Article:

Loaiza, V.M. and Borovanska, B.M. (2018) Covert retrieval in working memory impacts the phenomenological characteristics remembered during episodic memory. Consciousness and Cognition, 57. pp. 20-32. ISSN: 1053-8100

https://doi.org/10.1016/j.concog.2017.11.002

Covert Retrieval in Working Memory Impacts the Phenomenological Characteristics Remembered During Episodic Memory

Vanessa M. Loaiza

Borislava M. Borovanska

Word count: 8,439 words (excluding references)

Author Note

Vanessa M. Loaiza, Department of Psychology, University of Essex. Borislava M. Borovanska, Department of Psychology, King's College London.

Correspondence concerning this article should be addressed to Vanessa M. Loaiza, Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. Email: v.loaiza@essex.ac.uk

The data and analysis scripts are available on the Open Science Framework at:

https://osf.io/mu6an/

Abstract

Word count: 150

Much research has investigated the qualitative experience of retrieving events from episodic memory (EM). The present study investigated whether covert retrieval in WM increases the phenomenological characteristics that participants find memorable in EM using tasks that distract attention from the maintenance of memoranda (i.e., complex span; Experiment 1) relative to tasks that do not (i.e., short or long list lengths of simple span; Experiments 1 and 2). Participants rated the quality of the phonological, semantic, and temporal-contextual characteristics remembered during a delayed memory characteristics questionnaire (MCQ). Whereas an advantage of the complex over simple span items was observed for each characteristic (Experiment 1), no such difference was observed between short and long trials of simple span (Experiment 2). These results are consistent with the view that covert retrieval in WM promotes content-context bindings that are later accessible from EM for both objective performance and subjective details of the remembered information.

Keywords: working memory, episodic memory, memory characteristics

Retrieval from episodic memory (EM) is often accompanied by rich phenomenological details during the experience of mentally reliving the original event. Much research has explored the underlying processes that give rise to this experience of conscious recollection (Gardiner, 1988; Tulving, 1985), especially the conditions in which the remembered information was originally processed in working memory (WM). WM refers to the immediate memory system that maintains, updates, and manipulates information for brief intervals of time in the service of ongoing cognition. WM and EM are often strongly aligned both conceptually and empirically, as much work has shown that performance on their respective measures is significantly correlated (Unsworth, 2010; Unsworth & Spillers, 2010), especially recollection-based EM (Unsworth & Brewer, 2009). Some recent work has suggested that mechanisms underlying active maintenance of information in WM may not only promote later retrieval from EM, but also the experience of conscious recollection of the original event (Loaiza, Duperreault, Rhodes, & McCabe, 2015). In particular, we have proposed that measures of WM often entail the consistent covert retrieval of memoranda in order to keep them available despite other distracting events (Loaiza & McCabe, 2012; McCabe, 2008)[1]. Consequently, we have observed improvements in objective EM performance (e.g., free recall) as a function of opportunities to covertly retrieve memoranda in WM. The experiments in the current paper sought to determine whether subjective details associated with an event (i.e., phonological, semantic, and temporal-contextual) are likewise more phenomenologically memorable as a function of covert retrieval in WM.

Complex span tasks are often used to measure WM capacity, or the degree to which a person can maintain and manipulate information effectively. For example, the operation span

[1] It should be noted that in previous work we had specified covert retrieval as attentional refreshing (e.g., Loaiza & McCabe, 2012). However, in order to avoid any conflation of terms with other researchers using similar terminology but perhaps referring to different functions (e.g., Camos, Lagner, & Barrouillet, 2009; Johnson, 1992), we will use covert retrieval here. Much work remains to be done regarding how similar these proposed mechanisms are.

task (Turner & Engle, 1989) requires participants to solve basic arithmetic problems (e.g., 4 x 7 = 29?) that are interspersed among to-be-remembered information (e.g., concrete words). Simple span tasks such as word span instead successively present memoranda without any distraction or processing activity. Historically, complex span tasks were thought to measure WM capacity to a greater extent than simple span tasks due to the additional demands of the secondary task, whereas simple span tasks may only test passive storage of information. This view was largely predicated on the finding that complex span tasks predict performance on other measures of higher-order cognition, such as fluid intelligence, more strongly than simple span tasks (Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004).

However, further work has suggested that longer trial lengths of simple span tasks (i.e., more than four memoranda) may also reliably measure WM capacity (Unsworth & Engle, 2006, 2007a, 2007b). In their dual-component model, Unsworth and Engle (2007a, 2007b) proposed that WM capacity reflects the contributions of active maintenance of a limited amount of information in primary memory and the cue-based search and retrieval of secondary memory. In particular, attention must be devoted to sustaining the activation of memoranda or task goals in primary memory, and switching attention away from their maintenance to distraction or new incoming memoranda requires their retrieval from secondary memory. Accordingly, simple span tasks that present many more items than can be maintained in primary memory may be more similar to complex span tasks than previously considered to the extent that they also reflect the contribution of these two underlying functions. In support of this notion, Unsworth and Engle (2006) found that performance on simple span tasks resembled that of complex span tasks when the simple span tasks were more difficult in nature. Specifically, longer list lengths of simple span were more strongly correlated with fluid intelligence than shorter list lengths of simple span (Unsworth & Engle, 2006; Unsworth & Engle, 2007b). Using latent variable analysis, Unsworth and Engle (2006) also showed that the variability common to these same long list lengths of simple span loaded on a different factor than short list lengths of the same simple span task. This suggests that the distinctions typically made between simple and complex span tasks are not as straightforward as originally thought. Instead, Unsworth and Engle have argued that both simple and complex span tasks measure the same mechanisms, but to different degrees. Complex span and long simple span trials require retrieval from primary and secondary memory on each trial, whereas short simple span trials only require retrieval from primary memory. Thus, a cue-dependent search of secondary memory is necessary to retrieve the memoranda that have been displaced from active maintenance in primary memory due to distraction (complex span tasks) or new incoming memoranda (long simple span tasks).

McCabe (2008) further investigated this hypothesis by considering immediate and delayed retrieval from simple and complex span tasks. Specifically, McCabe administered trials of word span and operation span with two to four memoranda per trial. In addition to immediately recalling the words at the end of the trials, participants were also asked to try to recall the words after a delay. The results of his first experiment showed that while immediate recall of memoranda from simple span was predictably greater than that of complex span, the reverse was true for delayed recall, such that memoranda from complex span were more likely to be recalled than simple span. We henceforth refer to this finding of greater EM performance for complex span than simple span as *the McCabe effect*. The McCabe effect was also demonstrated even when immediate recall was precluded randomly for half the trials, thereby negating any possible differences in overt retrieval during the recall phase of the trials.

McCabe (2008) originally interpreted these findings in accordance with the dual-component model: given that the distraction phase of complex span tasks (e.g., arithmetic problems) displaces the memoranda from primary to secondary memory, participants must engage in covert retrieval to reactivate them so they are not forgotten by the end of the trial. This repeated covert retrieval from secondary memory during the interim between the distraction (i.e., the arithmetic problems) and the memoranda (i.e., the words) of complex span tasks in turn promotes strong cues for those items to be later accessed during delayed recall. Conversely, the simple span trials never exceeded the limits of primary memory (i.e., about four memoranda), and thus those items should remain within primary memory without being displaced. Accordingly, no controlled search of secondary memory is necessary for short list lengths of simple span, whereas complex span tasks necessitate covert retrieval that in turn promotes stronger retrieval cues to access that information later on during EM (McCabe, 2008). The most important evidence that supported the covert retrieval account was the finding that delayed recall declined as a function of serial position for the complex span but not the simple span trials. This is in line with the prediction that the memoranda presented earliest in the trials, which presumably had the most opportunities to be covertly retrieved during WM, were likewise the most likely to be recalled from EM.

Loaiza and McCabe (2012) further investigated the tacit prediction that long list lengths of simple span may also require covert retrieval because the earliest presented items should have been displaced from primary memory by new incoming memoranda within the same trial. That is, if long list lengths of simple span (i.e., eight words to recall) also require covert retrieval to keep the memoranda active in WM, then they should exhibit a similar McCabe effect as the complex span items in EM. Thus, both complex span and long trials of simple span should exhibit greater recall than short trials of simple span. However, Loaiza and McCabe showed that this was not the case in either delayed free or cued recall: there was no McCabe effect for the long list lengths of simple span, even the first four memoranda that are presumably displaced from and must be retrieved back into primary memory. This result conflicted with the suggestion that long simple span and complex span trials are similar in their requirement of covert retrieval to sustain the activation of the memoranda in WM. Instead, the distraction during complex span may serve as a unique opportunity to covertly retrieve the memoranda back into conscious awareness, and such opportunities are not available during simple span regardless of the list length. Loaiza and McCabe further investigated the use of internally-generated and externally-provided temporal-contextual cues under the notion that opportunities to covertly retrieve memoranda during complex span encourage the binding between the content of the memoranda and its context within the trial. Consistent with their predictions, Loaiza and McCabe showed that participants were more likely to make use of temporal associations between the memoranda originally studied during complex span relative to simple span tasks of any list length. This suggested that covert retrieval in WM promotes temporal-contextual processing such that it is particularly important to reinforcing content-context bindings that are later accessed during EM.

