Differential Information Transfer and Loss between Working Memory and Long-Term Memory Across Serial Positions

Alicia Forsberg[1], Dominic Guitard[2], Nathaniel R. Greene[3], Moshe Naveh-Benjamin[4], and Nelson Cowan[4]

[1] University of Sheffield

[2] Cardiff University

[3] University of Pennsylvania

[4] University of Missouri

Author Note

Contact details for the corresponding author: Dr. Alicia Forsberg, Cathedral Court, 1 Vicar Ln, Sheffield City Centre, Sheffield S1 2LT, United Kingdom. a.forsberg@sheffield.ac.uk,

Preregistrations for both experiments are available at https://osf.io/43hsd/?view_only=01456942b0ff4c31baac6bfd96a844aa. Anonymized data and materials are available at https://osf.io/y3qxf/?view_only=cdd71f1f49e54011bb549145edc87ebe

.We thank Bret Glass for assisting in data collection and acknowledge NIH Grant R01 HD-21338 to N. Cowan, NIH Grant F32AG087646 to N. R. Greene, and support by the Jacobs Foundation to A. Forsberg.

Author Contribution Statement

A. Forsberg developed the study concept with crucial input from all authors. All authors contributed to the study design. D. Guitard programmed the online experiment enabling data collection. N. Greene performed the Bayesian hierarchical *k* estimation, and A. Forsberg performed the remaining data analysis. A. Forsberg drafted the manuscript, and all authors provided critical revisions. All authors approved the final version of the manuscript for submission.

Abstract

Working Memory (WM) is the cognitive system that allows the temporary holding of mental representations for use in thought and action. Long-Term Memory (LTM) refers to our ability to remember a potentially unlimited amount of information over longer time periods. Understanding how these two memory systems interact has important implications for theories of cognition, learning, and education. Here, we examined (1) whether a shared perceptual bottleneck accounts for the relation between WM and LTM accuracy, and (2) whether serial position effects in WM are mirrored in LTM. In two experiments, participants studied sequences of objects at varying set sizes and completed old/new recognition tests for some items immediately after encoding (WM tests) and for other items after all WM trials were completed (LTM tests). In Experiment 1 (N = 80), LTM performance was better for items presented in lower rather than higher set-size sequences, indicating that limitations in WM capacity constrain LTM encoding, irrespective of perceptual bottlenecks. In Experiment 2 (N = 120), we observed WM and LTM recency effects, but primacy effects were only present in LTM and not in WM. Thus, serial position effects in WM did not consistently predict the relative rates at which items from different serial positions were preserved in LTM. These results reinforce accounts that view WM and LTM as having at least partially separate mechanisms, shedding light on the nature of these mechanisms.

Keywords: Working Memory, Short-Term Memory, Long-Term Memory, Primacy Effects, Recency Effects.

Differential Information Transfer and Loss between Working Memory and Long-Term Memory Across Serial Positions

The relationship between Working Memory[1] (WM) or Short-term Memory (STM) and Long-term Memory (LTM) has puzzled researchers for over a century (Broadbent, 1958; Cowan, 2019; Ebbinghaus, 1885/1913; Hebb, 1949; James, 1890; Shiffrin & Atkinson, 1969). WM is a system that allows the temporary holding of a limited number of mental representations for use in thought and action (Cowan, 2017; Logie et al., 2020), whereas LTM refers to our ability to retain large quantities of information over longer time periods (see Malmberg et al., 2019; Shiffrin & Atkinson, 1969).

Understanding how these two memory systems interact has crucial implications for theories of memory, cognition, and learning, with far-reaching practical implications for education, human information processing, and thinking more broadly. For example, limitations in how much information we can process initially may impact our ability to remember that information later. However, receiving too much information at once may impair long-term retention, compared to learning in more manageable chunks (Forsberg et al., 2021b). Consequently, numerous theoretical accounts have been put forward to explain the relationship between the two systems, but empirical evidence supporting these accounts has left unaddressed several important questions about this relation. Here, we first seek to answer whether previous

[1] Some researchers use WM and STM interchangeably, while others use the term WM only to refer to task paradigms which require concurrent memory and processing of information (for a discussion, see Cowan, 2017). We use the term WM to describe our memory task, which requires an active comparison between the probe item and the contents of the memory store and has a limited (i.e., within seconds), temporal duration. In this manuscript, for consistency and simplicity, we use the term WM to refer to research with paradigms that other researchers may refer to as measuring either WM or STM.

