recall

Fig.3b presents the data of the delayed recall test. There was a small advantage in delayed recall for words studied in complex span compared to the distraction-first condition. The distraction-last condition yielded somewhat lower recall than the distraction-first condition. Table 3 shows the evidence for the comparison of these two conditions against the distraction-first condition (akin to simple span). There was some evidence for better delayed recall of words studied in complex span, in line with a McCabe effect. There was, however, strong evidence against an advantage in recalling words studied in the distraction-last condition (aka Brown-Peterson task). Actually, delayed recall tended to be worse in this condition than in the distraction-first condition. This finding is the opposite of the one reported by Loaiza and McCabe (2012a).

There was a substantial advantage for delayed recall of words studied in slow-span compared to the blank-first condition, as shown in Table 3. As for the previous experiments reported herein, the effect size for this comparison was medium, and about twice as large as the effect size for the McCabe effect we observed in Experiment 4. Unlike the distraction-last condition, the blank-last condition tended to yield a slight advantage over the blank-first condition. The evidence was, however, ambiguous for this benefit.

In short, we found some evidence for a McCabe effect. However, this evidence was underwhelming even after collecting data of 60 participants - which is about twice as many participants as in most of the experiments reported by McCabe and colleagues (Loaiza & McCabe, 2012a; McCabe, 2008). Replicating Experiments 1–3, we found evidence for a slow-span benefit.

Discussion

Experiment 4 suggests that one determinant for our failure to replicate the McCabe effect in Experiments 1–3 was the mixing of the complex-span trials with the slow-span trials. When we separated these span tasks across different blocks, we found some evidence for a McCabe effect (pilot study and Experiment 4). The McCabe effect was, however, of relative small size (d = 0.33) compared to the medium-size (d = 0.62) and statistically strongly supported benefit for the slow-span trials.

Our results do not support the contention that during the distraction-last condition (Brown-Peterson task), participants engage in covert retrieval of the memoranda after the distraction episodes, leading to improved delayed recall. In our study, delayed recall was actually worse for this condition compared to the baseline. This results clashes with the one reported by Loaiza and McCabe (2012a). We have no ready explanation for the failure to

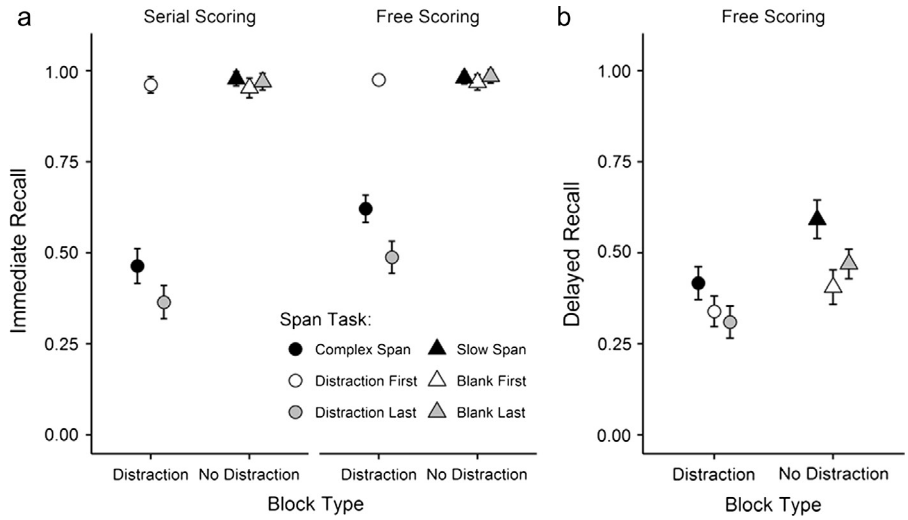

Fig. 3. Panel (a) proportion of list-words recalled in the correct list position (serial scoring) and proportion of list-words recalled regardless of list position (free scoring) in the immediate recall test of Experiment 4. Panel (b) proportion of words recalled in the delayed free recall test that were studied in each experimental condition in Experiment 4. Error bars show 95% within-subjects confidence intervals.

