



Secondary task engagement drives the McCabe effect in long-term memory

Kelly Cotton¹ · Joshua Sandry² · Timothy J. Ricker³

Accepted: 31 July 2023
© The Psychonomic Society, Inc. 2023

Abstract

Processing that occurs while information is held in working memory is critical in long-term retention of that information. One counterintuitive finding is that the concurrent processing required during complex span tasks typically impairs immediate memory, while also leading to improved delayed memory. One proposed mechanism for this effect is retrieval practice that occurs each time memory items are displaced to allow for concurrent processing during complex span tasks. Other research has instead suggested that increased free time during complex span procedures underlies this effect. In the present study, we presented participants with memory items in simple, complex, and slow span tasks and compared their performance on immediate and delayed memory tests. We found that how much a participant engaged with the secondary task of the complex span task corresponded with how strongly they exhibited a complex span boost on delayed memory performance. We also probed what participants were thinking about during the task, and found that participants' focus varied depending both on task type and secondary task engagement. The results support repeated retrieval as a key mechanism in the relationship between working memory processing and long-term retention. Further, the present study highlights the importance of variation in individual cognitive processing in predicting long-term outcomes even when objective conditions remain unchanged.

Keywords Working memory · Short-term memory · Long-term memory · Mind-wandering · Complex span

Introduction

People are often able to create memories that are useful in the moment, yet still able to be retrieved after considerable delay. The working memory system is key in creating and maintaining memories to be used in ongoing cognition, but the importance of this initial processing in the creation of lasting long-term memory and the mechanisms to support this relationship remain debated. While early models conceptualized a working memory system distinct from long-term memory (Atkinson & Shiffrin, 1968; Baddeley & Hitch, 1974), later models viewed working memory as being much more intertwined with the long-term memory

system (Cowan, 1988; Oberauer, 2002; but see Norris, 2017, 2019). The present study aimed to investigate the mechanisms that support a working memory processing benefit on long-term memory to better understand the relationship between these systems.

Processing that occurs while information is held in working memory improves long-term retention of this information (Cotton & Ricker, 2021; Hartshorne & Makovski, 2019; Jarjat et al., 2018; McCabe, 2008; Sandry et al., 2020; Souza & Oberauer, 2017). Memory span tasks have been frequently used to explore this relationship. In a typical simple span task, a series of memory items are presented sequentially before participants are asked to recall each memory item. In a complex span task, the memory items are interspersed with a secondary processing task, such as an arithmetic problem (Conway et al., 2005; Daneman & Carpenter, 1980; Turner & Engle, 1989). Previous research has found that while immediate recall is better in simple span compared to complex span tasks, the reverse is true for delayed-recall tests (Loaiza & McCabe, 2012, 2013; McCabe, 2008). In a series of experiments, McCabe (2008) tested participants' immediate and delayed recall for words presented in simple and complex span tasks. As expected, immediate-recall

✉ Kelly Cotton
kcotton@gradcenter.cuny.edu

¹ Department of Psychology, The Graduate Center, City University of New York, 365 5th Ave, New York, NY 10016, USA

² Department of Psychology, Montclair State University, Montclair, NJ, USA

³ Department of Psychology, University of South Dakota, Vermillion, SD, USA

performance was significantly worse on complex span trials compared to simple span trials. However, after a delay, participants were more likely to recall the words that had originally been presented during the complex span trials. Similar results were found regardless of whether the participants expected the delayed test or the presence of an immediate test. This pattern of results, now termed the McCabe effect, has been extended to delayed recognition (Loaiza et al., 2015).

The mechanism initially proposed to underlie the McCabe effect was covert retrieval, or the process of retrieving a recently displaced item and reactivating it in the focus of attention (McCabe, 2008). In simple span tasks, the memory items are not displaced from the focus of attention and so do not need to be covertly retrieved. However, in complex span tasks, the secondary processing task diverts attention away from the memory item and so the item must be retrieved from long-term memory after completing the processing task. This repeated retrieval enhances long-term activation of the items, making them easier to recall after a delay compared to items that were retrieved less often. However, Souza and Oberauer (2017) questioned the exact role of the intervening task in the McCabe effect. The authors conducted a similar series of experiments but included a slow span task, during which memory items are presented at a pace similar to complex span but without the intervening secondary task. They failed to replicate the McCabe effect, instead showing that the slow span task consistently resulted in better delayed recall compared to the other span tasks. The authors suggested that the extended free time available after memory item presentation during complex or slow span tasks allow for processes other than repeated retrieval to benefit long-term retention, such as consolidation (Cotton & Ricker, 2021; Ricker, 2015) or elaboration (Bartsch et al., 2018, 2019; Loaiza & Lavilla, 2021).

While the role of the intervening task during complex span tasks and thus the mechanism supporting a long-term benefit of working memory processing remains debated, there has been little work examining how participants engage with the secondary task. Previous studies have reported extremely high accuracy on the secondary task overall or have excluded participants who performed poorly, limiting the variability in participants. Given the extensive research demonstrating individual differences in working memory performance (Daneman & Carpenter, 1980; Unsworth & Engle, 2007), it may be informative to consider a wider distribution of participant performance to better understand the range in human memory ability. It is unlikely that individuals always engage fully with both tasks during real-life dual-tasking situations. Understanding cognitive performance across a wider range of ability and engagement will also help us better understand day-to-day cognitive performance in the general population. The present study includes data from a large online sample to explore mechanisms supporting the

effect of working memory processing on long-term memory across a greater range of task performance.

One explanation for variability in participant performance during a complex span task may be that participants have different task goals and focus their attention on different aspects of the task. If a participant prioritizes the primary memory task, they may sacrifice their performance on the secondary processing task to achieve their goal. This participant may not engage with the secondary task fully or at all, resulting in poor secondary task performance but good immediate memory performance. However, given the mechanisms outlined previously to explain the McCabe effect, it is unclear how this participant would perform on a delayed memory test. In two earlier experiments, we found that participants varied in their engagement with the secondary task during complex working memory tasks in an online setting (Cotton et al., 2023). Given the results of the earlier experiments, we designed the present study to leverage this variability in task engagement in an online setting to better understand the boundaries of long-term memory and, specifically, the McCabe effect.

Engagement with the secondary task should produce one of two potential outcomes in the present study. The first potential outcome is that only participants who engage with the secondary task during a complex span task show evidence of a McCabe effect. If the McCabe effect occurs because displaced items must be retrieved back into working memory or the focus of attention, then this pattern would be expected. Participants who either opt not to or are unable to engage with the secondary task should show no difference in delayed performance across different span tasks because they do not need to retrieve displaced memory items back into the focus of attention. This would provide supporting evidence that retrieval practice underlies the McCabe effect. Second, we may find that engaging or not engaging with a secondary task has no relationship with long-term retention. That is, even participants who are not fully engaged with the secondary task and perform poorly on it still exhibit evidence of the McCabe effect. This would suggest that the extended free time associated with complex span tasks relative to simple span tasks drives the long-term memory boost.