findings of a WM bottleneck on long-term retention in several recent studies (Forsberg et al. 2021a; 2022a; 2022b; 2023; Fukuda & Vogel, 2019) could instead be explained by a shared perceptual bottleneck between the two systems. Next, we address whether serial position effects (i.e., improved memory for items in certain positions within a sequence) transfer between WM and LTM in recognition procedures that minimize recall dynamics that may obscure or reduce evidence for this transfer. Our first goal has the potential to upend or strengthen current theoretical models of the WM-LTM relation, while our second goal will shed insight on whether the mechanisms that support successful WM for items from specific list positions also enhance LTM for items in those positions. Combined, these two goals will further the understanding of how the WM and LTM systems interact. We turn now to a brief overview of competing theoretical accounts on the relation between WM and LTM.

Theoretical Perspectives on the Relationship between WM and LTM

Theorists have long disagreed about the nature of the WM-LTM relation. Some theorists view WM and LTM as two separate memory systems (e.g., Atkinson & Shiffrin, 1968; Craik, 1970; Plancher & Barrouillet, 2020; Norris, 2017; Shallice & Warrington, 1970; Scoville & Milner, 1957). This view has been supported by findings that patients (e.g., Patient “H.M.”) who struggled to form new LTM representations nonetheless performed well on STM/WM tests (Scoville & Milner, 1957). Other patients, however, exhibited the opposite pattern, with impaired immediate memory despite normal LTM performance (Shallice & Warrington, 1970). These conflicting patterns of memory impairments represent a compelling neuropsychological double dissociation between WM and LTM. Moreover, the perception of WM and LTM as two separate memory systems is also supported by computational accounts (e.g., Norris, 2017) and by experimental data. For example, some studies have highlighted that continuously repeating the same memory array in a standard WM task does not necessarily result in improved LTM (e.g., Shimi & Logie, 2019). According to this view, there need not be a close correlation between WM and LTM performance patterns on the same task.

In contrast, other researchers propose that the most parsimonious account of existing behavioral and neuroimaging data is that both WM and LTM are managed by a single memory system and are both governed by the same principles and processes (e.g., Brown et al., 2007; McElree, 2006; Nairne, 2002; Surprenant & Neath, 2009). Such accounts emphasize that many important human memory phenomena are timescale-invariant, suggesting that most seemingly differential patterns can be explained by models that assume that the same processes operate for both WM and LTM recall (e.g., SIMPLE; *Scale Invariant Memory and Perceptual Learning;* Brown et al., 2002; Surprenant & Neath, 2009). These researchers would expect a close correlation between WM and LTM performance patterns.

In between these views, others envision the WM and LTM systems as closely intertwined, viewing WM as a momentarily activated, capacity-limited subset of LTM information (e.g., Cowan, 1988, 2019; Morey et al., 2013). We note that conceptualizations of the relationship between the two systems depend on definitions of STM/WM and LTM, which remain contentious (see Cowan, 2017). Hence, the differences among these contrasting theoretical frameworks may not be as extreme as they appear at first glance. For example, most researchers who posit that WM and LTM are supported by the same system would acknowledge that when memory is tested after a short delay (WM), the quality of the memory representations is likely to differ from when it is tested after a long delay (e.g., Greene & Naveh-Benjamin, 2022a, 2022b; Öztekin et al., 2010). On the other hand, researchers who emphasize differences between the two systems do not deny that the two systems may interact – for example, that information stored in LTM (e.g., word meanings) can be used to support WM performance (e.g., Hulme et al., 1991). According to this intermediate view, some aspects of WM are likely to be preserved in LTM (e.g., rapid learning and improvement in the memory representation: Cowan, 2019; Cowan et al., 2024; Craik, 2020; Ricker & Vergauwe, 2022). However, other aspects of WM are likely to reflect temporary maintenance mechanisms that do not facilitate LTM (e.g., refreshment of a decaying memory trace: Barrouillet & Camos, 2015; distinctiveness of very recently presented items: Glenberg & Swanson, 1986; Unsworth et al., 2008).