Table 3
Bayes Factor (BF for the Alternative Hypothesis over the Null Hypothesis), Bayesian Effect Size (d), and Highest-Density Interval (HDI) of the Effect size for the Pairwise Comparison Between Span Conditions in each Block Type in Experiment 4.

Score	Comparison	BF_{10}	d	95% HDI
Immediate Test				
Serial	Complex Span vs. Distraction First	2.6×10^{21}	−2.22	[−2.7, −1.7]
	Distraction Last vs. Distraction First	6.6×10^{25}	−2.90	[−3.7, −2.2]
	Complex Span vs. Distraction Last	21.9	0.46	[0.17, 0.74]
	Slow Span vs. Blank First	0.88	0.0	[−.14, 0.14]
	Blank Last vs. Blank First	0.27	0.0	[−0.15, 0.15]
	Slow Span vs. Blank Last	0.19	0.0	[−0.14, 0.14]
Free	Complex Span vs. Distraction First	2.9×10^{21}	−2.27	[−2.8, −1.7]
	Distraction Last vs. Distraction First	9×10^{23}	−2.55	[−3.1, −2.0]
	Complex Span vs. Distraction Last	1819.8	0.78	[0.40, 1.2]
	Slow Span vs. Blank First	0.31	0.0	[−0.14, 0.14]
	Blank Last vs. Blank First	0.54	0.0	[−0.14, 0.14]
	Slow Span vs. Blank Last	0.17	0.0	[−0.14, 0.14]
Delayed Test				
Free	Complex Span vs. Distraction First[a]	3.61	0.32	[0.05, 0.60]
	Distraction Last vs. Distraction First[a]	0.08	−0.12	[−0.38, 0.15]
	Complex Span vs. Distraction Last	62.6	0.49	[0.21, 0.77]
	Slow Span vs. Blank First[a]	2566.1	0.62	[0.33, 0.92]
	Blank Last vs. Blank First[a]	2.0	0.26	[−0.01, 0.52]
	Slow Span vs. Blank Last	48.3	0.51	[0.21, 0.80]

[a] One-tailed tests.

replicate their findings. Our finding cannot be interpreted as resulting from the mixing of complex-span trials and distraction-last trials, because this is exactly the same mixture as used by Loaiza and McCabe (2012a, Experiment 3). The delayed-recall benefit for words from a distraction-last condition has been reported only once. Our failure to replicate it suggests that it may be less robust than the McCabe effect.

Our results also refute the covert-retrieval model proposed by McCabe (2008): interspersing periods of distraction during the retention period are not necessary for yielding better delayed recall. If anything, omitting the distraction (while keeping the temporal separation of the memoranda) has a larger beneficial effect on episodic memory.

Taken together, the results of Experiment 4 are best explained by two variables that improve episodic memory: the temporal separation between items, and the duration for which each individual item is maintained and can be processed in WM (free time effect). Temporal distinctiveness could explain why delayed recall was better in complex span and slow span compared to their corresponding counterparts with uninterrupted presentation of the words. However, this account does not take into consideration the fact that slow span yielded better delayed recall than complex span, which indicates that the amount of free time also played a role. A consolidation account could explain both effects: participants needed free time after each individual item to consolidate this information in episodic memory. Given that consolidation only affects the last presented item and that it is an attentional demanding process, it can explain both the temporal separation effect and the free time effect. Alternatively, one could conceive that other attentional process (refreshing or elaboration) in combination with temporal distinctiveness could explain our data, but at present those accounts seems less parsimonious.