Methods

Study population and sample size determination

We aimed to collect usable data from approximately 250 participants. Once we reached this threshold, we conducted preliminary Bayes factor analyses. If the resulting Bayes factors were ambiguous (< 3), we planned to collect data from 50 additional participants and re-run the analyses, repeating this procedure until we had substantial Bayes factors. A total of 260 students (78% female, mean age = 19.3 years) who

were enrolled in an introductory psychology course participated online in exchange for course credit if they completed the study procedures fully. This experiment was approved by the Institutional Review Boards at the University of South Dakota and Montclair State University.

Materials

The word stimuli consisted of 687 nouns taken from the MRC Psycholinguistic Database (Coltheart, 1981). The words comprised four to six letters and one to two syllables, and had concreteness and familiarity ratings of 450–700. For each participant, the list of words was shuffled in a random order. In the first trial of the practice phase, the first four words were selected from the randomly shuffled list to be presented during that trial. The next trial then used the next four words in the list, and so on, until all practice trials were completed. This continued for the trials of the immediate-recall phase. Subjective experience of everyday mind-wandering was assessed with the Mindfulness Attention Awareness Scale (MAAS; Brown & Ryan, 2003). The experiment was created using PsychoPy software (Peirce et al., 2019).

Design

The experiment was divided into three phases, the immediate-recall phase, the distraction phase, and the delayed-recall phase. At the end of the experiment participants completed the MAAS scale questionnaire.¹ In the immediate-recall phase, participants completed three tasks, simple span, complex span, and slow span. In this phase, memory was assessed on each trial with immediate recall. In the delayed-recognition phase participants were tested on their recognition of words presented during the immediate phase. Details on the procedures for each task and scoring methods are included below.

During the experimental trials of the immediate-recall phase, there were six blocks of seven trials each. Each block consisted of only one type of trial (simple, complex, or slow span) and each block type was presented twice for a total of 14 trials per task. The blocks were randomly ordered. During the delayed phase, participants completed 168 trials of delayed recognition.

Procedure

Immediate-recall phase

At the beginning of the immediate-recall phase of the experiment, participants first completed a short practice session. During the practice session, they completed one trial of the

simple span task and one trial of the slow span task and four trials of the secondary parity judgment task separately and then two trials of the complex span task.² Prior to beginning the experimental trials, participants saw the following instructions: “Everything will be the same as in the practice. On some trials, you will just see the words and on other trials you will see the words and the numbers. Remember, both tasks are important! Please try to respond as quickly and accurately as possible.” An example of a single trial for each span task is presented in Fig. 1.

Simple span task At the beginning of each trial, a fixation cross appeared for 500 ms. Participants then saw four words presented sequentially. Each word was on screen for 900 ms, with a 100-ms inter-stimulus interval. After presentation of all four words, participants were asked to recall the words in the order they had been presented by typing each word. If they did not remember a word, they were instructed to simply press the return key. After completing the recall, participants were asked to report what they were thinking about during the preceding trial. Participants were given five categorical options: (1) the word task, (2) the number task, (3) both tasks, (4) task experience/performance, and (5) something else. Participants indicated via button press which categorical response best fits what they were thinking about. After responding to the thought probe, a new trial began.

Complex span task The procedure for the complex span task was the same as the simple span task except in the complex span task, participants completed the secondary task after each individual memory item. During the secondary task, a single digit was presented for 1,650 ms and participants indicated by button press if the number was odd or even. After a 100-ms interstimulus interval this procedure was repeated for a second digit resulting in a total secondary task length of 3,500 ms between each memory item. As in the simple span task, participants were asked to report what they were just thinking about. After responding to the thought probe, a new trial began.

Slow span task The procedure for the slow span task was the same as the simple span task except in the slow span task, there was a 3,500-ms delay with a blank screen between each individual memory item. This was the same amount of time between items as in the complex span task, except that in the slow span task there was no secondary task. As in the simple

¹ We do not report the results from the MAAS questionnaire as an error in data collection led to the loss of this data in roughly half of the participants.

² Participants saw the following instructions before beginning the complex span practice: “...On each trial, you will see a word, followed by two numbers. Again, you will decide if each number is even or odd. This will repeat four times. After you have seen all of the words, you will be asked to recall as many words as you can remember in the order they were originally presented in...”



Fig. 1 The procedure for a typical trial during each span task. (a) A single trial of the simple span task. (b) A single trial of the complex span task. (c) A single trial of the slow span task

span task, participants were asked to report what they were just thinking about. After responding to the thought probe, a new trial began.

Distractor phase

The distractor task consisted of approximately 5 min of a go/no-go task. A nonverbal task was chosen to reduce potential interference from the use of verbal material before the delayed recognition test. During this task, participants were instructed that if they saw a green circle on the screen, they should press the space bar, and if they saw a blue circle, they should not press any key. Each circle was presented for 2,000 ms, or until the participant pressed the space bar. Participants completed three blocks of 150 trials, for a total of 450 trials, of which 10% were no-go trials.

Delayed-recognition phase

Each trial began with a fixation cross for 500 ms. Participant then saw two words presented on the screen: a target word taken from the earlier immediate-recall task and a novel word. Participants indicated via button press which word they believed they had previously seen. All words from the immediate-recall phase were presented in random order during a single block.

Data analysis

For our statistical analyses, we used Bayes factors for t-tests (Rouder et al., 2009) and analysis of variance (ANOVA) effects (Rouder et al., 2012) as our measure of inference using R statistical computing software (R Core Team, 2019). Bayes factors indicate the probability of the data assuming an effect of the manipulation relative to the probability of the data assuming no effect. To compute Bayes factors, we used the BayesFactor package v0.9.12-4.3 (Morey & Rouder, 2018), with the default settings of the package with an effect size standard deviation of $(\sqrt{2})/2$.

Participant responses on the immediate-recall test were manually cleaned for typos and misspellings. If a response was one letter different from the correct answer (e.g., “srive” in place of “drive”), contained a letter swap (e.g., “peice” in place of “piece”), or was a pluralization (e.g., “women” in place of “woman”), it was counted as correct, unless the misspelling was a real word (e.g., “soup” in place of “soap”). We scored the immediate-recall tasks using both serial recall and free recall criteria. For serial recall accuracy, a response was marked as correct if it was recalled in the correct serial position. For free recall accuracy, a response was marked as correct if it was recalled during the trial, regardless of serial position.

One participant pressed an invalid key (“6”) in response to the mind-wandering probe for two trials and we coded these responses as thinking about “something else.” Data and analyses scripts are openly available at the project’s Open Science Framework page (<https://osf.io/e3fqu/>).

Results

In this section we identify whether statistical analyses were planned a priori or as post hoc analyses by labeling the sub-heading as either “(Planned)” or “(Exploratory).” As mentioned in the above sections, most of our analyses were planned based on previous findings reported in Cotton et al. (2023).