Here, we focus on the specific suggestion that STM/WM acts as a bottleneck for LTM encoding, initially proposed by Atkinson and Shiffrin (1968) and supported by recent experimental work (e.g., Čepukaitytė et al., 2023; Cotton & Ricker, 2021; Greene et al., 2024; Jeanneret et al., 2023; Forsberg et al., 2021a; 2022a; 2022b; 2023; Fukuda & Vogel, 2019, see also Bartsch et al., 2019; Loaiza et al., 2023). According to this account, WM limitations constrain LTM encoding, such that successful retention of information in the limited WM system also predicts the likelihood of successful retention of that information at a delayed LTM test. Forsberg et al. (2021a) developed a novel paradigm and analytical approach to test this hypothesis. This paradigm relies on a core distinguishing feature of the WM system: its capacity limit (3 – 4 items, Adam et al., 2017; Cowan, 2001, although see Williams et al., 2022, for a competing theoretical account). By manipulating the number of items presented in a trial – ranging from two items to as many as eight items presented at once – Forsberg et al. manipulated the probability that each item was held in WM. The logic was as follows: if the successful maintenance of items in WM predicts subsequent LTM recall, then items from lower set-size arrays in the WM procedure should be better remembered in a later LTM test, because an item in a smaller set-size array is more likely to be successfully held in WM. In these studies, WM set size was found to constrain LTM encoding, which was interpreted as evidence that WM acts as a bottleneck for LTM (Forsberg et al., 2021a; 2022a; 2022b).

In the present study, we made one critical change to this paradigm: rather than simultaneous presentation, we presented items sequentially in each WM trial. This allowed us to explore two key questions that may redefine our theoretical understanding of the relation between WM and LTM. We turn now to discussing the first of these questions.

WM-LTM Bottleneck: Perceptual Load or WM capacity?

Our first aim was to explore a potential alternative explanation for the WM-to-LTM bottleneck observed in previous studies with simultaneous presentation of items during encoding (e.g., Forsberg et al., 2021a; Fukuda & Vogel, 2019). Some past studies using sequential rather than simultaneous presentation methods did not find that WM load during encoding limited LTM for word pair bindings (Bartsch et al., 2019). On the other hand, evidence for a WM set size effect on LTM for items was found for smaller, sub-capacity sequences of two to four colorful items (Forsberg et al., 2023). However, the majority of recent studies supporting the WM-to-LTM bottleneck account have used simultaneous presentation (e.g., Forsberg et al. 2021a; 2022a; 2022b; 2023; Fukuda & Vogel, 2019). Under simultaneous presentation, the effect of WM set size on subsequent LTM retrieval may be caused by a shared perceptual encoding bottleneck rather than a WM capacity limitation, per se. Suppose that one is examining the memory of an individual with WM capacity k on a set of N items. Both an array and a list presumably impose a memory load of either k or N items, whichever is smaller, by the time of a WM recognition test. However, during the encoding period for each of the N items, the two presentation methods differ in the perceptual load imposed. The concurrent perceptual load is N in the case of an array, whereas it is always one in a list. If the perceptual load, rather than the WM load, was the limiting factor on LTM performance for items presented simultaneously during encoding in previous studies (e.g., Forsberg et al., 2021a), then LTM performance should not differ for items presented sequentially in lower versus higher set-size lists.

In both simultaneous and sequential presentation of items, complex attentional processes are likely used to keep memoranda in mind, and while presentation can be controlled experimentally, it is notoriously difficult to distinguish serial (one at a time) from parallel (concurrent) processing (see Cowan & Guitard, 2024; Townsend, 1971, 1990; Townsend & Wenger, 2004). Moreover, higher set size arrays are inherently more visually complex given the simultaneous presentation of items. Although Forsberg et al. adjusted the encoding time for each item (e.g., a two-item array was shown for 500 ms and a four-item array for 1000 ms), the arrays were arguably still qualitatively different. For example, arrays containing more items may cause inter-item interference (see Oberauer et al., 2012), or participants may apply different strategies to such arrays, such as choosing to focus their attention on only a small subset of items (see Atkinson et al., 2022; Jeanneret et al., 2023).

Cowan et al. (2011) found comparable age differences in WM change-detection accuracy between participants who studied four colored shapes simultaneously versus those who studied the shapes sequentially. These findings suggest that age differences in WM capacity, rather than differences in perceptual load imposed by sequential versus simultaneous presentation methods, was the limiting factor on recognition accuracy immediately after encoding. Accordingly, it is conceivable that WM capacity was also the limiting factor on LTM in the previous studies exploring the WM-to-LTM bottleneck with simultaneous presentation. If so, we should observe a set size effect for LTM performance with sequential presentation in the current study. Such a finding would provide stronger support for theories that view WM and LTM as closely intertwined systems, such as models that view WM as a capacity-limited subset of LTM (e.g., Cowan, 1988; Cowan et al., 2024). However, if the set size effect does not replicate with sequential presentation, the conclusions of Forsberg et al. (2021a) and others (e.g., Fukuda & Vogel, 2019) would be called into question, suggesting that the perceptual load – rather than WM capacity limitations – limited both WM and LTM performance. Such a finding would pose a challenge for models predicting that WM is the "gateway" through which new LTM representations are formed (e.g., Atkinson & Shiffrin, 1968; Cowan, 1988). In contrast, such a finding may align well with unitary models that predict timescale-invariance in the processes that affect memory (Brown et al., 2007; McElree, 2006; Nairne, 2002; Surprenant & Neath, 2009).