Experiment 5

The goal of our last experiment was to assess whether the McCabe effect could be replicated more robustly when only simple-span and complex-span trials are included in the experiment. Experiment 4 showed evidence for a McCabe effect when slow-span and complex-span trials were presented in different blocks of trials. However, one may wonder whether the somewhat weak evidence for the McCabe effect we obtained in Experiment 4 may be due to the presence of slow-span trials in the same experiment or due to some sort of particularity of our experimental design. Hence the goal of Experiment 5 was to provide a more direct replication of the McCabe effect using the same basic experimental set-up as in Experiment 4 (i.e., the same pool of words; the same spatial filler task; the same immediate and delayed recall set-up). If we can more robustly replicate the McCabe effect when we remove the slow-span trials from the picture, this would provide strong support for the contention that words from slow-span trials competed for recall with words studied in complex span, hence limiting the observation of the McCabe effect.

Method

Participants

Thirty-one students took part in Experiment 5 (mean age = 24 - years; 6 men). Three participants had to be excluded from the experiment because they were not reading the equations in complex span, leaving a final sample of n = 28.

Materials and procedure

Participants completed four experimental blocks consisting of an immediate serial recall task phase, the filler spatial WM task, and the free delayed-recall test. The tasks were the same as in Experiment 4 with one exception: the immediate serial recall phase comprised 2 simple-span trials and 2 complex-span trials presented in alternated fashion (with the order of alternation of the trials counterbalanced across blocks and participants).

Results

Descriptive of performance in the equation-verification task and spatial WM filler task can be found in the Online Supplementary Materials.

As shown in Fig. 4a, immediate recall was higher in simple span compared to complex span both in terms of serial scoring ($BF_{10} = 3.5 \times 10^9$; d = −2.49, 95% HDI [−3.55, −1.56]) and free scoring ($BF_{10} = 6.4 \times 10^8$; d = −2.15, 95% HDI [−2.94, −1.38]). Performance levels across these two types of span tasks were similar to the ones obtained in Experiments 3 and 4. In contrast to immediate recall, delayed recall (see Fig. 4b) was larger for words studied in complex-span trials than simple-span trials ($BF_{10} = 881.6$; d = 1.06, 95% HDI [0.47, 1.67]), replicating the McCabe effect.

Discussion

Our final experiment showed that the McCabe effect can be more robustly observed when only simple-span and complex-span trials are intermixed. This unambiguously indicates that the reason for our previous failure in replicating the McCabe effect was indeed related to the inclusion of the slow-span trials in the same experimental set-up. Overall our data suggests that slow span improves episodic memory so substantially that it interferes with the observation of a McCabe effect.

General discussion

Which conditions promote the formation of strong episodic retrieval cues?

The present study investigated how the maintenance of information in WM for an immediate serial recall task impacts the formation of episodic long-term traces. The motivation of this study was to test whether interspersing periods of distraction – assumed to disrupt active maintenance of the memoranda in WM –, foster the creation of strong episodic retrieval cues, as suggested by the McCabe effect. This is the assumption at the core of the covert-retrieval model proposed by McCabe, Loaiza, and colleagues (Loaiza & McCabe, 2012a; Loaiza et al., 2013, 2014, 2015; McCabe, 2008). As an alternative explanation, we reasoned that maintenance in WM (irrespectively of distraction) may be the critical variable in yielding strong episodic retrieval cues. This account could explain the McCabe effect as follows: in complex span trials participants use some of the time allowed for processing of the distractor to attend to and further process the memoranda. If they do so, then the amount of time words can be processed in WM is longer in complex span than in simple span, resulting in better encoding into episodic memory. To assess this possibility, we created a slow-span condition in which an unfilled interval was inserted in between presentation of every two words, nominally reducing the presentation rate of the memoranda (a condition that we called slow-span). Across four experiments, we observed a slow-span benefit for delayed recall. When assessed in the same experimental set-up (see Experiment 4), the slow-span benefit was about two times larger than the McCabe effect.

Our findings challenge one of the main assumptions of the covert-retrieval model (McCabe, 2008), namely that periods of distraction promote the displacement from and covert retrieval of items into WM, thereby providing retrieval practice that fostered the creation of strong episodic-retrieval cues. Our results point to an alternative interpretation: Free time to attend and further process information in WM is key in promoting episodic memory.