We have further investigated the possibility that covert retrieval in WM is particularly important for the promotion of content-context bindings that may give rise to conscious recollection in EM. Following a block of simple and complex span trials, we administered a delayed remember/know recognition test that asked participants to reflect on the subjective experience that accompanied any tested items that they detected as old (Loaiza et al., 2015). These subjective reports are used to approximate the relative contributions of recollection (i.e., retrieval of specific, contextual details of an event) and familiarity (i.e., recognition in the absence of specific details) to their memory performance (see Yonelinas, 2002 for a review). As is typical in the remember/know paradigm (Gardiner, 1988; Tulving, 1985), participants were asked to decide whether they could consciously recollect specific, contextual details from when they had studied the word (i.e., *remember*) or if they "just knew" the word was presented in the absence of any specific details (i.e., *know*). Consistent with the notion that covert retrieval facilitates content-context binding in WM, the results indicated that the McCabe effect was evident for remember but not know responses. Moreover, just as in the delayed recall tests, remember responses declined as a function of serial position for complex span, but serial position had little effect for remember responses for simple span and know responses for both trial types. This suggests that the subjective experience of conscious recollection in which details from the original event can be retrieved and relived may be influenced by the extent to which the information is subject to the brief reactivations in WM that covert retrieval affords.

Although these results collectively suggest that covert retrieval is not only important to objective EM performance, but also facilitates the retrieval of subjective details, much more work is necessary in order to substantiate this proposal. For example, we have not yet inquired about the kinds of subjective details that participants can remember about the memoranda. Memory characteristics questionnaires (MCQs; Johnson, Foley, Suengas, & Raye, 1988) provide a meaningful method of ascertaining the kinds of characteristics that are remembered for different kinds of events. During a typical MCQ, participants rate the quality of their memory for different characteristics of recognized information (e.g., auditory or perceptual detail, associations, feelings/reactions, etc.). Sometimes in conjunction with or in comparison to the remember/know paradigm, MCQs have often been used to dissociate the characteristics remembered of veridical and illusory memories (Mather, Henkel, & Johnson, 1997; Neuschatz, Payne, Lampinen, & Toglia, 2001; Norman & Schacter, 1997). For example, participants often report “remembering” semantically related but non-presented lures during a remember/know paradigm, sometimes as frequently as the presented memoranda (Payne, Elie, Blackwell, & Neuschatz, 1996; Roediger & McDermott, 1995). Research with MCQs has elucidated that the underlying characteristics distinguish these true and false recollections, such that presented memoranda are more likely to be endorsed with regard to their perceptual detail and associated thoughts/feelings than non-presented lures

(Mather et al., 1997; Neuschatz et al., 2001; Norman & Schacter, 1997) and imagined events (Johnson et al., 1988). Thus, MCQs have helped to inform the theoretical understanding of the underlying characteristics of memories of false or imagined events, especially those that underlie false recollection.

Likewise, the use of MCQs to investigate the McCabe effect may also facilitate the theoretical understanding of how covert retrieval improves long-term retention, and in particular, subjective recollection. That is, the overall greater recollection of memoranda studied during complex span versus simple span (Loaiza et al., 2015) may be attributable to specific underlying characteristics that are phenomenologically more memorable than others. Given our previous work (Loaiza & McCabe, 2012), it may be expected that complex span tasks that promote covert retrieval during WM likewise increase the likelihood of recalling temporal-contextual details in particular compared to other details, such as semantic or phonological information. Moreover, because covert retrieval is thought to occur exclusively after the processing phases of the complex span tasks (McCabe, 2008), the opportunity to immediately recall the memoranda should not differentially affect the accessibility of these details. Such results provide further insight into the importance of covert retrieval in WM for the subjective experience of remembering during EM.

Current Study

The current study explored whether the aforementioned McCabe effect is also evident for the subjective characteristics of memoranda that were once maintained in WM. In particular, we examined whether such characteristics are differently emphasized during tasks that presumably encourage covert retrieval of memoranda (e.g., complex span tasks like operation span) relative to tasks that do not (e.g., simple span tasks of short or long list lengths like word span). We tested this by administering a delayed MCQ that inquired about the quality of the phonological, semantic, and temporal-contextual characteristics remembered for memoranda originally studied during operation span and word span (Experiment 1) or short and long trials of word span (Experiment 2). Given our aforementioned studies suggesting that covert retrieval promotes content-context bindings (Loaiza et al., 2015; Loaiza & McCabe, 2012), we were interested in whether temporal-contextual characteristics would be particularly sensitive to covert retrieval compared to phonological and semantic characteristics. Furthermore, we explored whether the impact of immediate recall from WM may moderate this effect by having participants immediately recall the memoranda or not for half of the trials in each experiment.

Consistent with our prior work, we expected that qualitative characteristics of the memoranda, especially temporal-contextual characteristics, would be sensitive to tasks that require covert retrieval (i.e., operation span) versus tasks that do not (i.e., short or long list lengths of word span). This would be evident in an overall increase in the memorability of memoranda studied during operation span relative to short and long list lengths of word span, particularly for earlier serial positions of the operation span trials. Conversely, the memorability of word span memoranda should not vary as a function of serial position regardless of the trial length.[2] Such findings would be consistent with the suggestion that covert retrieval facilitates content-context binding in WM, and that these bindings are evident not only in objective retrieval but also subjective, phenomenological memorability of the memoranda (Loaiza et al., 2015). Furthermore, although there may likely be an overall benefit of immediate recall on the characteristics' memorability, the benefit of operation span over word span should be consistent regardless of the original immediate recall condition or even having successfully recalled the memoranda in the first place. That is, covert retrieval in

[2] It should be noted that this serial position investigation was considered after the initial design of the study, and thus in the interest of transparency we wish to qualify it as technically exploratory. However, it is not inconsistent with our prior work wherein we have extensively investigated the pattern of performance across serial position as evidence for the importance of the original context of the studied memoranda during complex span tasks (e.g., Loaiza et al., 2015; Loaiza & McCabe, 2012; McCabe, 2008).

WM should improve the subjective characteristics that are remembered even if participants were not given the opportunity to recall the memoranda or if they did not successfully recall the memoranda (i.e., correcting for initial recall).[3]

Experiment 1

Method

Participants and Design. Twenty-four participants were recruited from the subject pool at Colorado State University in exchange for partial course credit. Participants were young adults with normal or corrected-to-normal vision. All participants in both experiments gave informed consent and were fully debriefed at the conclusion of the experiment.

The independent variables of trial type (word span vs. operation span) and immediate recall (immediate recall vs. no immediate recall) were manipulated within-subjects. The principal dependent variable was the ratings on the delayed MCQ. We also report on immediate recall (i.e., serial and free recall scoring).

Materials and Procedure. The memoranda for both experiments were 144 concrete, high frequency nouns (letters: M = 5.22, SD = 1.33, range = 3 – 9; syllables: M = 1.42, SD = 0.60, range = 1 – 4; log HAL frequency: M = 10.49, SD = 0.82, range = 7.42 – 12.67; Balota et al., 2007). The memoranda were randomly arranged and counterbalanced across the trial type and immediate recall conditions.

An initial arithmetic task contained similar single-digit multiplication problems (e.g., 7 x 4 = 28?) that were to appear during the operation span task. The arithmetic task was given to participants in order to familiarize them with the processing component used in the operation span task. Participants were instructed to read each problem aloud and verify whether the answer given was true or false. The experimenter advanced the screen once the

[3] The correction for initial recall was considered after planning the study given recent research encouraging the use of the correction to ensure that any patterns in delayed performance are not simply an artifact of differential rates of initial recall across conditions (Rose, Buschsbaum, & Craik, 2014).

response was given. Numbers 4 – 9 were used in various combinations to compose the problems. Afterward, participants also practiced the digit matching task to familiarize themselves with another element that was to be included on the operation span and word span trials that did not require immediate recall. In this task, participants were presented with two-digit numbers (e.g., 64, 27, 53) and verified whether both digits were even or not. If both digits were even, participants were instructed to respond “yes” aloud. If both digits were odd or one was even and the other was odd, participants were instructed to respond “no” aloud.