Primacy and Recency Effects: How do WM Maintenance Processes Affect LTM Retention?

Our second aim was to explore how WM maintenance processes affect LTM retention by investigating whether serial position effects (i.e., better memory for items presented at certain positions in the sequence) in WM are mirrored in LTM. If serial position effects in WM mirror those in LTM, this would suggest that the processes that support successful WM for items in a sequence (i.e., WM maintenance processes) translate into improved LTM for those items, as well.

The most widely reported serial position effects are primacy effects (enhanced memory for items from early list positions) and recency effects (enhanced memory for items from the last few list positions), popularized by Ebbinghaus (1885/1913, also see Siegler, 1978 for an account of earlier work by Nipher). Primacy and recency effects have been reported in verbal free recall by Murdock (1962) and have been observed in numerous WM/STM paradigms using verbal stimuli (e.g., Guitard & Cowan, 2023; Saint-Aubin & Poirier, 2000) and visuospatial stimuli

(Jones et al., 1995; Smyth & Scholey, 1996). Extensive prior research has established that primacy and recency effects are sensitive to various experimental manipulations (for a review, see Kahana, 2017). For example, the recency effect is attenuated by a brief, distractor-filled delay, compared to immediate free recall (Glanzer & Cunitz, 1966; Postman & Phillips, 1965), and primacy effects are reduced when a continuous distractor task is introduced between each study event (Kahana, 2012) but persist over long time periods when participants have ample free time between item presentations to process each item (Bjork & Whitten, 1974; Howard & Kahana, 1999). The attenuation of recency with delay, in particular, may lead naturally to the prediction that serial position effects in WM would not be mirrored in LTM. However, these classic studies relied on free recall paradigms with verbal materials and occasionally administered both immediate and delayed free recall tests on the same lists (e.g., Craik, 1970). In immediate free recall, participants tend to initiate recalls from recency items, but recall initiation is driven by primacy items for delayed free recall (Healey et al., 2014; Kahana, 2017). Transitions between items are driven primarily by shared temporal or semantic relations among the items in the list (Howard & Kahana, 2002; Kahana, 1996, 2020), such that participants are less likely to make temporally or semantically distant transitions. The exception is that when subjects initiate recall with end-of-list items, they then tend to transition to the beginning of the list and proceed to recall items in an asymmetrically forward direction (i.e., transitioning to the next item in the sequence rather than to an earlier serial position; see Kahana, 1996). Accordingly, participants who initiate recall with primacy items (as in the case of delayed free recall) are less likely to transition to recency items. Because participants recall items as they come to mind in free recall tasks, output interference may further constrict their ability to recall end-of-list items when recall is initiated with beginning-of-list items (cf., Cowan et al., 2002). It is conceivable, however, that in recognition procedures like ours, recency effects would transfer from WM to LTM tests. Such tests minimize output interference due to recalling multiple items and the strong temporal and semantic dynamics that drive recalls potentially away from end-of-list items in classic delayed free recall procedures where recall is initiated with primacy items. Indeed, prior recognition studies have observed enhanced LTM for recency items, relative to middle list-position items, at least under incidental encoding conditions (Jiang & Cowan, 2020). In the procedure to-be-employed in the present study, participants are aware that their memory will be tested at the end of each WM trial, but they are not explicitly informed about the LTM tests at the end of the experiment (see Forsberg et al., 2021a). Thus, although participants intentionally encode items for immediate WM tests, their LTM encoding (i.e., encoding the items in such a way to promote long-term retention) is relatively more incidental.

Several theories can account for the primacy and recency effects observed in the serial position curve. Page and Norris (1998) proposed that primacy arises because the amount of attention to each list item declines as items are added to the memory load during list presentation (see also Azizian & Polich, 2007; Brown et al., 2007; Sederberg et al., 2006). Primacy effects have also been attributed to the differential rehearsal of beginning-of-list items relative to subsequent items (e.g., Fischler et al, 1970; Rundus & Atkinson, 1970; Tan & Ward, 2008). Other accounts suggest that an item's strength of encoding in LTM is a direct function of its length of stay in the short-term store. In such accounts, it is assumed that primacy effects transfer to LTM because early list items enjoy the longest stay in the WM store (see Craik, 1970).