Memory task performance

Immediate serial recall accuracy by span task (planned)

Overall, participant performance on the immediate serial recall test was high ($M = .79$, $SD = .18$). Serial recall was highest in the slow span task ($M = .82$, $SD = .20$), followed by the simple span ($M = .80$, $SD = .17$). Performance was worst in the complex span task ($M = .74$, $SD = .22$). These results are depicted in Fig. 2a. A Bayesian one-way ANOVA found evidence for a difference in serial recall accuracy between

block conditions, $F(2,776) = 10.2$, $BF = 95$ in favor of an effect of span task. Follow-up t-tests indicated that performance was different between simple and complex span ($BF = 15$ in favor of an effect of span task), and slow and complex span ($BF = 211$ in favor of an effect of span task), but not simple and slow span ($BF = 6.1$ in favor of a null effect).

Immediate free recall accuracy by span task (planned)

Immediate free recall was also high overall ($M = .85$, $SD = .13$). Free recall was again highest in the slow span task ($M = .88$, $SD = .15$) followed by the simple span ($M = .86$, $SD = .13$), and worst in the complex span task ($M = .82$, $SD = .17$). These results are depicted in Fig. 2b. A Bayesian one-way ANOVA found evidence for a difference in serial recall accuracy between block conditions, $F(2,776) = 10.3$, $BF = 127$ in favor of an effect of span task. Follow-up t-tests indicated a similar pattern of results as serial recall: performance was different between simple and complex span ($BF = 18$ in favor of an effect of span task) and slow and complex span ($BF = 221$ in favor of an effect of span task), but not simple and slow span ($BF = 6.2$ in favor of a null effect).

Delayed recognition accuracy by original span task (planned)

Across all participants, performance on the delayed-recognition task was high ($M = .73$, $SD = .14$). Accuracy was similar across all span tasks, though it was slightly higher

for words originally presented during the slow span task ($M = .74$, $SD = .15$), followed by the complex span task ($M = .73$, $SD = .15$), and lowest for the simple span task ($M = .71$, $SD = .14$). These results are depicted in Fig. 2c. A Bayesian one-way ANOVA failed to find evidence for a difference in recognition accuracy between originally presented span tasks, $F(2, 776) = 2.48$, $BF = 12$ in favor of a null effect.

Distractor task performance (planned)

Performance on the distractor task was high ($M = .96$, $SD = .07$).

Secondary task performance (planned)

Based on the results from earlier experiments (Cotton et al., 2023), we analyzed performance on the secondary task as a measure of participant engagement with the task. Overall, performance on the secondary parity judgement task during the immediate-recall task was low ($M = .58$, $SD = .40$). However, individual participant performance varied considerably. Of the total 260 participants, 129 performed relatively poorly ($\leq 80\%$ accuracy), with 62 participants failing to respond to any secondary task item.

We divided participants into “high” ($> 80\%$ accuracy, $n = 131$), “low” ($\leq 80\%$ accuracy, $n = 67$), or “zero” (0% accuracy, $n = 62$) secondary task performers. Immediate serial and free recall and delayed recognition performance across the three span tasks and secondary task performance groups is presented in Fig. 3. High secondary task

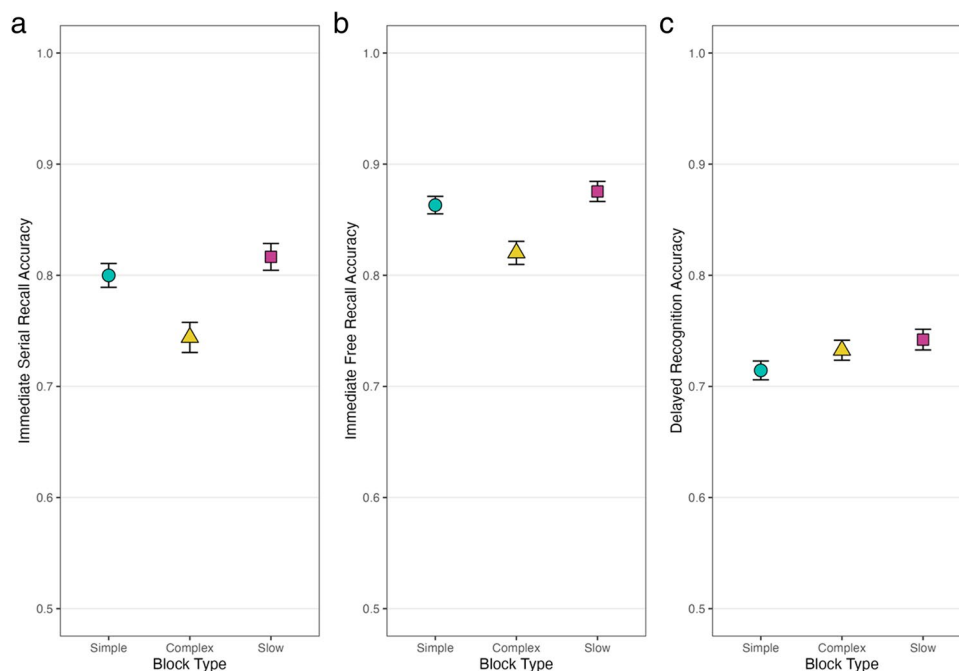


Fig. 2 Overall memory performance across span tasks. Error bars represent standard error of the mean

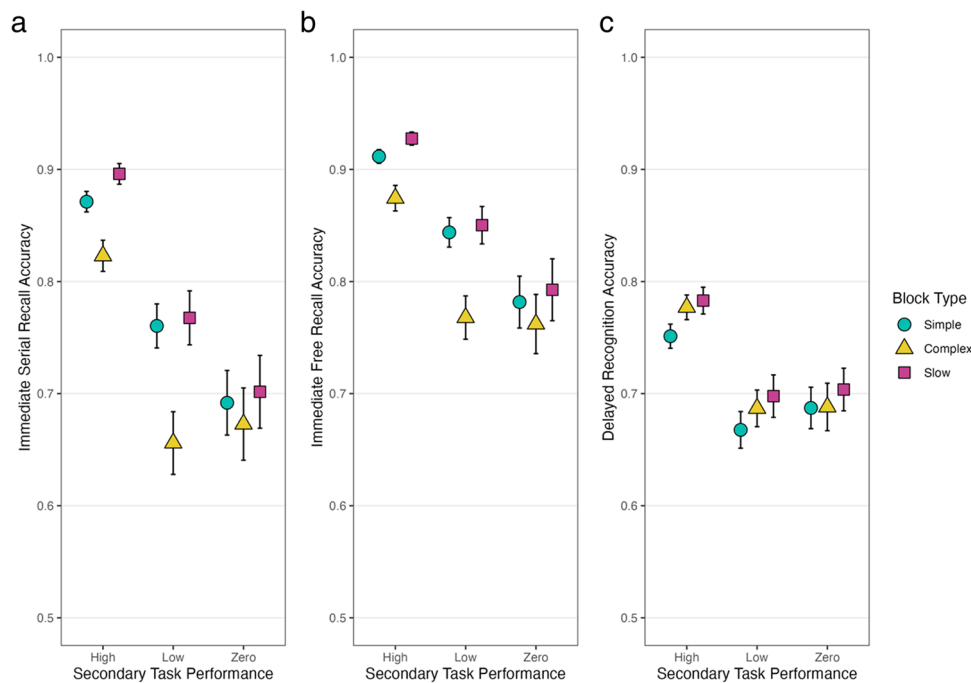


Fig. 3 Memory performance across secondary task performance and span tasks. Error bars represent standard error of the mean

performers displayed the typical McCabe effect pattern: better immediate memory in simple span (serial recall: $M = .87$, $SD = .11$; free recall: $M = .91$, $SD = .07$) compared to the complex span (serial recall: $M = .82$, $SD = .16$; free recall: $M = .87$, $SD = .13$), and better delayed memory in complex span ($M = .78$, $SD = .13$) compared to simple span ($M = .75$, $SD = .13$). Bayesian t-tests found evidence for a difference in both immediate serial recall, $t(130) = 4.3$, $BF = 500$ in favor of an effect, and free recall, $t(130) = 3.7$, $BF = 55$ in favor of an effect, as well as delayed recognition, $t(130) = -3.3$, $BF = 18$ in favor of an effect.