Following the practice tasks, participants completed three blocks of word span and operation span trials that were each followed by a delayed characteristics questionnaire. Each block comprised 12 randomly presented word span and operation span trials (six of each type). During the word span trials, four memoranda appeared successively for 1 s each. Participants were instructed to read them silently and try to remember them. During the operation span trials, the four memoranda were each preceded by an arithmetic problem presented for 3.5 s that participants read and solved aloud as previously described. At the end of half of the word span and operation span trials, participants were cued to try to immediately recall the memoranda in their original order of appearance (i.e., serial recall). For the other half of the trials, participants responded “yes” or “no” to five double-digit numbers successively presented for 2 s each as they had done earlier during the digit matching task. Accordingly, there were three trials of each condition per block.

At the end of each block, participants engaged in a distracter-filled delay activity for approximately 2 min (i.e., questionnaires, crosswords). Afterward, all 48 of the memoranda from the previous block were randomly re-presented one at a time during the three-question MCQ. Participants were instructed to rate each word with regard to what they could remember in terms of its phonological, semantic, and temporal-contextual characteristics from when the words had been presented during the original task. Specifically, participants responded regarding each item's physical characteristics (i.e., "How much can you remember about the PHYSICAL CHARACTERISTICS (e.g., what the word sounded like, what the word looked like)?"), the meaning of the word (i.e., "How much can you remember about the MEANING (e.g., the definition of the word, your emotional reaction to the word, related words the word made you think of)?"), and its original place on the list (i.e., "How much can you remember about WHERE IT WAS ON THE STUDY LIST (e.g., the word that came before or after it, whether it was at the beginning or end of a trial?"). The order of the presentation of the questions was random, and participants responded to each question on a 1 – 9 Likert scale. The full scale was shown on the screen with the question, with the words "very little," "some," and "very much" presented respectively underneath points 1, 5, and 9 on the scale.

Results and Discussion

The data and analysis scripts for both experiments are available on the Open Science Framework at osf.io/mu6an. The results of both experiments were primarily analyzed with Bayesian *t*-tests (Rouder, Speckman, Sun, Morey, & Iverson, 2009) and Bayesian analysis of variance (BANVOA; Rouder, Morey, Speckman, & Province, 2012) with the independent variables (i.e., trial type, immediate recall condition) as fixed effects and participant as a random effect using the BayesFactor package with its default settings (e.g., the "medium" scale factor for the prior) in R (Morey & Rouder, 2015). Bayesian inferential statistics allow for the comparison of the data given one model (e.g., the null model assuming only a random effect of participant, M_0) to that of another model (e.g., an alternative model assuming an effect of trial type, M_1). The ratio of these likelihoods is the Bayes factor (BF) that expresses the relative evidence for the alternative model (BF_{10}) or the null model (BF_{01}).

Due to experimenter error, 1.07% of the data for the MCQ test were missing and therefore excluded from analysis. We first examined immediate recall performance according to trial type. Although immediate serial recall (ISR) was instructed to the participants, we also scored participants' recall without regard to the original serial order of the memoranda (i.e., immediate free recall, IFR) in both experiments. As expected, there was overwhelming evidence that participants' ISR and IFR performance was greater for word span compared to operation span trials (see Table 1). This is consistent with a great deal of research indicating that complex span tasks like operation span impair immediate recall to a greater degree than simple span tasks like word span (e.g., McCabe, 2008; Unsworth & Engle, 2006).

The principal analyses concerned the ratings given during the MCQ of the originally studied memoranda (see Figure 1). The ratings given to the three characteristic types (phonological, semantic, and temporal-contextual) were each submitted to a 2 (trial type: word span, operation span) x 2 (recall condition: immediate recall, no immediate recall) repeated measures BANOVA. The results of the analyses indicated overwhelming evidence for main effects of trial type and immediate recall condition for each characteristic relative to the null model (see Table 2), such that operation span memoranda were more phenomenologically memorable overall than word span memoranda. This finding replicates the McCabe effect that has been demonstrated in objective measures of EM (Loaiza et al., 2015; Loaiza & McCabe, 2012; McCabe, 2008). Furthermore, memoranda that were immediately recalled were also rated as more memorable than memoranda that were not immediately recalled. Importantly, the main effects models for each characteristic except for phonological characteristics were at least substantially preferred to the other models, in particular the interaction model. Given this overall main effect of recall condition, we next averaged across that factor and used Bayesian estimation software (BEST; Kruschke, 2013) to estimate the size of the McCabe effect in order to assess its consistency across characteristic type. Although the effect size was numerically largest for temporal-contextual characteristics, the McCabe effect was consistent across the characteristics, evident in the overlap of the 95% highest-density intervals (HDI; i.e., the range of credible values of the effect size) between the characteristics: phonological *d* = 0.71 [0.24, 1.17], semantic *d* = 1.05 [0.52, 1.57], and temporal-contextual *d* = 1.22 [0.59, 1.85]. Finally, in order to account for the impact of initial recall when immediately recalling the words, we also assessed the ratings for each characteristic for only the memoranda that were correctly recalled during the immediate recall attempt (see Table 3). Once again, there was overwhelming evidence for a benefit of studying memoranda in the context of operation span versus word span trials even when correcting for initial recall. The Bayesian estimates of effect size also mirrored the full analyses, with a consistent McCabe effect across characteristic type.

We also considered the MCQ ratings as a function of serial position. To foreshadow, as in the overall analyses there was only an overall main effect of recall condition, and thus the results represented in Figure 2A are collapsed across recall condition for the sake of clarity. The ratings given to the three characteristic types were each submitted to a 2 (trial type: word span, operation span) x 2 (recall condition: immediate recall, no immediate recall) x 4 (serial position: 1, 2, 3, 4) repeated measures BANOVA. For the sake of brevity, we focus only on the best models. For the temporal-contextual characteristics, the best model included main effects of all three factors and a trial type x serial position interaction (BF_{10} = 2.27 x 10^{29}), and this simpler model was anecdotally preferred (BF = 2.33) to the next best model that further included an interaction term between recall condition and serial position (BF_{10} = 9.75 x 10^{28}). For the semantic characteristics, the best model only included main effects of the three factors (BF_{10} = 6.01 x 10^{20}), and this simpler model was anecdotally preferred (BF = 1.37) to the next best model that further included an interaction term between recall condition and serial position (BF_{10} = 4.40 x 10^{20}). Finally, for the phonological characteristics, the best model included main effects of all three factors and a trial type x serial position interaction

(BF_{10} = 1.50 x 10^{13}), and this model was anecdotally preferred (BF = 2.10) to the next best but simpler main effects model (BF_{10} = 7.16 x 10^{12}).

As before, in order to account for the impact of initial recall when immediately recalling the words, we also considered these serial position analyses for only the memoranda that were correctly recalled during the immediate recall attempt (see Figure 2B). In general, these results mirror and amplify the full analyses: once again, the best model for the temporal-contextual characteristics was a full model of main effects of and an interaction between trial type and serial position (BF_{10} = 1.89 x 10^{16}) that was overwhelmingly preferred (BF = 119) to the next best main effects only model (BF_{10} = 1.59 x 10^{14}). This full model was also the best for the semantic characteristics (BF_{10} = 1.32 x 10^{9}), but only anecdotally preferred (BF = 2.97) to the simpler main effects only model (BF_{10} = 4.42 x 10^{8}). Finally, the best model for the phonological characteristics was a main effect of only trial type (BF_{10} = 1.42 x 10^{5}), and this simpler model was strongly preferred (BF = 14.04) to the next best model of main effects of both trial type and serial position (BF_{10} = 10,146).

In summary, these results replicate and extend the McCabe effect to include the characteristics of the memoranda that are phenomenologically remembered during EM. Moreover, the results also replicate the previous findings that immediate recall boosts both retrieval and memorability of the memoranda overall but does not moderate the magnitude of the McCabe effect (i.e., McCabe, 2008; Experiment 3). Finally, the serial position analyses also replicated the finding that items presented at the beginning of the operation span trials are not only more likely to be recalled than later serial positions (McCabe, 2008), but are also rated as more memorable, particularly in terms of their temporal-contextual characteristics. The findings are in line with the suggestion that participants engage in covert retrieval during tasks that distract attention from the memoranda (e.g., operation span) relative to tasks in which all of the memoranda are presented without distraction (e.g., word span). This covert retrieval has important consequences not only for the ability to retrieve information during EM (Loaiza et al., 2015; Loaiza & McCabe, 2012), but also according to the qualitative characteristics (i.e., phonological, semantic, and temporal-contextual) about the memoranda that participants are able to retrieve.

In Experiment 2, we explored whether the McCabe effect is specific to tasks that involve distraction, like complex span, or if longer trial lengths of simple span tasks may also yield a benefit to the subjective memorability of memoranda in EM. We expected to replicate our previous findings regarding objective retrieval performance (Loaiza & McCabe, 2012), such that long simple span trials (i.e., eight memoranda) should not differ from short simple span (i.e., four memoranda), even when examining the first four memoranda of the long trials that are presumably displaced from primary memory. Moreover, correcting for the initial recall rates of the different list lengths of simple span should not change this pattern of results. Finally, analysis of the ratings across serial position should yield relatively flat slopes for both short and long trials, thereby providing converging evidence that covert retrieval during WM is specifically important to the subjective details remembered during EM.