By examining the serial position function, one can observe the buildup of the WM load over time (e.g., Chen & Cowan, 2009). This allows a distinction between the load at the time that each item is encoded (presumably equal to either *k* or the serial position in the list minus one, whichever is smaller) versus the load by the time of the test (*k* or the set size *N*, whichever is smaller). Some recent work on consolidation suggests that the load at the time of encoding could be important because WM limitations constrain item consolidation processes (for a review, see Ricker et al., 2018; see also Ricker & Hardman, 2017). For example, Vergauwe et al. (2014) asked participants to maintain series of letters of different lengths during a 12-second delay filled by a processing task (i.e., a parity judgment task on digits appearing successively). They found that processing time was longer for the first, compared with the subsequent digits, and they attributed this delay to the consolidation process of the letters. Crucially, they found this postponement increased with the number of memory items.

There are reasons to expect enhanced memory for the final list items at various time delays as well. Recency effects could arise due to distinctiveness of the items near both ends of the list (Brown et al., 2007; Kuhn et al., 2018) or the displacement of earlier items as new items arrive (Maskarinec & Brown, 1974). Similarly, end-of-list items may be retained in an activated, more accessible state, which may enhance recall (e.g., Azizian & Polich, 2007; Davelaar et al., 2005; Glanzer & Cunitz, 1966; Shriffin & Atkinson, 1969). If the final item presented in the WM set is 'fresher' in the participants' minds, they may be more likely to rehearse it during the retention interval, compared to earlier list-position items that may have been forgotten or overwritten by later items. For example, participants may rehearse by visualizing or by repeating an item's verbal label, if the retention interval is not occupied by another task or activity. This, in turn, could result in more durable LTM representations for end-of-list items, yielding preserved recency effects in LTM tests.

Finally, Craik and Lockhart (1972) suggested that LTM performance depends on the level – rather than the duration – of processing of the memoranda. They found that keeping information in an active state did not bolster memory performance, but instead, information that was processed more deeply was better remembered than information that was processed in a shallower way.

Combined, extant theoretical accounts and empirical observations of primacy and recency effects yield several interesting predictions for our study. For example, while a shallow, verbal rehearsal strategy is likely to support WM performance (e.g., Forsberg et al., 2020), it would likely exert a less meaningful impact on LTM performance (e.g., Craik & Lockhart, 1972). Thus, if WM primacy or recency effects are underpinned by shallow processes, we may not observe a similar boost (relative to middle list-position items) for those items in subsequent LTM tests. In contrast, if WM primacy or recency effects are supported by deep rehearsal mechanisms, we would expect items presented in those positions to also fare better in LTM, bolstered by deeper, more attentive encoding (Craik et al., 1970; Glanzer & Cunitz, 1966; Page & Norris, 1998).

Finally, we note that the role of attention in shallow and deep encoding processes is not theoretically straightforward. During item encoding, the semantic or conceptual representation of an item can be formed rapidly, with little need for sustained attention, as indicated by evidence of rapid gist extraction in both the verbal (e.g., Draine & Greenwald, 1998) and the visual domain (Potter, 1976; Tatler, 2003), and the reduced attentional demands of encoding semantic and gist representations (e.g., Greene & Naveh-Benjamin, 2023; Odegard & Lampinen, 2005). However, in order to firmly establish those representations, items appear to be held at least some of the time within the Focus of Attention (see Cowan et al., 2024), as recent research suggests that long-term retention of gist (semantic) representations is affected by capacity constraints during initial encoding (Greene et al., 2024). Thus, while the initial establishment of a semantic representation embedded within a specific episodic context may be relatively automatic, the formation of a more durable representation (what we refer to as 'deep maintenance processes') likely requires some more active and focused commitment of attention.

To gain insight into the mechanisms used to support memory performance in these two systems, we explored whether patterns of primacy and recency effects generalized from WM to LTM trials .