Could our results be explained as an effect of temporal distinctiveness?

We only observed benefits for delayed recall when memoranda were presented in a temporally spaced fashion (complex span and slow span), but not when the memoranda were presented in a temporally crowded fashion (distraction-last or blank-last conditions). The fact that better delayed recall required spaced presentation is in line with a temporal distinctiveness explanation (Brown et al., 2007; Unsworth, Heitz, & Parks, 2008). We note, however, that the temporal separation of the memoranda was the same between the complex-span trials and the slow-span trials, yet delayed recall was larger for the latter than the former. Hence even if temporal distinctiveness contributes to episodic memory, it remains to be explained why the slow-span trials led to superior delayed recall. Our argument is that focusing attention on the memoranda is necessary to consolidate/refresh/elaborate on this information, and slow-span trials offer a larger opportunity for these processes to take place. This finding also dovetails with a recent study by Camos and Portrat (2015) showing that the amount of free time in between processing of two words – varied by including distractor tasks of high or low attentional demand – impacts delayed recall, whereas opportunities for articulatory rehearsal do not. Together these findings support the contention that participants need time to focus attention on information (one by one) in WM in order to create efficient cues for retrieval over the long-term.

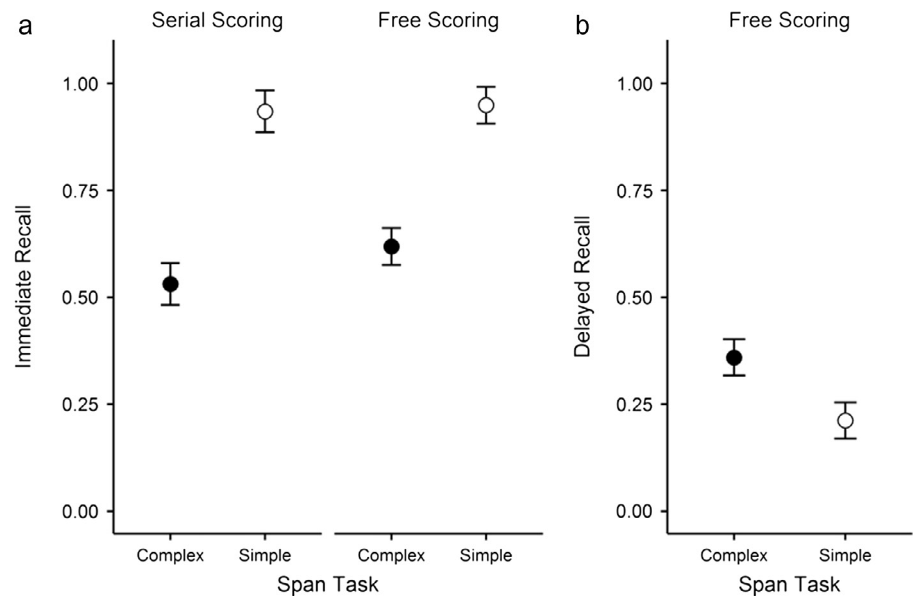

Fig. 4. Proportion of recalled words from complex-span and simple-span trials in the (a) immediate memory test and (b) delayed memory test. Error-bars depict 95% within-subjects confidence intervals.

What is free time being used for?

Our data shows that participants need free time to attend to each presented item in order to create strong episodic retrieval cues. What do people do in this free time to improve episodic memory?

Our findings can be easily accommodated by a consolidation account: Time after each individual word is used to consolidate the last present item into WM and episodic memory (Ricker, 2015). In contrast, a refreshing account of our data is less compelling. Refreshing is assumed to cycle across all presented items, and it is only prevented when participants engage in an attention-demanding task. Hence, refreshing is assumed to take place during an unfilled retention interval following presentation of all memoranda, and there is evidence supporting that assumption (Souza & Oberauer, 2017; Vergauwe, Camos, & Barrouillet, 2014; Vergauwe & Langerock, 2017). Therefore, we should expect that participants attempted refreshing of all words in the distraction-last and blank-last conditions in Experiment 4. If refreshing helps to establish stronger episodic retrieval cues, we would expect a better delayed recall in these conditions, which we however did not observe.