Experiment 2

Method

Participants and Design. Thirty-six participants[4] were recruited from the subject pool at Colorado State University in exchange for partial course credit. Participants were young adults with normal or corrected-to-normal vision, and none had participated in the previous experiment.

The independent variables of trial type (list length 4 vs. 8 of word span) and immediate recall (immediate recall vs. no immediate recall) were manipulated within-

[4] Note that the increase in sample size from Experiment 1 is only a consequence of balancing the same number of participants across the increased number of counterbalances in Experiment 2.

subjects. The principal dependent variable was the ratings on the delayed MCQ. We also report on immediate recall (i.e., serial and free recall scoring).

Materials and Procedure. The materials and procedure were very similar to Experiment 1. The principal difference was that the participants instead studied the memoranda in trials of short or long list lengths of word span. There were four trials of each list length per block (three blocks total), with half of each randomly ending with either an immediate recall attempt or the digit matching task. As in Experiment 1, participants completed a MCQ following a distracter-filled delay at the end of each block wherein they reported on the phonological, semantic, and temporal-contextual characteristics they remembered on a scale of 1 to 9 for each individual item from the previous block.

Results and Discussion

Due to experimenter error, 0.08% of the data for the MCQ test were missing and therefore excluded from analysis. With regard to immediate recall performance, there was overwhelming evidence that participants were more likely to recall memoranda from shorter than longer list lengths of word span (see Table 1). This is consistent with a great deal of previous research indicating that increasing the list length reduces immediate recall performance (e.g., Ward, Tan, & Grenfell-Essam, 2010).

As in Experiment 2, the principal analyses concerned participants' subjective ratings to the re-presented memoranda according to their remembered phonological, semantic, and temporal-contextual characteristics (see Figure 2). These respective ratings were each submitted to a 2 (trial type: 4, 8) x 2 (recall condition: immediate recall, no immediate recall) repeated measures BANOVA (see Table 2). As mentioned previously, the first four memoranda of list length 8 trials have been presumed to be retrieved from outside of primary memory (Unsworth & Engle, 2007b), and thus we also compared recall of the first four memoranda of list length 8 trials to the list length 4 trials. We repeated these analyses for memoranda that were initially recalled during the immediate recall attempt in order to ensure the results were consistent even when accounting for differences in immediate recall performance (see Table 3). Finally, we also considered the MCQ ratings as a function of serial position.

When assessing the ratings given to all of the memoranda, there was overwhelming evidence for main effects of immediate recall condition for the semantic and temporal-contextual characteristics. The phonological characteristics showed strong evidence of both main effects of trial type and immediate recall condition. However, this model was not strongly preferred to a simpler model positing only a main effect of immediate recall condition. Thus, an immediate recall attempt improved the ratings of the characteristics later remembered overall (see Figure 3). The BEST estimation of the effect size comparing short and long trials collapsing across recall condition was consistent for each characteristic type, with each 95% HDI including 0: phonological d = -0.29 [-0.64, 0.06], semantic d = -0.24 [-0.60, 0.13], temporal-contextual d = -0.09 [-0.45, 0.27]. This pattern of results was similar when comparing the ratings given to the first four memoranda of the list length 8 relative to list length 4 trials (see Figure 4). The best models for semantic and temporal-contextual characteristics showed substantial and overwhelming evidence for a main effect of immediate recall condition, respectively. The effect of immediate recall condition was at least substantially preferred to the other models. For phonological characteristics, there was very little evidence for any model, with BFs in favor of the null ranging from 1.4 to 5 (i.e., ambiguous to substantial evidence in favor of null effects). When comparing trial types for memoranda that had been initially recalled (see Table 3), there was at least substantial evidence for a benefit of list length 8 trials relative to list length 4 trials for semantic and temporal-contextual characteristics. If the first four memoranda of list length 8 trials must be retrieved from secondary memory, then these effects should be even stronger when comparing the memorability of list length 4 trials relative to the first four memoranda of list length 8 trials. Critically, however, there was only ambiguous evidence in favor of greater ratings for these memoranda relative to list length 4 trials.

Finally, we considered the MCQ ratings across serial position, but with particular attention to comparing the short trials to the first four memoranda of the long trials for brevity and to address the specific theoretical issue regarding whether these particular items may follow the same pattern as complex span from Experiment 1. As in Experiment 1, the results are represented in Figure 5 collapsed across recall condition due to the lack of or simply an overall effect. The ratings given to the three characteristic types were each submitted to a 2 (trial type: 4, 8) x 2 (recall condition: immediate recall, no immediate recall) x 4 (serial position: 1, 2, 3, 4) repeated measures BANOVA (see Figure 2A). For the sake of brevity, we focus only on the best models. For the phonological characteristics, the null model was at least substantially preferred to the models positing any effect or interaction ($BFs_{01} \geq 5$). For the semantic characteristics, the best model included main effects of recall and serial position ($BF_{10} = 6.21$), but this model was not substantially preferred ($BF = 1.83$) to a simpler model with only a main effect of recall condition ($BF_{10} = 3.39$). Finally, for the temporal-contextual characteristics, the best model included main effects of recall and serial position ($BF_{10} = 1.76 \times 10^5$), and this model was substantially preferred ($BF = 7.51$) to the next best model that further included an interaction between recall and serial position ($BF = 23{,}467$). We also observed that correcting for initial recall did not change this pattern of results (see Figure 2B)[5]: the null effects model was still substantially preferred for phonological characteristics ($BFs_{01} \geq 5$), and there was only ambiguous evidence of an effect of trial type for semantic ($BF_{10} = 1.01$) and temporal-contextual ($BF_{10} = 2.91$) characteristics.

[5] Note that this analysis had a reduced number of observations due to lack of initial recall for some serial positions during the list length 8 trials, and thus these data were excluded (2.1% in total).

Overall, these results are consistent with previous research indicating that the McCabe effect is specific to tasks that comprise distraction of attention from maintenance of memoranda, such as complex span tasks, rather than increased trial lengths of simple span tasks (Loaiza & McCabe, 2012). Thus, just as is the case with objective retrieval of information, the phenomenological characteristics that are retained in EM are uniquely promoted during complex span tasks that require covert retrieval relative to simple span tasks of any list length.

General Discussion

The present experiments investigated the phenomenological characteristics that are accessible in EM as a function of having covertly retrieved the memoranda in WM. Importantly, the results suggest a distinction between complex span tasks and simple span tasks of any list length: items studied in the context of a complex span task were more phenomenologically memorable than those of a simple span task across all three of the measured characteristics (i.e., phonological, semantic, and temporal-contextual), whereas no such difference occurred between shorter and longer trials of simple span. This suggests a unique influence of the use of distraction during complex span to promote not only objective performance (Loaiza & McCabe, 2012; McCabe, 2008), but also the subjectively memorable characteristics of that retained information in EM. As we have argued previously, we attribute this beneficial effect for objective performance and subjective memorial experience to participants' increased opportunity to engage in covert retrieval to retain the memoranda in WM, in turn promoting the recollected details of the events in later EM (Loaiza et al., 2015).

Furthermore, these results were generally consistent across opportunity for immediate recall and characteristic type. First, participants' ratings were sensitive overall to the immediate recall condition, indicating that the self-report ratings are reliable insofar as they are susceptible to task characteristics in a predictable manner (i.e., an overall advantage of tested over non-tested items; Rowland, 2014). Importantly, however, whether the memoranda were immediately recalled or not (i.e., either via the manipulation of immediate recall in the experiment or the participants' actual immediate recall performance) did not change the overall pattern of these results. This further corroborates the covert retrieval account of the McCabe effect, suggesting that the locus of the benefit is specific to the encoding rather than the retrieval phase of complex span tasks. Moreover, while temporal-contextual cues have been noted as being particularly relevant to objective retention in EM (Loaiza & McCabe, 2012), the consistency of the McCabe effect in Experiment 1 across the different characteristics (i.e., phonological, semantic, and temporal-contextual) suggests that phenomenological experience does not distinguish across these aspects of an event episode. Interestingly, the interaction between trial type and serial position that has been observed in previous objective performance (Loaiza et al., 2015; McCabe, 2008) was most evident for temporal-contextual characteristics and to a much lesser extent or not at all for phonological and semantic characteristics. That is, the subjective memorability of the temporal-contextual characteristics was the most sensitive to the original serial position during complex span, and presumably the relative opportunities to covertly retrieve the memoranda in WM (McCabe, 2008). These results collectively suggest that multiple retrieved features may contribute to the recollective experience that has been associated with the McCabe effect (Loaiza et al., 2015). However, the content-context bindings that may be reinforced as a function of covert retrieval opportunities may result in subjective memorability that is more sensitive to those temporal associations than the other characteristics that accompany recollective experience. Thus, consistent with previous studies (Mather et al., 1997; Norman & Schacter, 1997), the novelty of these results further extend the utility of MCQs as a method of elucidating the underlying characteristics of recollective experience as a function of covert retrieval in WM.