Experiment 1

The first goal of Experiment 1 was to test whether the findings of Forsberg et al. (2021a) would replicate when to-be-remembered items were presented sequentially (as opposed to simultaneously). The switch to a sequential presentation method enabled us to test competing hypotheses about the WM-to-LTM bottleneck observed in previous studies. According to one hypothesis, if WM capacity is the limiting factor for LTM retention, memory performance in LTM tests should be better for items presented as part of a smaller relative to a larger sequences in the WM task – replicating prior results with simultaneous presentation (Forsberg et al., 2021a; 2022a; 2022b; see also, Forsberg et al., 2023, which included only sub-capacity set size sequences). However, because these previous studies used simultaneous presentation, the limiting factor may not be WM capacity, but instead, a shared perceptual bottleneck that is magnified for arrays with more items to-be-encoded at once. According to this alternative hypothesis, we should observe evidence against a set size effect on LTM under sequential presentation, which equates the concurrent perceptual bottleneck for smaller and larger sequences.

Our second aim was to better understand the mechanisms driving both WM and LTM primacy and recency effects by comparing serial position effects during the initial WM test and subsequent LTM test. Specifically, we explored the following pre-registered hypotheses:

Primacy Effect Hypotheses

$H_{Primacy1}$: If the WM primacy effect is driven by deep maintenance processes (e.g., participants actively holding early items in the focus of attention, or keeping them in a semantic form, which may be possible because of the availability of attention[2]), we would expect evidence for primacy effects in LTM.

$H_{Primacy2}$: If the WM primacy effect is driven by shallow processes (e.g., verbal rehearsal of early items), we would expect evidence against primacy effects in LTM.

Recency Effects Hypotheses

$H_{Recency3}$: If the WM recency effect is driven by deep maintenance processes (e.g., final items actively held in focus of attention, or kept in a semantic form, which may be possible because of the availability of attention, during the delay), we would expect to observe evidence for recency effects in LTM.

$H_{Recency4}$: If the WM recency effect is driven by shallow processes (e.g., less interference and decay compared to other items) we would expect evidence against recency effects in LTM.

Method

[2] In our pre-registered hypotheses, we had included items being held in the focus of attention as a sole example of a "deep" maintenance process. In the levels of processing theory, deep processing often refers to forming a conceptual or semantic representation. Our theoretical assumption was that for an item to enjoy any type of 'deeper' processing (including semantic), it needs to be held in the focus of attention. However, we acknowledge that holding information in the focus of attention, in itself, may not fit the classic definition of a deep maintenance process, and we have amended our hypotheses for clarity.

Transparency and Openness

The methods and analyses reported below were pre-registered on the Open Science Framework at [https://osf.io/43hsd/?view_only=01456942b0ff4c31baac6bfd96a844aa], except analyses labeled as 'exploratory'. Data, analysis code, and study materials are available at [https://osf.io/y3qxf/?view_only=cdd71f1f49e54011bb549145edc87ebe].We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study below.

Participants

Sample Size Determination. To determine our experimental sample size, we used Bayes Factors design analysis (BFDA; Schönbrodt & Stefan, 2018). We used a sequential sampling procedure with a pre-specified, minimum sample size of N = 80, and a maximum N of 140 participants. Our BFDA was based on 10,000 simulations for a within-subjects design (using a Bayes Factor (BF) > 3 as decision criteria). The BFDA tested whether we could detect evidence against an effect under the assumption that no effect exists in the population, using a non-directional within-subjects Bayesian t-test, with a minimum number of 80 participants. Assuming that there is no effect (d = 0.00), these simulations revealed that 84.9% of samples found evidence for H_0 (i.e., correctly finding evidence against an effect), 1.0% of samples found evidence for H_1 (i.e., a false positive), and 14.1% of samples were inconclusive. Next, we ran similar simulations under the assumption of a true effect of d = 0.59, based on the effect size (Cohen's d) of WM set size (set size 2 versus set size 6) on LTM performance (adjusted p(LTM)) from a similar prior study using simultaneous presentation (Forsberg et al., 2021a).

Using a BF > 6 as decision criteria revealed that, with 80 participants, 98.9% of simulated studies found evidence for H_1 (i.e., correctly finding evidence for an effect), while 1.1% of studies were inconclusive, and 0.0% of studies found evidence for H_0 (false negative). Finally, we specified that if evidence for or against the central hypothesis in our pre-registered analysis was inconclusive (defined as a BF between 0.33 and 3, thus failing to provide convincing evidence either for or against an effect), we would collect 20 more participants until the BF is > 3, or until we reach our maximum *N* = 140 (Rouder, 2014). This threshold was reached at the first stopping point, resulting in a final *N* of 80 participants.