There is previous evidence consistent with a role of refreshing in improving episodic memory. M.K. Johnson and colleagues reported that refreshing of a just-presented item yielded better long-term memory compared to the re-presentation of the item for reading (Grillon, Johnson, Krebs, & Huron, 2008; Johnson, Reeder, Raye, & Mitchell, 2002; Johnson et al., 2013). We note however that Johnson and colleagues have a definition of refreshing that is quite close to the one of consolidation: namely, that refreshing prolongs the activation of a just-seen item. Under this definition, refreshing and consolidation become very similar and difficult to tease apart. Hence it may well be that the episodic memory benefits observed for refreshing in their experiments could be, at least partially, due to consolidation. Further studies will be needed to assess whether refreshing multiple items can lead to better episodic memory, and to better delineate the boundaries between consolidation and refreshing.

A third possibility is that elaboration contributes to the effects we observed. Loaiza and collaborators have observed that the McCabe effect only occurs for words already established in LTM, but not for novel words or nonwords (Loaiza et al., 2013, 2015). This leaves open the possibility that elaborative strategies are contributing to the better delayed recall we observed in complex-span and slow-span trials. Although viable, we have no evidence that participants attempted to elaborate on the presented items.

Altogether, our data seems more consistent with a consolidation account, but it does not provide evidence against the viability of elaboration as contributing to free time benefit we observed. We note here some possible routes to empirically distinguish between the roles of consolidation, refreshing, and elaboration in future research. First, we have not attempted to control the strategies participants may apply during the free-time periods. One promising way to do so is to provide strategy instructions. For example, participants may be instructed to focus only on the last presented item (consolidation), or they may be asked to think of all memory items presented so far during the free-time (refreshing), or alternatively asked to think of the semantic aspects of the memoranda (elaboration). One could also vary characteristics of the memoranda that may facilitate or hinder the use of some of those strategies, such as the use of abstract versus concrete words or the use of words vs. non-words to constrain elaboration, and also the type of domain (verbal vs. visual) for testing for domain-general processes such as consolidation and refreshing. Another possibility may be related to the manipulation of the locus of free-time: before or after processing the distractors. Arguably, free time before processing the distractors could be used both for consolidation and refreshing, but free time after processing the distractor could only be used for refreshing. Using this logic, Bayliss et al. (2015) have shown that free time before processing the distractors (aka consolidation time) was more beneficial to immediate recall than time after the distractors (aka refreshing time). More recently, De Schrijver and Barrouillet (2017) have extended this finding indicating that consolidation time and refreshing time produced additive effects that linearly predicted immediate serial recall. It remains an open question whether the same relationship holds for episodic memory.

Finally, it is worth considering the implications of our results for two models of how free time during complex span tasks is used: In the Time-Based Resource-Sharing theory (TBRS, Barrouillet &

Table 4
Proportion of Recalled Words in the Delayed Recall Test that Were Also Recalled in the Immediate Recall Test, and the Evidence (BF) for the Difference between Conditions.

	Span Task			BF_{10}		
Exp.	Complex	Simple	Slow	Complex vs. Simple	Slow vs. Simple	Complex vs. Slow
1	0.40 [0.35, 0.45]	0.38 [0.33, 0.44]	0.54 [0.48, 0.59]	0.36	149.6	42.6
2	0.46 [0.41, 0.52]	0.37 [0.32, 0.42]	0.54 [0.46, 0.62]	23.5	45.9	0.69
3	0.56 [0.47, 0.65]	0.36 [0.28, 0.44]	0.54 [0.46, 0.62]	21.9	30.3	0.26
4a	0.57 [0.51, 0.63]	0.34 [0.30, 0.39]		49795.5		0.17
4b		0.42 [0.36, 0.47]	0.60 [0.54, 0.66]		772	
5	0.69 [0.62, 0.75]	0.39 [0.33, 0.45]		136374.8		

Note. In Experiment 4, there were two types of simple span: the distraction-first condition occurred in the same block of trials as complex span, whereas the blank-first condition occurred in the same block of trials as slow-span.