In contrast, those participants who failed to engage with the secondary task at all (the “zero” group) showed no evidence of the McCabe effect. Immediate memory performance was similar in both the simple span (serial recall $M = .69$, $SD = .23$; free recall $M = .78$, $SD = .18$) and the complex span (serial recall: $M = .67$, $SD = .25$; free recall: $M = .76$, $SD = .21$) tasks. Bayesian t-tests failed to find evidence for a difference in either immediate serial recall, $t(61) = .87$, $BF = 5.0$ in favor of a null effect, or free recall, $t(61) = 1.2$, $BF = 3.7$ in favor of a null effect. Delayed memory performance was also similar in both the simple span ($M = .69$, $SD = .15$) and the complex span ($M = .69$, $SD = .17$) tasks, $t(61) = -.07$, $BF = 7.2$ in favor of a null effect.

The middle group, those who were somewhat engaged but still exhibited poor performance on the secondary task, showed mixed evidence of the McCabe effect. Immediate memory performance was better in the simple span (serial recall $M = .76$,

$SD = .16$; free recall $M = .84$, $SD = .11$) compared to the complex span (serial recall $M = .66$, $SD = .23$; free recall: $M = .77$, $SD = .16$) task. Bayesian t-tests found evidence for a difference in both immediate serial recall, $t(66) = 5.04$, $BF = 4314$ in favor of an effect, and free recall, $t(66) = 5.2$, $BF = 6668$ in favor of an effect. However, we found only ambiguous evidence that delayed memory performance was higher in the complex span ($M = .69$, $SD = .13$) compared to the simple span ($M = .67$, $SD = .13$) task, $t(66) = -2.3$, $BF = 1.6$ in favor of an effect.

Thought probe response

Overall reported thought (planned)

Table 1 presents the percentage of responses to the thought probe by presentation span task. Notably, most participants reported thinking about “the word task” most of the time in the simple (75.1% of all responses) and slow span (69.8%) tasks. Participants reported similar levels of thinking about “the word task” (35.5%) or “both tasks” (34.7%) in the complex span task. Participants reported not thinking directly about either task (i.e., “task experience/performance” or “something else”) most often in the slow span task (24.1%).

The results below describe in detail the variation across both secondary task performance group and span task, both in what participants report thinking about and how it affects their performance on the memory tasks. In summary, simple and slow span tasks show similar patterns of responses to

Table 1 Responses to thought probe across all trials by span task

Thought probe	Simple span %	Complex span %	Slow span %
“The word task”	75.1	35.5	69.8
“The number task”	2.30	7.10	1.90
“Both tasks”	4.20	34.7	4.20
“Task experience/ performance”	7.60	9.60	8.70
“Something else”	10.8	13.0	15.4

the thought probe, regardless of secondary task performance, while the complex span task varies considerably more. Further, we found that thinking about the primary word task led to the best immediate-recall performance, while thinking about the secondary task led to lower performance. Finally, thinking about neither (i.e., “task experience” or “something else”) resulted in the lowest performance across all span tasks.

Thought probe response across secondary task performance groups (planned)

Responses to the thought probe varied across secondary task performance groups and span tasks. These results are plotted in Fig. 4. Notably, the simple span (Fig. 4a) and slow span (Fig. 4c) tasks show a similar pattern of results in all three secondary task performance groups. In these span tasks, only thinking about “the word task” can be considered “on-task” thought, as there was no secondary task. Participants in

all three secondary task performance groups reported thinking about “the word task” most often in the simple span trials (High $M = 78.1, SD = 23.9$; Low $M = 76.8, SD = 22.7$; Zero $M = 66.9, SD = 32.2$) and slow span trials (High $M = 73.9, SD = 24.5$; Low $M = 69.4, SD = 28.8$; Zero $M = 61.5, SD = 32.6$). To compare the proportion of trials in which participants selected “the word task” across the three secondary task performance groups for simple and slow span trials, we conducted a Bayesian two-way ANOVA. We found evidence of a main effect of secondary task performance group, $BF = 28$ in favor of an effect, but no main effect of span task, $BF = 1.2$ in favor of a null effect. There was no evidence to support an interaction effect, $BF = 44$ in favor of a null effect.

In contrast, the complex span trials showed more variability, which is expected as there are multiple ways to define “on-task” thought in this span task. We expected that participant engagement with the secondary task should drive how often they report thinking about each “on-task” thought probe. A participant who was fully engaged with all components of the task would report thinking about “both tasks” most often. Participants who performed very well on the secondary task performance reported thinking about “both tasks” ($M = 45.1, SD = 30.1$) most often, while those participants who failed to engage with the secondary task at all reported thinking about “both tasks” ($M = 16.9, SD = 21.6$) much less often. Participants who somewhat engaged with but performed poorly on the secondary task were in the middle ($M = 30.7, SD = 23.8$). A Bayesian one-way ANOVA found evidence of an effect of secondary task performance