It is important to emphasize that the MCQs administered in this study re-presented all of the memoranda, with no new stimuli, and thus there was very little retrieval demand on the participants compared to a more standard and objective measure of EM, such as recall or recognition. Even still, the McCabe effect was evident across characteristics, and its predicted sensitivity to the original serial positions for the memorability of the temporal-contextual characteristics in particular was even more striking. Thus, the robustness of this effect across multiple methods of measuring EM, both objective and subjective, indicates both a meaningful finding as well as a testament to the utility of subjective measures as providing informative insights for understanding human memory (Gardiner, 1988; Johnson et al., 1988; McCabe, Geraci, Boman, Sensenig, & Rhodes, 2011; Tulving, 1985). A great deal of research has investigated how self-reports of subjective memorial experience, as in the remember/know paradigm, are susceptible to a variety of theoretically meaningful factors, such as aging (McCabe, Roediger, McDaniel, & Balota, 2009) and source confusions during false recollection (McCabe & Geraci, 2009). Much of this work has helped to clarify that two processes of recollection and familiarity underlie episodic memory performance (see Yonelinas, 2002, for a review). However, the remember/know paradigm does not often ask participants to indicate the details that underlie their subjective experience. MCQs provide a meaningful method of isolating the characteristics that contribute to participants' memory performance. For example, research using MCQs has revealed that false recollections and imagined events are distinguishable from accurate recollections and real events in terms of the characteristics that participants find subjectively memorable (Johnson et al., 1988; Mather et al., 1997; Norman & Schacter, 1997).

Our research regarding the McCabe effect mirrors this literature: although we previously have found greater recollective experience for complex span over simple span memoranda (Loaiza et al., 2015), the underlying characteristics that contribute to this effect were not known. Such an investigation is theoretically meaningful for understanding whether certain characteristics (i.e., temporal-contextual) are especially relevant to the later recollection of information that was presumably covertly retrieved in WM. That is, if covert retrieval operates on content-context bindings during the interim of the presentation of the memoranda and the processing component, then temporal-contextual characteristics may be the most sensitive to covert retrieval compared to other characteristics and regardless of the opportunity to recall the memoranda. Accordingly, the current study using a MCQ makes a novel contribution to the literature by showing that, regardless of the opportunity for immediate recall, multiple characteristics (phonological, semantic, and temporal-contextual) contribute to the greater long-term retention and recollective experience that covert retrieval in WM affords. The study also extends the use of MCQs beyond discriminating real versus illusory or imagined events to also informing the characteristics that are recollected during EM as a function of covert retrieval in WM. Although phonological and semantic characteristics were largely used as a basis of comparison for temporal-contextual characteristics in the current study, it would be interesting to examine in future research whether the patterns reported here could be dissociated in meaningful ways. For example, an emphasis of the temporal-contextual characteristics may shift with variations of the stimuli in terms of their physical characteristics (e.g., presenting the words in different styles of font) or semantic characteristics (e.g., varying the concreteness or relatedness of the words).

The results also have implications for the dual-component model of WM (Unsworth & Engle, 2007a, 2007b). According to this framework, WM capacity reflects two components of active maintenance in primary memory and cue-driven search and retrieval of secondary memory. Thus, the model assumes that tasks that similarly tax both components, such as complex span tasks and simple span tasks of longer list lengths, are comparable measures of WM capacity and predictors of other higher-order cognition, such as fluid intelligence. Indeed, Unsworth and Engle have shown strong overlap between complex span and long trials of simple span in terms of the predictive utility of the tasks. Moreover, the detrimental effect of increasing list length on recall was largely similar when equating the number of items to retrieve from secondary memory between simple and complex span tasks (Unsworth & Engle, 2006). Accordingly, if covert retrieval functions by retrieving displaced information from secondary memory (McCabe, 2008), then a similar McCabe effect for complex span should be demonstrated for long simple span trials that likewise exceed the capacity of primary memory and thereby require retrieval from secondary memory. Congruent with our previous findings using objective measures of memory performance (Loaiza & McCabe, 2012), this was not the case: only the immediate recall condition affected ratings overall in Experiment 2, whereas there was no evidence of greater retention of long simple span trials compared to short trials, even when considering the first four items that were presumably displaced from primary memory. Thus, these results conflict with the notion that new incoming information in long simple span trials similarly distracts attention as the processing components of complex span trials.

It is important to note that although Unsworth and Engle (2006) observed that short and long list lengths of simple span loaded on distinct factors, long list lengths of simple span did not load on the same factor as complex span trials. Furthermore, a significant proportion of the variance was unique to long list lengths of simple span and complex span trials, suggesting that while both trial types significantly predict fluid intelligence, there is still a large degree of variance that is not shared between them. Thus, the two tasks may engage different types of processing that are differently important to other higher-order cognition, such as fluid intelligence. Consequently, it may not be the case that the tasks share some similar features, but differ at least with regard to retrieval of displaced information. For example, it may be the case that the shift in mental task set that the processing task necessarily engenders during complex span provides a more substantial shift in attention than an increased load from the same task set of maintaining memoranda during long trials of simple span. This may in turn differently emphasize the opportunity to covertly retrieve the previously presented memoranda between the two tasks. Models that assume that WM represents the activated content of long-term memory (e.g., Cowan, 1999; Oberauer, 2002) may provide a means of understanding how covert retrieval in WM occurs: less activated content that was shunted from immediate accessibility due to its irrelevance to the task at hand (e.g., solving the processing task) must be retrieved as a single chunk. This may not occur to the same degree for simple span trials of longer list lengths as the new incoming information represents the consistently relevant task of attempting to maintain and retrieve all of the memoranda. Accordingly, the earliest presented memoranda that are presumably displaced into secondary memory by new incoming memoranda do not have the same opportunity to be covertly retrieved throughout the trial, and their cue-driven search can only occur during the immediate recall attempt. This would suggest that the mental task set shifts during complex span provide more retrieval practice via covert retrieval overall than long simple span trials, thereby yielding differences in the retention and phenomenological memorability of the memoranda. More work will be necessary in order to consider these possibilities.

There may be more optimal explanations of these findings than our covert retrieval account (Souza & Oberauer, 2017). Although we have addressed alternative explanations in prior work, such as the possibility that covert retrieval is simply rehearsal (Loaiza & McCabe, 2013) or that the McCabe effect reflects a general advantage of spacing and/or temporal distinctiveness (Loaiza & McCabe, 2012), more work is necessary to unequivocally reject these and other accounts. For example, Souza and Oberauer (2017) recently showed that memoranda that were presented in trials with a fixed blank period of equal duration to the distraction of operation span trials were most likely to be recalled after a delay. Thus, rather than covert retrieval being the principal underlying source of the McCabe effect, it may be the case that complex span affords greater opportunity for elaborative rehearsal or consolidation that promotes the long-term retention of these items. However, if elaborative rehearsal explained the effect, it would be expected that the memorability of the semantic characteristics would yield the greatest McCabe effect, with very little effect for phonological characteristics in Experiment 1. This was not the case, thereby casting doubt on the elaboration explanation. Rose and colleagues (2014) also showed that an advantage for memoranda similar to the McCabe effect did not change as a function of the level of processing (i.e., shallow or deep) engaged during encoding. Another possibility is that the McCabe effect may reflect increased opportunity for consolidation (e.g., Bayliss, Bogdanovs, & Jarrold, 2015) due to the greater time between the presentation of the memoranda during complex versus simple span. Contrary to the covert retrieval explanation that all of the memoranda are cumulatively covertly retrieved (McCabe, 2008), the consolidation account only presumes the last-presented item in WM is consolidated, and thus there should only be an overall advantage of complex span items that does not change with serial position. However, the pattern of negative recency of complex span exhibited in the current results (Figure 2) and in our other studies (e.g., Loaiza & McCabe, 2012; McCabe, 2008) conflicts with this account, and instead comports with the covert retrieval account. Thus, while more work remains to be done, so far there is less substantial support for these alternative accounts in the present study.