Camos, 2012), free time is used for refreshing, thereby reactivating decaying memory traces. In the SOB-CS model (Oberauer et al., 2016), free time is used for removing distractor representations from memory, thereby reducing interference. At first glance, the contention that free time is used for consolidation, and perhaps elaboration, might appear to stand in opposition to these proposals. However, there is no reason why free time could not be shared between multiple processes. Therefore, our data imply that the assumptions about the use of free time in the TBRS model and in the SOB-CS model are likely to be incomplete, but not necessarily wrong.

Contribution of a testing effect?

Two reviewers of this paper raised the concern that our data may be explained by a testing effect. A testing effect is the observation that retrieved information is better retained in long-term memory than information presented for re-study (Roediger & Butler, 2011; Roediger & Karpicke, 2006). Given that slow-span trials were associated with better immediate recall, this condition may also have lent more opportunities for words to benefit from the retrieval practice that is involved in the testing effect, which in turn would have promoted their encoding into episodic memory. We have two arguments against this interpretation of our findings. First, performance was near ceiling for both the slow-span and the simple-span condition in all experiments reported here. This means that almost all words studied in these conditions were retrieved in the immediate recall test (thus they benefited equally from retrieval practice). Nevertheless, delayed recall was better for words from slow-span compared to simple-span. Second, when we conditionalize delayed recall on immediate recall, the effects critical for our conclusions remain unchanged. Table 4 presents the proportion of words that were recalled in the immediate test (according to the free recall scoring) that were also recalled in the delayed recall test.[4]

As shown in Table 4, correcting for immediate recall baseline differences showed evidence for a McCabe effect in four out of five experiments, and this evidence was larger in the conditions in which slow-span competition was reduced (Experiment 4) or removed (Experiment 5). Correcting for immediate recall performance did not change the slow-span benefit. When delayed recall is conditionalized on immediate recall, slow-span and complex-span yield similar levels of performance. This is to be expected if the processing of distractors impairs immediate and delayed recall to about the same extent. Therefore, although our data do not rule out a contribution of the testing effect to episodic memory performance, the testing effect does not play a role in explaining our findings.

[4] We thank Vanessa Loaiza for providing us with an R-script to perform such analysis.

Across all of our experiments, we have consistently observed that the better way to improve episodic memory is to give plenty of free time for information to be processed in WM – the inclusion of distraction periods only reduces the amount of free time, and yield weaker episodic traces than when no distraction is presented.

How robust is the McCabe effect?

Across the first three experiments of this series we were unable to replicate the McCabe effect with the unconditionalized delayed-recall data. This effect has been replicated multiple times before (Loaiza et al., 2013, 2015; Loaiza & McCabe, 2012a, 2012b; McCabe, 2008), so our repeated failure to replicate it is likely to reflect a systematic boundary condition of the McCabe effect. With Experiments 4 and 5 we identified that boundary condition: The presence of the simple-slow condition in the experiment substantially weakens the McCabe effect. Because this was an unanticipated finding, we did not yet determine the reason for this boundary condition. We offer one plausible explanation: Words from the simple-slow condition out-compete words from complex-span trials in the free-recall test. Other explanations are certainly possible, and we must leave it to future research to adjudicate between them.

Conclusion

When we are attempting to create better episodic cues for retrieval over the long-term, providing ample free time for information to remain in WM is the most effective way of creating enduring memories.

Acknowledgment

This research was supported by a grant from the Swiss National Science Foundation (project 149193) to Klaus Oberauer.

A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jml.2017.07.002.