Recruitment. We recruited participants online via Prolific.co, found to produce comparable results to in-person laboratory studies (Germine et al., 2012; Peer et al., 2017). Participants were pre-screened via Prolific to fit the following criteria: (1) Native speaker of English, (2) British, American or Canadian Nationality, (3) Normal or corrected-to-normal vision, (4) No cognitive impairment or dementia, (5) No language-related disorders, (6) Aged between 18 and 30 years, and (6) Approval rating of at least 90% on prior submissions at Prolific.co. The study was approved by the local Ethics (IRB) committee at the University of Missouri, and participants were compensated according to Prolific.co recommendations.

Pre-registered Pilot. As specified in our pre-registration, we first collected data from 10 pilot participants via Prolific.co, to ensure that the task difficulty level was appropriate (i.e., that it did not result in chance-level performance). Specifically, we checked the data from these initial 10 participants to ensure that average WM performance was greater than .55 for set sizes of two or four items among at least eight of the 10 participants, and to ensure that all data was saved as intended. These conditions were met, and we proceeded with data collection, with these 10 initial participants included in the final sample.

Exclusions. One participant was excluded and replaced under our pre-registered exclusion criteria, for taking two > 10-minute breaks within a memory trial. All participants performed over .55 accuracy in the easiest WM condition (Set Size 2), and therefore, no one was excluded for poor WM recognition performance.

Demographics. The mean age of the participants was 24.0 years (*SD* = 3.29, range 18–30 years; 55.0% female, 42.5% male, 1.25% other, and 1.25% 'prefer not to say', and 16.2% Asian, 3.75% Black or African American, 6.25% More than one Race, 72.5% White or European, 1.25% Other). The average experiment completion time was 24.8 minutes (*SD* = 8.0).

Procedure

Each participant completed three tasks in the following fixed order: (1) a WM probe-recognition task, (2) a one-minute arithmetic distraction task, and (3) a second probe-recognition memory task to assess LTM for previously untested items from the WM task. Figure 1 shows example trials for the WM (Panel A) and LTM (Panel B) phases. The crucial manipulation was the WM set size (i.e., the number of items presented in each trial), with two (SS2), four (SS4), or six (SS6) items per sequence. The general procedure was similar to that in previous studies (Forsberg et al., 2021a; 2022a; 2022b), with one key difference: each memory item was presented on its own (i.e., sequential presentation). Participants were informed that they would complete two memory tests but were not explicitly told that their memory for the WM items would be tested at a later point. We programmed the experiment using PsyToolkit (Stoet, 2010, 2017).

Working Memory (WM) Task. All memory stimuli were selected from the Microsoft Office 'Icons', and consisted of easily recognizable images (e.g., animals, fruits, and furniture).

The background was light grey, and the to-be-remembered items were presented in black (image resolution: 96 × 96 pixels). Each participant studied a total of 288 unique memory items in the WM task, at varying set sizes (two, four, or six items per sequence).

Each WM trial started with a 250 ms central fixation cross before the presentation of the first memory item. Items in each set were presented sequentially for 250 ms per item, with a 250 ms blank inter-stimulus interval separating each item**.** After all the items in the trial had been presented, there was a 2000 ms blank delay before the probe item and response options were presented, allowing comparison with previous studies using sequential presentation (see Forsberg et al., 2021a). On half of the trials at each set size, the probe item was a novel, previously unseen item. On the other half of the trials at each set size, the probe item was the same as one of the studied items from that sequence. The number of trials at SS2, SS4, and SS6 was 48, 24, and 16, respectively, to ensure that an equal number of 96 unique memory items was presented at each set size, resulting in a total of 88 WM test trials. For the 24 same probes in SS2 tests, half were drawn from each serial position. The 12 same probes in SS4 tests were evenly drawn from each of the four serial positions in the studied sequence (i.e., three same probes per serial position). The eight same probes for SS6 tests included two probes each from the first two serial positions and one probe from each remaining serial position (positions three through six) in the study sequence. The order of trials and the selection of items for each trial were randomized for each participant. Participants responded by clicking on one of the following options, which were presented on the screen along with the probe: ‘I'm Sure I Saw It’, ‘I Think I Saw It’, ‘I Guess I Saw It’, ‘I Guess I Didn't See It’, ‘I Don't Think I Saw It’, or ‘I'm Sure I Didn't See It’ (see Figure 1A).