In summary, these results replicate and extend prior work indicating a long-term objective (Loaiza & McCabe, 2012, 2013; McCabe, 2008) and subjective (Loaiza et al., 2015) memorial advantage for information presented in tasks during which attention to representations in WM is regularly distracted. The findings are in line with the covert retrieval account suggesting this advantage of complex span over simple span memoranda (i.e., the McCabe effect) is due to the increased accessibility of retrieval cues that are reinforced via covert retrieval during the encoding phase of a complex span task. This was particularly evident in the finding that immediate recall did not moderate the McCabe effect, emphasizing it as an encoding-specific effect, and that the earliest serial positions were particularly memorable in terms of their temporal-contextual characteristics. Conversely, tasks that do not introduce any opportunity to covertly retrieve memoranda in WM (e.g., simple span tasks at any list length) do not confer such benefits to either objective performance or the details that are phenomenologically available during EM. As such, covert retrieval in WM appears to be an important underlying factor that promotes later subjective recollective experience in EM.

References

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

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. https://doi.org/10.1016/j.jml.2014.12.004

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language*, *61*(3), 457–469. https://doi.org/10.1016/j.jml.2009.06.002

Cousineau, D. (2005). Confidence intervals in within-subjects designs: A simpler solution to Loftus and Masson’s method. *Tutorials in Quantitative Methods for Psychology*, *1*, 42–45. https://doi.org/10.20982/tqmp.01.1.p042

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

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*(4), 309–313. https://doi.org/10.3758/BF03197041

Johnson, M. K. (1992). MEM: Mechanisms of Recollection. *Journal of Cognitive Neuroscience*, *4*(3), 268–280. https://doi.org/10.1162/jocn.1992.4.3.268

Johnson, M. K., Foley, M. A., Suengas, A. G., & Raye, C. L. (1988). Phenomenal characteristics of memories for perceived and imagined autobiographical events. *Journal of Experimental Psychology. General*, *117*(4), 371–376.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. https://doi.org/10.1037/0096-3445.133.2.189

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

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. https://doi.org/10.3758/s13423-014-0673-7

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. https://doi.org/10.3758/s13421-011-0148-2

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.

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition*, *25*(6), 826–837. https://doi.org/10.3758/BF03211327

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. https://doi.org/10.1016/j.jml.2007.04.004

McCabe, D. P., & Geraci, L. (2009). The role of extralist associations in false. *Memory & Cognition*, *37*(2), 130–142. https://doi.org/10.3758/MC.37.2.130

McCabe, D. P., Geraci, L., Boman, J. K., Sensenig, A. E., & Rhodes, M. G. (2011). On the validity of remember–know judgments: Evidence from think aloud protocols. *Consciousness and Cognition*, *20*(4), 1625–1633. https://doi.org/10.1016/j.concog.2011.08.012

McCabe, D. P., Roediger III, H. L., McDaniel, M. A., & Balota, D. A. (2009). Aging reduces veridical remembering but increases false remembering: Neuropsychological test correlates of remember–know judgments. *Neuropsychologia*, *47*(11), 2164–2173. https://doi.org/10.1016/j.neuropsychologia.2008.11.025

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs. (Version 0.9.12-2). Retrieved from http://CRAN.R-project.org/package=BayesFactor

Neuschatz, J. S., Payne, D. G., Lampinen, J. M., & Toglia, M. P. (2001). Assessing the effectiveness of warnings and the phenomenological characteristics of false memories. *Memory*, *9*(1), 53–71. https://doi.org/10.1080/09658210042000076

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, *25*(6), 838–848. https://doi.org/10.3758/BF03211328

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411–421. https://doi.org/10.1037//0278-7393.28.3.411

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory Illusions: Recalling, Recognizing, and Recollecting Events that Never Occurred. *Journal of Memory and Language*, *35*(2), 261–285. https://doi.org/10.1006/jmla.1996.0015

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 803–814. https://doi.org/http://dx.doi.org/10.1037/0278-7393.21.4.803

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., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://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. https://doi.org/10.3758/PBR.16.2.225

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Souza, A. S., & Oberauer, K. (2017). Time to process information in working memory improves episodic memory. *Journal of Memory and Language*, *96*, 155–167. https://doi.org/10.1016/j.jml.2017.07.002

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*(1), 1–12. https://doi.org/10.1037/h0080017

Turner, M. L., & Engle, R. W. (1989). Is Working Memory Capacity Task Dependent? *Journal of Memory and Language*, *28*(2), 127–154.

Unsworth, N. (2010). On the division of working memory and long-term memory and their relation to intelligence: A latent variable approach. *Acta Psychologica*, *134*(1), 16–28. https://doi.org/10.1016/j.actpsy.2009.11.010

Unsworth, N., & Brewer, G. A. (2009). Examining the relationships among item recognition, source recognition, and recall from an individual differences perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1578–1585. https://doi.org/10.1037/a0017255

Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, *54*(1), 68–80. https://doi.org/10.1016/j.jml.2005.06.003

Unsworth, N., & Engle, R. W. (2007a). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038–1066. https://doi.org/10.1037/0033-2909.133.6.1038

Unsworth, N., & Engle, R. W. (2007b). 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. https://doi.org/10.1037/0033-295X.114.1.104

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*(4), 392–406. https://doi.org/10.1016/j.jml.2010.02.001

Ward, G., Tan, L., & Grenfell-Essam, R. (2010). Examining the relationship between free recall and immediate serial recall: The effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1207–1241. https://doi.org/10.1037/a0020122

Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Table 1. Means (and standard deviations) of immediate serial recall (ISR) and immediate free recall (IFR) scoring methods in Experiments 1 and 2.

Experiment	Trial Type	ISR	IFR
1	Word Span	0.94 (0.07)	0.96 (0.04)
	Operation Span	0.50 (0.15)	0.71 (0.11)
	BF_{10}	2.47×10^{9}	3.49×10^{8}
	d	2.84	2.68
	95% HDI	[1.69, 4.03]	[1.55, 3.98]
2	Word Span List Length 4	0.94 (0.08)	0.97 (0.04)
	Word Span List Length 8	0.27 (0.14)	0.50 (0.10)
	BF_{10}	1.34×10^{24}	9.63×10^{22}
	d	5.28	4.99
	95% HDI	[3.87, 6.69]	[3.59, 6.51]

Note. Bayes factors (BF_{10}; the relative evidence for the alternative model with a difference between trial types over the null model), Bayesian effect size (*d*), and Highest-Density Interval (HDI) of the effect size are displayed. See text for details.

Table 2. Results of the BANOVAs for each memory characteristics questionnaire (MCQ) ratings and experiment.

			Fixed Effects Model			
Experiment	Characteristic	Model (M) Ratio	Trial Type	Immediate Recall Condition	Trial Type + Immediate Recall Condition	Trial + Recall + Trial x Recall
1	Phonological	BF_{10}	3265.2	8.3	**1.0×10^{5}**	39,995.6
		Best M/M	30.8	12,077.6	Best	2.5
	Semantic	BF_{10}	3.1×10^{6}	12.7	**7.1×10^{8}**	2.0×10^{8}
		Best M/M	232.1	5.6×10^{7}	Best	3.5
	Temporal-Contextual	BF_{10}	3.5×10^{7}	64.9	**3.7×10^{11}**	1.1×10^{11}
		Best M/M	10538.4	5.7×10^{9}	Best	3.3
2 (all recall)	Phonological	BF_{10}	1.1	10.1	**12.4**	4.0
		Best M/M	11.7	1.2	Best	3.1
	Semantic	BF_{10}	0.3	**274.2**	74.6	35.9
		Best M/M	1074.3	Best	3.7	7.6
	Temporal-Contextual	BF_{10}	0.2	**1,676.8**	415.8	100.2
		Best M/M	7,240.3	Best	4.0	16.7
2 (first 4 of list length 8)	Phonological	BF_{10}	0.4	0.7	0.3	0.2
		Best M/M	-	-	-	-
	Semantic	BF_{10}	0.2	**9.1**	1.8	1.8
		Best M/M	46.1	Best	5.0	5.1
	Temporal-Contextual	BF_{10}	0.2	**110.2**	22.7	6.1
		Best M/M	528.9	Best	4.9	18.1

Note. All models include participant as a random effect. The Bayes factor (BF) refers to the evidence for the alternative model (BF_{10}) of each effect (shown in different columns) relative to the null model (i.e., intercept-only model). The best model is shown in boldface font in the first row for each measure, and the second row for each measure compares the best model in the numerator to each of the other models in the denominator.

Table 3. Means (and standard deviations) of the memory characteristics questionnaire (MCQ) ratings correcting for initial recall in Experiments 1 and 2.