References

Barrouillet, P., & Camos, V. (2012). As time goes by temporal constraints in working memory. *Current Directions in Psychological Science, 21*(6), 413–419. http://dx.doi.org/10.1177/0963721412459513.

Barrouillet, P., Portrat, S., & Camos, V. (2011). On the law relating processing to storage in working memory. *Psychological Review, 118*(2), 175–192. http://dx.doi.org/10.1037/a0022324.

Bayliss, D. M., Bogdanovs, J., & Jarrold, C. (2015). Consolidating working memory: Distinguishing the effects of consolidation, rehearsal and attentional refreshing in a working memory span task. *Journal of Memory and Language, 81*, 34–50. http://dx.doi.org/10.1016/j.jml.2014.12.004.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114*(3), 539–576. http://dx.doi.org/10.1037/0033-295X.114.3.539.

Camos, V., & Portrat, S. (2015). The impact of cognitive load on delayed recall. *Psychonomic Bulletin & Review, 22*(4), 1029–1034. http://dx.doi.org/10.3758/s13423-014-0772-5.

Craik, F. I. M. (2002). Levels of processing: Past, presentEllipsis and future? *Memory, 10*(5–6), 305–318. http://dx.doi.org/10.1080/09658210244000135.

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.

Grillon, M.-L., Johnson, M. K., Krebs, M.-O., & Huron, C. (2008). Comparing effects of perceptual and reflective repetition on subjective experience during later recognition memory. *Consciousness and Cognition, 17*(3), 753–764. http://dx.doi.org/10.1016/j.concog.2007.09.004.

Johnson, M. H. (2012). Executive function and developmental disorders: the flip side of the coin. *Trends in Cognitive Sciences, 16*(9), 454–457. http://dx.doi.org/10.1016/j.tics.2012.07.001.

Johnson, M. R., Higgins, J. A., Norman, K. A., Sederberg, P. B., Smith, T. A., & Johnson, M. K. (2013). Foraging for thought: An inhibition-of-return-like effect resulting from directing attention within working memory. *Psychological Science, 24*(7), 1104–1112. http://dx.doi.org/10.1177/0956797612466414.

Johnson, M. K., Reeder, J. A., Raye, C. L., & Mitchell, K. J. (2002). Second thoughts versus second looks: An age-related deficit in reflectively refreshing just-activated information. *Psychological Science, 13*(1), 64–67. http://dx.doi.org/10.1111/1467-9280.00411.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General, 142*(2), 573–603. http://dx.doi.org/10.1037/a0029146.

Loaiza, V. M., Duperreault, K. A., Rhodes, M. G., & McCabe, D. P. (2014). Long-term semantic representations moderate the effect of attentional refreshing on episodic memory. *Psychonomic Bulletin & Review, 1–7*. http://dx.doi.org/10.3758/s13423-014-0673-7.

Loaiza, V. M., Duperreault, K. A., Rhodes, M. G., & McCabe, D. P. (2015). Long-term semantic representations moderate the effect of attentional refreshing on episodic memory. *Psychonomic Bulletin & Review, 22*(1), 274–280. http://dx.doi.org/10.3758/s13423-014-0673-7.

Loaiza, V. M., & McCabe, D. P. (2012a). Temporal–contextual processing in working memory: Evidence from delayed cued recall and delayed free recall tests. *Memory & Cognition, 40*(2), 191–203. http://dx.doi.org/10.3758/s13421-011-0148-2.

Loaiza, V. M., & McCabe, D. P. (2012b). The influence of aging on attentional refreshing and articulatory rehearsal during working memory on later episodic memory performance. *Aging, Neuropsychology, and Cognition, 20*(4), 471–493. http://dx.doi.org/10.1080/13825585.2012.738289.

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. http://dx.doi.org/10.1037/a0023923.

Loaiza, V. M., Rhodes, M. G., & Anglin, J. (2013). The influence of age-related differences in prior knowledge and attentional refreshing opportunities on episodic memory. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, gbt119*. http://dx.doi.org/10.1093/geronb/gbt119.