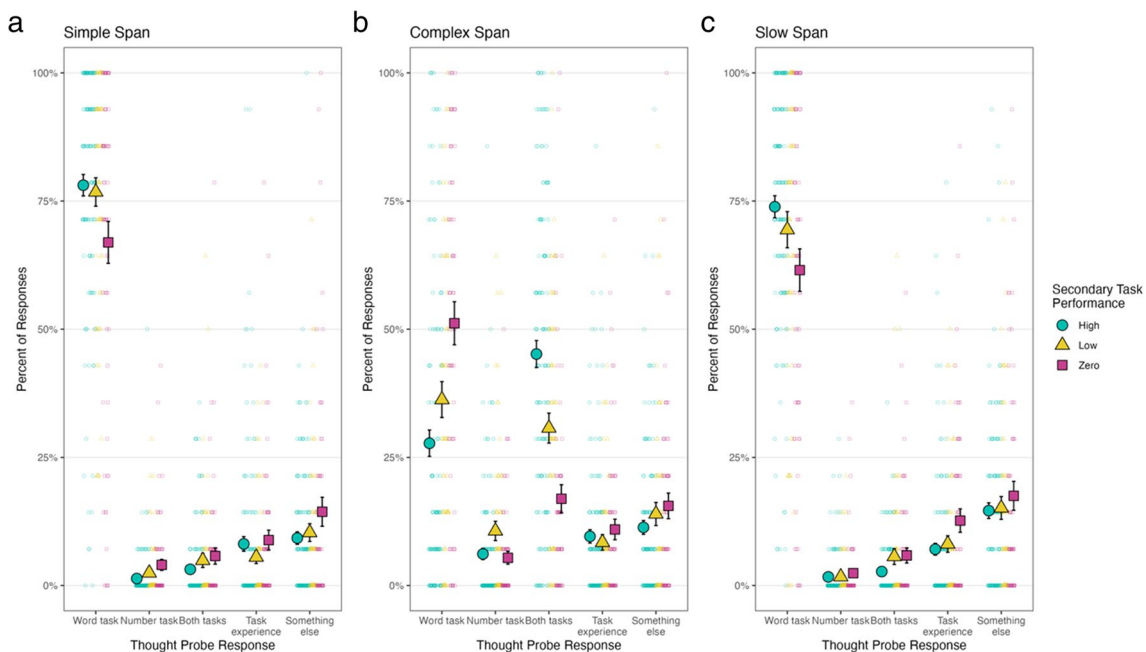


Fig. 4 Responses to thought probe by secondary task condition block. Error bars represent standard error of the mean

group on proportion of “both task” complex span trials, $BF = 5.0 \times 10^7$ in favor of an effect.

Another way to be considered “on-task” would be thinking about either “the word task” or “the number task,” which should also differ depending on secondary task engagement. Participants who were not engaged with the secondary task at all reported thinking about “the word task” most often ($M = 51.2$, $SD = 33.1$), followed by the somewhat engaged ($M = 36.3$, $SD = 28.6$), and finally the high secondary task performers ($M = 27.7$, $SD = 29.3$). A Bayesian one-way ANOVA found evidence of an effect of secondary task performance group on proportion of “the word task” complex span trials, $BF = 2593$ in favor of an effect. Thinking about “the number task” was overall very low across all groups (High $M = 6.15$, $SD = 12.4$; Low $M = 10.7$, $SD = 15.3$; Zero $M = 5.41$, $SD = 9.74$), and a Bayesian one-way ANOVA found only ambiguous evidence against an effect of secondary task performance group on proportion of “the number task” complex span trials, $BF = 1.6$ in favor of a null effect.

What is more typically considered “off-task” thoughts are the “task experience/performance” and “something else” responses. Thinking about the “task experience/performance” was similar between the high (simple $M = 8.12$, $SD = 16.3$; complex $M = 9.60$, $SD = 15.0$; slow $M = 7.09$, $SD = 13.2$), low (simple $M = 5.54$, $SD = 10.2$; complex $M = 8.41$, $SD = 12.4$; slow $M = 8.10$, $SD = 13.1$), and zero secondary task performance groups (simple $M = 8.87$, $SD = 15.2$; complex $M = 10.9$, $SD = 15.7$; slow $M = 12.7$, $SD = 18.0$). To compare the proportion of trials in which participants selected “task experience/performance” across the three secondary task performance groups and span tasks, we conducted a Bayesian two-way ANOVA. We found no evidence of main effects of either secondary task performance group, $BF = 5.9$ in favor of a null effect, or span task, $BF = 10$ in favor of a null effect, or an interaction effect, $BF = 685$ in favor of a null effect.

The final probe may be considered the most “off-task” thought and may be influenced by participant engagement or span task. However, we found that thinking about “something else” was similar between the high (simple $M = 9.27$, $SD = 14.2$; complex $M = 11.3$, $SD = 15.4$; slow $M = 14.6$, $SD = 17.4$), low (simple $M = 10.3$, $SD = 14.2$; complex $M = 14.0$, $SD = 18.5$; slow $M = 15.1$, $SD = 18.2$), and zero secondary task performance groups (simple $M = 14.4$, $SD = 22.3$; complex $M = 15.5$, $SD = 19.7$; slow $M = 17.5$, $SD = 22.2$). A Bayesian two-way ANOVA found evidence for a main effect of span task, $BF = 6.0$ in favor of an effect, but no main effect of secondary task performance group, $BF = 9.0$ in favor of a null effect, or an interaction effect, $BF = 377$ in favor of a null effect. However, this type of off-task thinking was generally low in the present study.

Immediate recall accuracy and reported thought (exploratory)

On trials in which participants reported only thinking about “the word task,” serial recall performance was high across all block conditions (simple: $M = .84$, $SD = .16$; complex: $M = .84$, $SD = .20$; slow: $M = .86$, $SD = .16$), $BF = 32$ in favor of a null effect. Free recall performance was also generally high for these trials (simple: $M = .89$, $SD = .10$; complex: $M = .89$, $SD = .15$; slow: $M = .91$, $SD = .11$), $BF = 14$ in favor of a null effect. For complex span trials, serial recall performance was lower when participants reported thinking about “both tasks” ($M = .74$, $SD = .26$) compared to “the word task,” $BF = 569$ in favor of an effect, and even worse when they reported thinking about “the number task” ($M = .61$, $SD = .36$) compared to “the word task,” $BF = 3.3 \times 10^9$ in favor of an effect, though this was a much smaller percentage of trials. A similar pattern was found in free recall (“both tasks” $M = .83$, $SD = .19$, $BF = 133$ in favor of an effect; “the number task” $M = .72$, $SD = .29$, $BF = 3.4 \times 10^8$ in favor of an effect).

When participants engaged in conventional “off-task” thought, performance dropped similarly across all span tasks. Serial recall performance, when participants reported thinking about “task experience/performance,” was similar across block conditions (simple: $M = .72$, $SD = .27$; complex: $M = .72$, $SD = .31$; slow: $M = .74$, $SD = .29$), $BF = 47$ in favor of a null effect, as was free recall performance (simple: $M = .79$, $SD = .22$; complex: $M = .80$, $SD = .23$; slow: $M = .81$, $SD = .22$), $BF = 51$ in favor of a null effect. Performance was similarly low in all block conditions for the trials in which participants reported thinking about “something else” in both serial recall (simple: $M = .57$, $SD = .30$; complex: $M = .53$, $SD = .32$; slow: $M = .63$, $SD = .32$), $BF = 1.2$ in favor of a null effect, and free recall (simple: $M = .70$, $SD = .24$; complex: $M = .65$, $SD = .28$; slow: $M = .74$, $SD = .25$), $BF = 1.6$ in favor of an effect. Across all block conditions, performance was considerably lower when participants reported thinking about “something else” compared to any other option in both serial recall, $BF = 1.7 \times 10^{20}$ in favor of an effect, and free recall, $BF = 2.37 \times 10^{21}$ in favor of an effect. These results are depicted in Fig. 5a for serial recall and Fig. 5b for free recall.