Experiment	Trial Type	phonological	semantic	temporal-contextual
1	Word Span	4.49 (1.57)	4.72 (1.39)	3.69 (0.71)
	Operation Span	5.16 (1.97)	5.64 (1.50)	4.91 (1.15)
	BF_{10}	111.96	2,848.1	188,123.5
	d	0.85	1.14	1.61
	95% HDI	[0.36, 1.33]	[0.68, 1.69]	[0.93, 2.29]
2	Word Span List Length 4	4.06 (1.63)	4.61 (1.79)	3.52 (1.46)
	Word Span List Length 8	4.20 (1.60)	4.92 (1.81)	3.97 (1.56)
	BF_{10}	0.4	4.3	5.6
	d	0.21	0.45	0.47
	95% HDI	[-0.20, 0.61]	[0.10, 0.82]	[0.11, 0.84]
	Word Span List Length 8 (first 4)	4.11 (1.64)	4.93 (1.94)	3.91 (1.59)
	BF_{10}	0.2	1.5	1.3
	d	0.04	0.36	0.34
	95% HDI	[-0.32, 0.38]	[-0.01, 0.71]	[-0.03, 0.70]

Note. Bayes factors (BF_{10}; the relative evidence for the alternative model with a difference between trial types over the null model), Bayesian effect size (*d*), and Highest-Density Interval (HDI) of the effect size are displayed. See text for details.

Figure 1. Mean ratings on the delayed memory characteristics memoranda originally studied during complex span and simple span in Experiment 1. Error bars reflect 95% within-subjects confidence intervals (Cousineau, 2005; Morey, 2008).

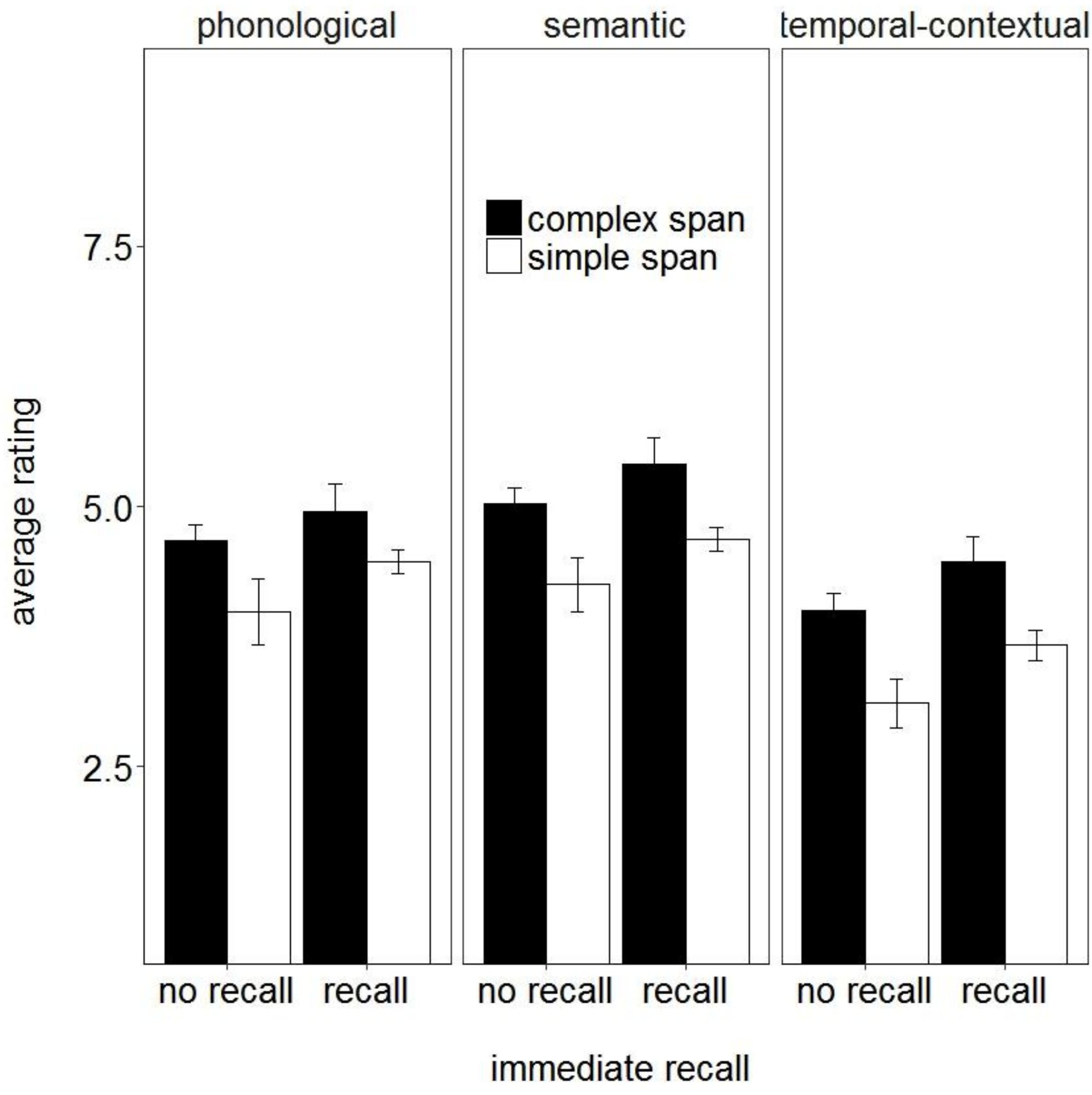

Figure 2. Mean ratings on the delayed memory characteristics (A) and corrected for initial immediate recall (B) of memoranda originally studied during complex span and simple span as a function of serial position in Experiment 1. Error bars reflect 95% within-subjects confidence intervals (Cousineau, 2005; Morey, 2008).

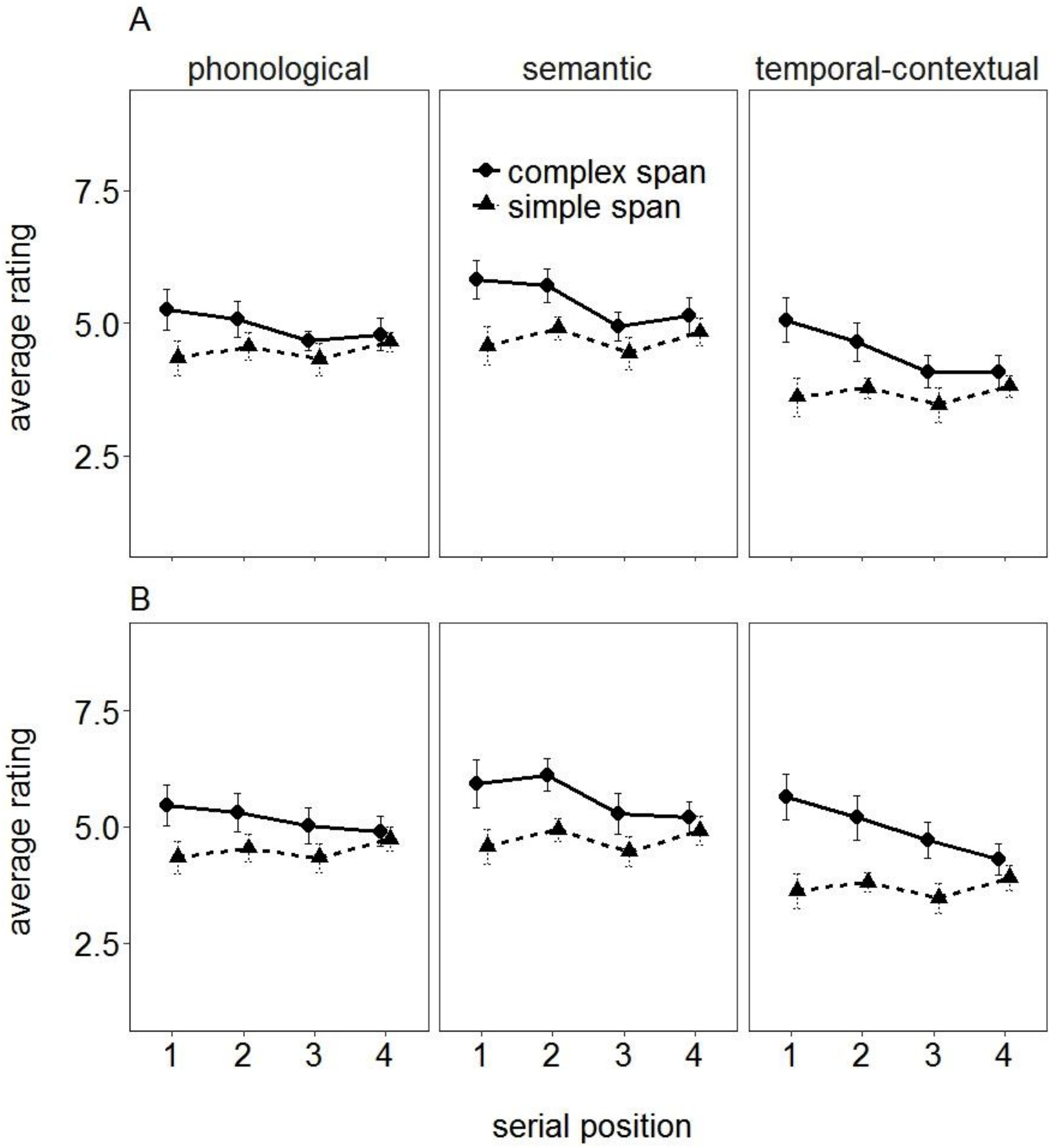

Figure 3. Mean ratings on the delayed memory characteristics questionnaire (MCQ) as a function of recall condition, trial type, and characteristic of memoranda originally studied during list length 4 and 8 trials. Error bars reflect 95% within-subjects confidence intervals (Cousineau, 2005; Morey, 2008).