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS One, 7*(8), e43230. http://dx.doi.org/10.1371/journal.pone.0043230.

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*(2), 480–494. http://dx.doi.org/10.1016/j.jml.2007.04.004.

Morey, R. D. & Rouder, J. N. (2014). BayesFactor: Computation of Bayes factors for common designs (Version 0.9.9). Retrieved from <http://CRAN.R-project.org/package=BayesFactor>.

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin, 142*(7), 758–799. http://dx.doi.org/10.1037/bul0000046.

R core team. (2014). *R: A language and environment for statistical computing*. Vienna: Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>.

Ricker, T. J. (2015). The role of short-term consolidation in memory persistence. *AIMS Neuroscience, 2*(4), 259–279. http://dx.doi.org/10.3934/Neuroscience.2015.4.259.

Ricker, T. J., & Cowan, N. (2014). Differences between presentation methods in working memory procedures: A matter of working memory consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(2), 417.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27. http://dx.doi.org/10.1016/j.tics.2010.09.003.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational Practice. *Perspectives on Psychological Science, 1*(3), 181–210. http://dx.doi.org/10.1111/j.1745-6916.2006.00012.x.

Rose, N. S., Buchsbaum, B. R., & Craik, F. I. M. (2014). Short-term retention of a single word relies on retrieval from long-term memory when both rehearsal and refreshing are disrupted. *Memory & Cognition, 42*(5), 689–700. http://dx.doi.org/10.3758/s13421-014-0398-x.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*(5), 356–374. http://dx.doi.org/10.1016/j.jmp.2012.08.001.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237. http://dx.doi.org/10.3758/PBR.16.2.225.

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends in Cognitive Sciences, 4*(3), 108–115. http://dx.doi.org/10.1016/S1364-6613(00)01445-5.

Schrijver, S. D., & Barrouillet, P. (2017). Consolidation and restoration of memory traces in working memory. *Psychonomic Bulletin & Review, 1–7*. http://dx.doi.org/10.3758/s13423-017-1226-7.

Souza, A. S., & Oberauer, K. (2017). The contributions of visual and central attention to visual working memory. *Attention Perception & Psychophysics*. http://dx.doi.org/10.3758/s13414-017-1357-y.

Souza, A. S., Rerko, L., & Oberauer, K. (2015). Refreshing memory traces: Thinking of an item improves retrieval from visual working memory. *Annals of the New York Academy of Sciences, 1339*(1), 20–31. http://dx.doi.org/10.1111/nyas.12603.

Tulving (1985). *Elements of episodic memory*. New York: Oxford University Press, USA.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127–154. http://dx.doi.org/10.1016/0749-596X(89)90040-5.

Unsworth, N., Heitz, R. P., & Parks, N. A. (2008). The importance of temporal distinctiveness for forgetting over the short term. *Psychological Science, 19*(11), 1078–1081. http://dx.doi.org/10.1111/j.1467-9280.2008.02203.x.

Vergauwe, E., Barrouillet, P., & Camos, V. (2010). Do mental processes share a domain-general resource? *Psychological Science, 21*(3), 384–390. http://dx.doi.org/10.1177/0956797610361340.

Vergauwe, E., Camos, V., & Barrouillet, P. (2014). The impact of storage on processing: How is information maintained in working memory? *Journal of Experimental Psychology: Learning, Memory & Cognition*. Advance online publication.

Vergauwe, E., & Cowan, N. (2014). Attending to items in working memory: Evidence that refreshing and memory search are closely related. *Psychonomic Bulletin & Review, 22*(4), 1001–1006. http://dx.doi.org/10.3758/s13423-014-0755-6.

Vergauwe, E., & Langerock, N. (2017). Attentional refreshing of information in working memory: Increased immediate accessibility of just-refreshed representations. *Journal of Memory and Language, 96*, 23–35. http://dx.doi.org/10.1016/j.jml.2017.05.001.