Delayed recognition accuracy and reported thought (exploratory)

In contrast to immediate-recall performance, delayed recognition performance was generally less affected by what the participant reported thinking about during initial learning. For words that were originally presented during trials in which participants reported only thinking about “the word task,” delayed recognition performance was similarly high

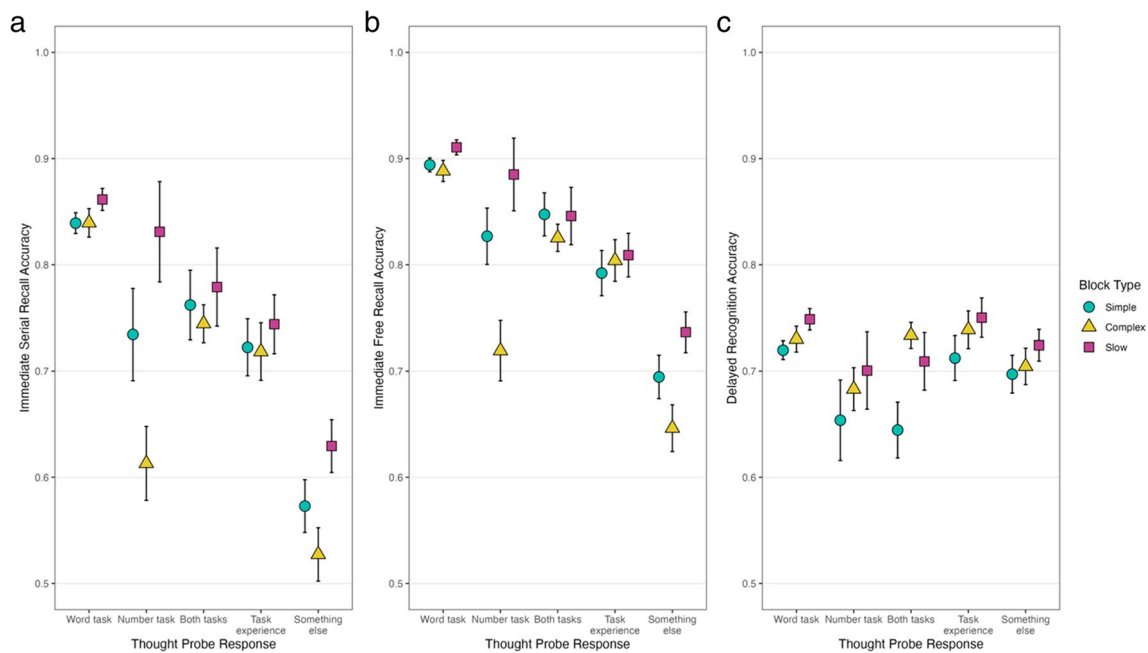


Fig. 5 Memory performance across thought probe responses and span tasks. Error bars represent standard error of the mean

across all block conditions (simple: $M = .72$, $SD = .14$; complex: $M = .73$, $SD = .18$; slow: $M = .75$, $SD = .16$), $BF = 15$ in favor of a null effect. For words originally presented in complex span trials, delayed recognition performance was similar when participants reported thinking about “the word task” and “both tasks” ($M = .73$, $SD = .18$), $BF = 9.2$ in favor of a null effect, and similar when they reported thinking about “the number task” ($M = .68$, $SD = .21$) compared to “the word task,” $BF = 1.1$ in favor of an effect. Performance for “task experience/performance” trials was similar across all block conditions (simple: $M = .71$, $SD = .21$; complex: $M = .74$, $SD = .21$; slow: $M = .75$, $SD = .19$), $BF = 24$ in favor of a null effect. Delayed recognition performance was similarly low for words originally presented in the trials in which participants reported thinking about “something else” compared to any other option, $BF = 3.4$ in favor of a null effect, and was similar across block conditions (simple: $M = .70$, $SD = .21$; complex: $M = .70$, $SD = .22$; slow: $M = .72$, $SD = .19$), $BF = 40$ in favor of a null effect. These results are depicted in Fig. 5c.

Discussion

The goal of the present study was to determine how engagement with the secondary processing task during complex span tasks modulates the McCabe effect in a large online sample. Our initial analyses found that while immediate-recall performance was better in simple span compared to complex span tasks, there was no evidence for the typical

reversal in delayed memory performance. However, follow-up analyses indicated that not all participants engaged with the task in the same way. While many participants performed well on the secondary task, an equal number of participants performed poorly on the task and a considerable number failed to engage with the secondary task at all. By examining performance for these three groups of participants separately, we replicated the McCabe effect only in those participants who fully engaged with the secondary task. Performance in the other groups either showed ambiguous evidence supporting or strong evidence against any McCabe effect. These results underscore the importance of engaging with the secondary task to receive the McCabe effect performance boost in long-term retention.

Previous research has not explored the relationship between secondary task engagement or performance and successful observation of the McCabe effect. However, this is complicated by a lack of reporting of secondary performance or any exclusion criteria in many studies (e.g., Jarjat et al., 2018; Loaiza et al., 2015; Loaiza & McCabe, 2012, 2013; McCabe, 2008). When secondary task performance is reported, it is typically quite high (> 80% accuracy) or the researchers required participants to maintain high secondary task performance to continue (Camos & Portrat, 2015; Loaiza et al., 2021; Loaiza & Halse, 2018). In one study, which reported finding a McCabe effect, five participants were excluded based on a 60% accuracy threshold (Rose et al., 2014). In contrast, Souza and Oberauer (2017) reported typically high secondary task performance (> 80%) and excluded participants for poor performance on

the secondary task (though the exact criteria are not consistently reported), but found no evidence of the McCabe effect in several experiments. However, using similar criteria, Souza and Oberauer (2017) did find evidence of a McCabe effect in other experiments. Still, it is unclear exactly why the participants in the present study performed much more poorly than previous similar studies. It is possible that this was caused by the remote nature of the experiment, as many previous studies have been conducted in-person, which may contribute to the overall high secondary task performance. In one experiment that was conducted online, Loaiza et al. (2021) found evidence of the McCabe effect, but this experiment forced high secondary task performance, and as such there was little variability of secondary task performance in contrast to the present study.

Beyond simple secondary task accuracy measures, we also conducted an exploratory analysis that investigated how participants self-report of task focus influences the McCabe effect. We posited that participants may engage with the task differently and that some may shift their attentional focus away from completing all task requirements on a considerable number of trials. To understand how participants were choosing to complete the task, we asked participants to report “what they were just thinking about” during every trial and allowed them to report whether they were thinking “on-task” thoughts (and which task they were thinking about), “task-related” thoughts, or “off-task” thoughts. Participants varied considerably in what they reported thinking about depending on both the trial type and how well they performed on the secondary task overall. Notably, most participants thought about the primary task most often during the simple and slow span tasks. However, during the complex span trials, when participants should ideally be thinking about both tasks, we saw marked differences. The frequency of thinking about both tasks corresponded with how much the participant engaged with the secondary task, and immediate-recall performance dropped when participants reported thinking about both tasks. In contrast, the delayed-recognition task was much less affected by any variation in reported thought during initial learning. Past research has found mind-wandering to affect performance on many cognitive tasks (for review, see Mooneyham & Schooler, 2013), including working memory performance (e.g., Krinsky et al., 2017; Mrazek et al., 2012), but mind-wandering is often operationalized as only task-related versus task-unrelated thought. By asking our participants not only if their thought was related to the task, but also which task they were thinking about specifically, we gathered a more nuanced sample of participant focus during the experiment. Together, these results suggest that what participants choose to focus on may strongly affect working memory performance, but that long-term memory may be less susceptible to these variations. However, an alternative is also

possible, that the type of thought does influence long-term memory and our recognition measure was not sensitive enough to capture subtle differences. Future research should consider the various aspects of a task that may draw participant focus when studying the effects of mind-wandering on task performance and the influence on different outcome measures. While the present study demonstrates that individuals engage in tasks differently, the reasons underlying this variation remain unclear.