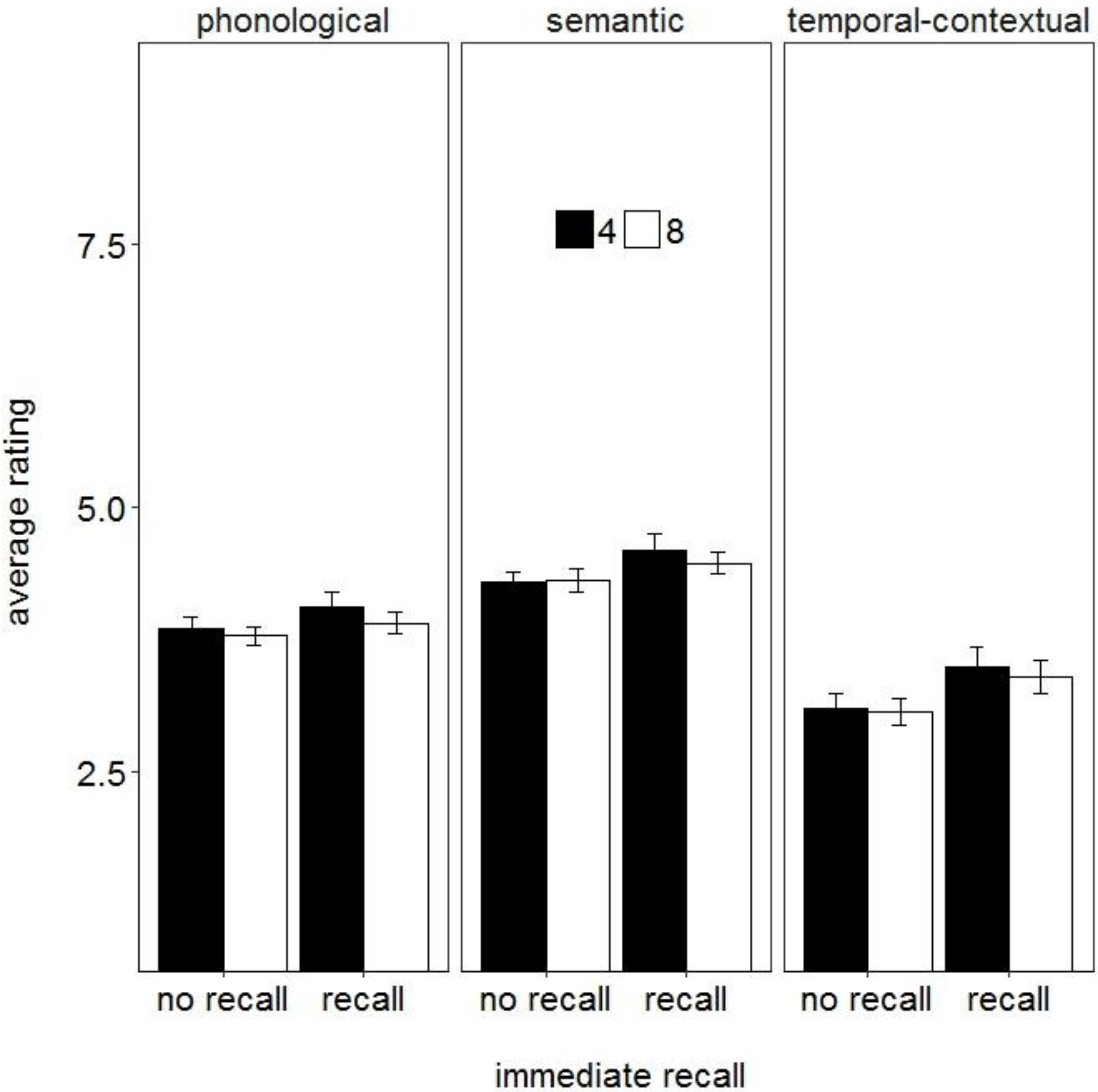

Figure 4. Mean ratings on the delayed memory characteristics questionnaire (MCQ) as a function of recall condition, trial type, and characteristic from list length 4 and the first four memoranda of list length 8. Error bars reflect 95% within-subjects confidence intervals (Cousineau, 2005; Morey, 2008).

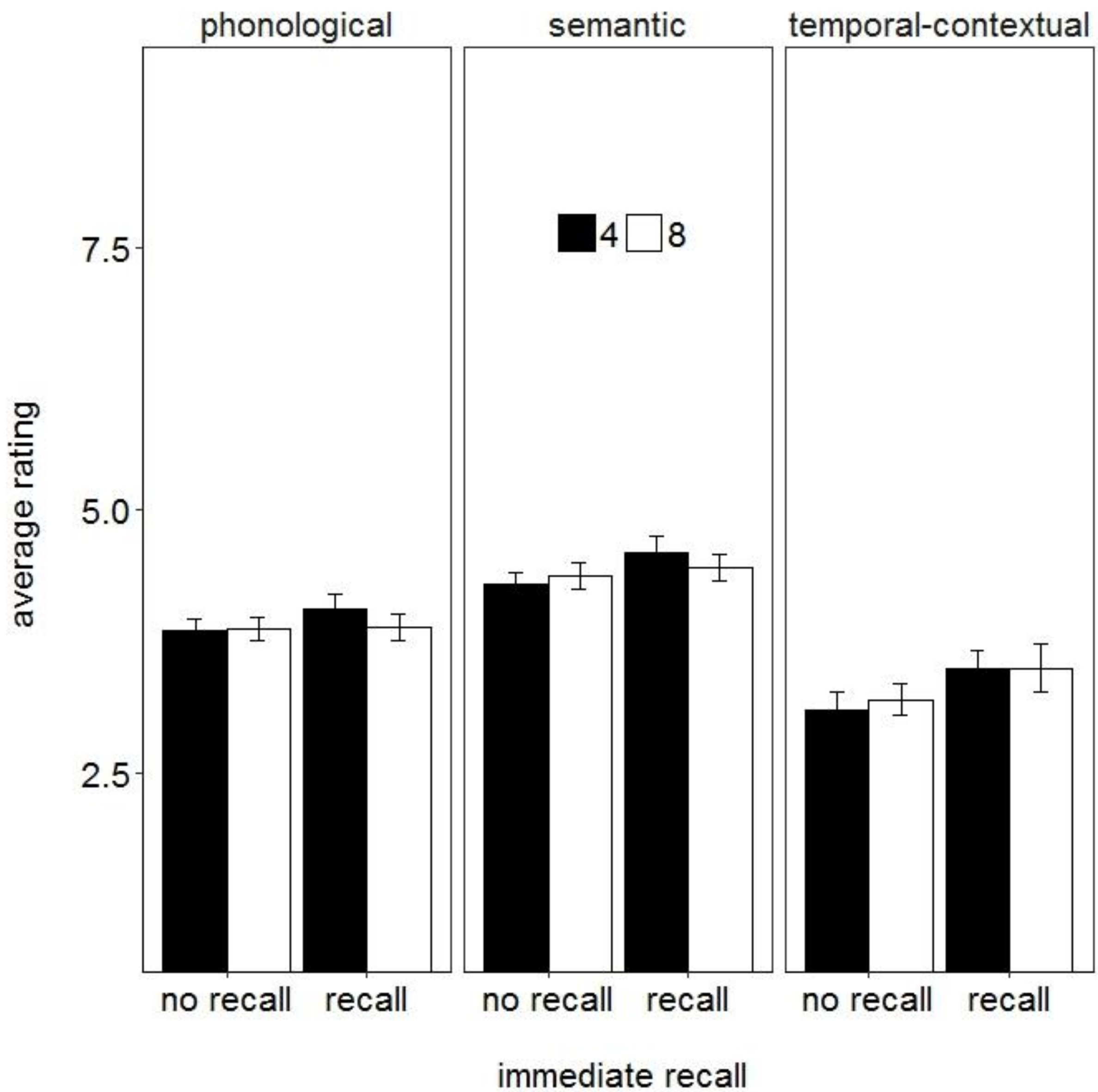

Figure 5. Mean ratings on the delayed memory characteristics (A) and corrected for initial immediate recall (B) of memoranda originally studied during list length 4 and the first four memoranda of list length 8 as a function of serial position in Experiment 2. Error bars reflect 95% within-subjects confidence intervals (Cousineau, 2005; Morey, 2008).