Notably, we also found that there was a small increase in fully off-task thinking (thinking about “something else”) during the slow span task compared to the simple span and complex span tasks. One potential explanation for the increased free time benefit found in Souza and Oberauer (2017) is that during a slow span task, participants engaging in mind-wandering, displacing memory items. They must then intermittently retrieve the memory items back into the focus of attention, resulting in a pattern of results similar to that of the McCabe effect. However, given that the increase in this mind-wandering was quite modest in the present study (15% in slow span compared to 10% in simple span and 13% in complex span), it is likely that any long-term memory benefit would be extremely small as well. This contrasts with displacement of memory items in the complex span task, which necessitates shifting the focus of attention to the secondary task after each memory item on every trial. Further, as depicted in Fig. 5c, delayed memory performance on the slow span trials in which participants reported thinking about “something else” was not better than the other task types. Thus, if participants are engaging in some form of retrieval during working memory processing via mind-wandering, it does not appear to benefit long-term memory in the same way that a typical complex span task does.

The present results have important impacts on at least three fronts. First, our results indicate that retrieval of displaced information back into the focus of attention is critical as a mechanism underpinning the McCabe effect. When participants engaged less with the secondary task, the McCabe effect decreased despite more free time being available during initial processing. Rather than increased free time as found in Souza and Oberauer (2017), engaging with secondary task performance drove the effect of the long-term benefit of working memory processing.

Second, the nature of our sample also demonstrates the importance of setting for understanding human cognition. While many traditional psychology experiments take place in a controlled laboratory setting, most real-world cognitive tasks do not. Further, recent shifts to online setting in education and work underscore the importance of understanding how cognitive performance changes across settings. Prior research has found that many basic attention and choice paradigms show similar patterns of performance both in-person and online (Barnhoorn et al., 2015; Crump

et al., 2013; Hilbig, 2016; Semmelmann & Weigelt, 2017). The present work differs from these basic studies in that it involves more complex cognitive strategies for task completion, including task and effort tradeoffs between concurrent tasks. It seems that setting effects can change how some participants allocate their effort and attention across tasks, leading to different patterns of performance compared to laboratory studies. Other studies have explored this question in working memory (Bui et al., 2015; Ruiz et al., 2019; Uittenhove et al., 2022), but generally ignore any metrics beyond primary task accuracy. This is a critical missed opportunity as we demonstrate that individuals vary widely in their approach to the task when allowed to choose their own level of task engagement outside the lab. These differences have long-term consequences, at least for memory retention. The present findings imply that cognition in the service of remote self-monitored work and education may often be far from optimal even when initial tests of performance on the primary task metric indicate sufficient performance. Future research should explore these effects in other types of complex cognitive tasks to obtain a more complete picture of changes in cognitive engagement across different settings.

Third, in contextualizing these findings in the broader landscape of science including the replication crisis in psychology and open science initiatives, these data clearly show the importance of analytic decisions. Specifically, these data show strikingly different performance patterns depending on how the data were treated, with evidence swinging in favor of or against certain hypotheses, depending on how the data were preprocessed (Gelman & Loken, 2014). If we had collapsed across the entire sample our main conclusions would have likely landed in the file drawer with no evidence for the McCabe effect due to factors unknown. Alternatively, if we had excluded a large proportion of our sample based on the $\leq 80\%$ low performance exclusion rule and just reported high performers, we would have concluded the McCabe effect was robust. Clearly, what may seem like a small decision based on good intentions and sound logic (that is, inferring low performance as an indicator of participants not engaged on the task and that they should be excluded) has quite a profound downstream theoretical impact. Investigators should consider these implications in their own research and what it may mean for different models of cognition. In online studies where collection of large samples is relatively quick and easy, researchers should consider presenting supplemental analyses of the full sample broken by performance, attention checks, or some alternative performance measures. This may provide unique insights into human behavior generally, and specific insight into the boundary conditions of different theories of human cognition.

One potential criticism of the present study may be that due to the online setting, our participants were not fully

engaged with the experiment, resulting in poor quality data. However, we view this performance variability as a strength. Though some of our participants performed poorly on some components of the experiment, performance on both immediate-recall and delayed-recognition tasks was still well above chance. This finding suggests that the participants were not fully disengaged from all components of the experiment. We intentionally designed our experiment to leverage the variability in participant engagement to test the importance of full engagement with the complex span secondary task on delayed memory performance. We used the unengaged participants to demonstrate that shifting attention to the secondary task is what drives the McCabe Effect. While the field of psychology has begun to acknowledge the importance of broadening the diversity of research samples (Apicella et al., 2020; Henrich et al., 2010), there is room to consider variation even within a typical sample. Using a typical 80% accuracy threshold would result in the exclusion of half of our online sample. Such an approach assumes that so-called “bad” participants have little to offer our understanding of human cognition. Outside of the lab, circumstances are rarely ideal, and everyone is a “bad” participant sometimes. Understanding how a variety of people process information across a variety of contexts is important for broadening cognitive models to reflect how people function in daily life.

Another criticism of the present findings could be that we cannot know what cognitive process led to poor secondary task performance and so we cannot interpret the meaning of the data in our low and zero engagement groups. In complex span tasks, memory researchers typically only include data from participants that are performing at or above 80% accuracy under the assumption that we cannot know what participants are doing when they get secondary task performance wrong. This caution should certainly apply here in our low engagement participants, those with secondary task performance greater than 0% but less than 80% correct. A broad mix of cognitive processes could occur in any given participant to reach this behavioral result. The same issue is not at play in the zero-engagement group for the following reason. Chance performance is 50% in the task, but there was no requirement forcing participants to enter any response during the secondary task to proceed with the experiment. They could omit responses when they did not engage in the task. The only process that leads to achieving 0% accuracy in the secondary task is to simply not do the task at all. If a participant engages with the task in any way, even randomly pressing buttons on some trials, then performance will be above 0% due to a 50% chance each response is correct. Compared to the typical requirement that participants perform above 80% for data inclusion, we can be more confident in the present results because there is more than one way to achieve high performance in the task, but only one way to achieve 0% performance.

Conclusion

The present study demonstrated that dual tasking while maintaining information in working memory is beneficial to long-term retention because of repeated retrieval of information displaced from the focus of attention. We found that only participants who fully engaged with both components of a dual task exhibited an improvement in delayed memory performance for items initially encountered in a complex span task compared to a simple span task. Individuals differed in what they reported thinking about during the task, depending on both the span task and secondary task performance. These results emphasize that participants employ different strategies to complete a task based on task difficulty, motivation, and likely other factors. This seems especially true in self-monitored online contexts and has implications for how we assess performance across a variety of cognitive tasks. The present work provides further evidence of the importance working memory processing for long-term memory but also underscores the need to consider a wide variety of metrics when assessing cognitive performance, particularly in online samples.

Open Practices Statement Data and analyses scripts are openly available at the project's Open Science Framework page (<https://osf.io/e3fqu/>). None of the experiments were preregistered.

Declarations

Conflicts of interest We have no conflicts of interest to disclose.

References

- Apicella, C., Norenzayan, A., & Henrich, J. (2020). Beyond WEIRD: A review of the last decade and a look ahead to the global laboratory of the future. *Evolution and Human Behavior, 41*(5), 319–329. <https://doi.org/10.1016/j.evolhumbehav.2020.07.015>
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89–195). Elsevier.
- Baddeley, A., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47–89). Elsevier.
- Barnhoorn, J. S., Haasnoot, E., Bocanegra, B. R., & van Steenbergen, H. (2015). QRTEngine: An easy solution for running online reaction time experiments using Qualtrics. *Behavior Research Methods, 47*(4), 918–929. <https://doi.org/10.3758/s13428-014-0530-7>
- Bartsch, L. M., Singmann, H., & Oberauer, K. (2018). The effects of refreshing and elaboration on working memory performance, and their contributions to long-term memory formation. *Memory & Cognition, 46*(5), 796–808.
- Bartsch, L. M., Loaiza, V. M., Jäncke, L., Oberauer, K., & Lewis-Peacock, J. A. (2019). Dissociating refreshing and elaboration and their impacts on memory. *NeuroImage, 199*, 585–597.
- Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*(4), 822. <https://doi.org/10.1037/0022-3514.84.4.822>
- Bui, D. C., Myerson, J., & Hale, S. (2015). Age-related slowing in online samples. *The Psychological Record, 65*(4), 649–655. <https://doi.org/10.1007/s40732-015-0135-2>
- Camos, V., & Portrat, S. (2015). The impact of cognitive load on delayed recall. *Psychonomic Bulletin & Review, 22*(4), 1029–1034. <https://doi.org/10.3758/s13423-014-0772-5>
- Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A, 33*(4), 497–505. <https://doi.org/10.1080/146407481084008>
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769–786.
- Cotton, K., & Ricker, T. J. (2021). Working memory consolidation improves long-term memory recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(2), 208–219.
- Cotton, K., Sandry, J., & Ricker, T. (2023). *The Effects of Mind-Wandering, Cognitive Load and Task Engagement on Working Memory Performance in Remote Online Experiments*. PsyArXiv.
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*(2), 163.
- Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a Tool for Experimental Behavioral Research. *PLOS ONE, 8*(3), e57410.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450–466.
- Gelman, A., & Loken, E. (2014). The statistical crisis in science data-dependent analysis—A “garden of forking paths”—Explains why many statistically significant comparisons don't hold up. *American Scientist, 102*(6), 460.
- Hartshorne, J. K., & Makovski, T. (2019). The effect of working memory maintenance on long-term memory. *Memory & Cognition, 47*(4), 749–763.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Hilbig, B. E. (2016). Reaction time effects in lab-versus Web-based research: Experimental evidence. *Behavior Research Methods, 48*(4), 1718–1724. <https://doi.org/10.3758/s13428-015-0678-9>
- Jarjat, G., Hoareau, V., Plancher, G., Hot, P., Lemaire, B., & Portrat, S. (2018). What makes working memory traces stable over time? *Annals of the New York Academy of Sciences, 1424*(1), 149–160. <https://doi.org/10.1111/nyas.13668>
- Krimsky, M., Forster, D. E., Llabre, M. M., & Jha, A. P. (2017). The influence of time on task on mind wandering and visual working memory. *Cognition, 169*, 84–90.
- Loaiza, V. M., & Halse, S. C. (2018). Where Working Memory Meets Long-Term Memory: The Interplay of List Length and Distractors on Memory Performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Loaiza, V. M., & Lavilla, E. T. (2021). Elaborative strategies contribute to the long-term benefits of time in working memory. *Journal of Memory and Language, 117*, 104205. <https://doi.org/10.1016/j.jml.2020.104205>
- Loaiza, V. M., & McCabe, D. P. (2012). Temporal–contextual processing in working memory: Evidence from delayed cued recall and delayed free recall tests. *Memory & Cognition, 40*(2), 191–203.
- Loaiza, V. M., & McCabe, D. P. (2013). 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.

- Loaiza, V. M., Doherty, C., & Howlett, P. (2021). The long-term consequences of retrieval demands during working memory. *Memory & Cognition*, *49*(1), 112–126.
- 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.
- Mooneyham, B. W., & Schooler, J. W. (2013). The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *67*(1), 11–18.
- Morey, R. D., & Rouder, J. N. (2018). *BayesFactor: Computation of bayes factors for common designs*. R package version 0.9.12-4.2. <https://CRAN.R-project.org/package=BayesFactor>
- Mrazek, M. D., Smallwood, J., Franklin, M. S., Chin, J. M., Baird, B., & Schooler, J. W. (2012). The role of mind-wandering in measurements of general aptitude. *Journal of Experimental Psychology: General*, *141*(4), 788. <https://doi.org/10.1037/a0027968>
- Norris, D. (2017). Short-term memory and long-term memory are still different. *Psychological Bulletin*, *143*(9), 992–1009.
- Norris, D. (2019). Even an activated long-term memory system still needs a separate short-term store: A reply to Cowan (2019). *Psychological Bulletin*, *145*(8), 848–853. <https://doi.org/10.1037/bul0000204>
- Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ricker, T. J. (2015). The role of short-term consolidation in memory persistence. *AIMS Neuroscience*, *2*(4), 259–279.
- 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. <https://doi.org/10.3758/s13421-014-0398-x>
- 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. <https://doi.org/10.3758/PBR.16.2.225>
- 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.
- Ruiz, S., Chen, X., Rebuschat, P., & Meurers, D. (2019). Measuring individual differences in cognitive abilities in the lab and on the web. *PLOS ONE*, *14*(12), e0226217.
- Sandry, J., Zuppichini, M. D., & Ricker, T. J. (2020). Attentional flexibility and prioritization improves long-term memory. *Acta Psychologica*, *208*, 103104. <https://doi.org/10.1016/j.actpsy.2020.103104>
- Semmelmann, K., & Weigelt, S. (2017). Online psychophysics: Reaction time effects in cognitive experiments. *Behavior Research Methods*, *49*(4), 1241–1260. <https://doi.org/10.3758/s13428-016-0783-4>
- Souza, A. S., & Oberauer, K. (2017). Time to process information in working memory improves episodic memory. *Journal of Memory and Language*, *96*, 155–167.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154. [https://doi.org/10.1016/0749-596X\(89\)90040-5](https://doi.org/10.1016/0749-596X(89)90040-5)
- Uittenhove, K., Jeanneret, S., & Vergauwe, E. (2022). *From lab-based to web-based behavioural research: Who you test is more important than how you test*. PsyArXiv. <https://doi.org/10.31234/osf.io/uy4kb>
- Unsworth, N., & Engle, R. (2007). The Nature of Individual Differences in Working Memory Capacity: Active Maintenance in Primary Memory and Controlled Search From Secondary Memory. *Psychological Review*, *114*(1), 104–132.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.