Arithmetic Distraction Task. After completing all 88 WM trials, participants completed a one-minute distraction task. They were asked to verify arithmetic equations of the form a × b + c = d, where a, b, and c were random single-digit integers and d was equal to a × b + c or differed from that expression by ±1. Participants responded by clicking ‘Correct’ or ‘Incorrect’ on the screen. On average, participants attempted 14.5 problems (*SD* = 6.2, range 2 – 34) during this one-minute distraction task, and the average accuracy was 73.9% (*SD* = 18.9).

Long-Term Memory (LTM) Task. Lastly, participants completed the LTM task. Participants were asked to respond to 244 single probe items by indicating whether they had seen each item during the initial WM task, using the same response scale as in the WM task (see Figure 1B). To avoid repeated exposure, items that were probed in the WM task were not probed in the LTM task. The 244 LTM test probes consisted of 136 new items, 36 SS2 items, 40 SS4 items, and 60 SS6 items. At SS2, 18 items were probed for each serial position. At SS4 and SS6, 10 items were probed for each position.

Analyses

We report all pre-registered analyses and some additional exploratory analyses, which are labeled as ‘exploratory’. Statistical analyses were conducted within a Bayesian statistical framework, which confers many advantages over null hypothesis significance testing, such as the ability to quantify evidence for (or against) a null effect. Our primary analyses concerned the effects of WM set size on both WM and LTM accuracy. These analyses included (1) comparisons of response accuracy in the WM tests using hierarchical Bayesian logistic regression models in the brms package in R (Bürkner; 2017), and (2) recognition model-based estimates of the probability that an item from a given set size was held in WM (by the time of WM testing) or LTM (by the time of LTM testing) based on formulas from Forsberg et al.

(2021a). To ensure that our findings generalize when using different theoretical approaches to memory data (e.g., Williams et al., 2022), we also report measures of *d*-prime and *a*-prime for all analyses using *p*(WM) or *p*(LTM) measures (see Supplementary Materials, Section 1, for more details, and tables summarizing these values).

The hierarchical Bayesian logistic regression models leveraged trial-level response accuracy within each subject. Our model estimated the effect of WM set size (coded as a continuous factor) on memory performance (the parameter η (*eta)* in our model), accounting for the binary distribution of the data (correct or incorrect), using a Bernoulli distribution. We used a normally distributed prior for η (memory performance), specified by set_prior("normal(0,5)"). Participant identity and trial number were both included as random intercepts, to account for individual variation and trial variability. The dependent variable was Correct versus Incorrect responses (1 or 0) on each trial, and responses marked as guesses were coded as incorrect for this analysis. Such multilevel models allow modeling of data that take complex dependency structures into account and yield not only the mean but also a measure of the uncertainty of each parameter (the Bayesian Credible Interval), which conveys the range of values in which we can be certain, with a specified probability (e.g., 95%), that the "true" estimate of the parameter can be found within the population (Kruschke & Liddell, 2018).

Our recognition modeling was based on the analytical approach of Forsberg et al. (2021a). We estimated the proportion of items from a given encoding set size that was observable in memory at the time of (1) WM testing (*p*(WM)), and (2) LTM testing (*p*(LTM)). The *p*(WM) estimates were based on the observed rates at which participants correctly identified studied items (i.e., the hit rate, *h*) and incorrectly identified new items as old (i.e., the false alarm rate, *f*). The model for estimating *p*(WM) was based on formulas from Pashler (1988) that were applied to single-probe recognition test situations by Cowan et al. (2013), dubbed the "reverse Pashler" formula. This model assumes that when the probed item is in WM, participants respond correctly, and when it is not, they guess that the item is new with a certain rate (g). Then, the correct detection rate for old (studied) items, h, equals the probability that the probe item is in WM plus the probability that it is not in WM but that a correct "old" guess g is given:

$$h = p(\mathrm{WM}) + (1 - p(\mathrm{WM}))(g)$$

New items cannot match a representation in WM, so performance depends on the guessing rate, and an incorrect response (f) is made at the rate, $f = g$. By combining these formulas, it can be shown that:

$$p(WM) = \left(\frac{h - f}{1 - f}\right)$$

We used a similar formula to estimate the proportion of items from each WM sequence that were subsequently accessible in LTM. In the formula for p(LTM), l denotes that the hits (h) and false alarms (f) are based on correct and erroneous "old" recognition responses in the LTM tests:

$$p(LTM) = \left(\frac{hl - fl}{1 - fl}\right)$$

If the false alarm rates exceed the hit rate, p(WM) or p(LTM) becomes an implausible negative value, which likely reflects a combination of poor memory and unlucky guessing. In our pre-registration, we specified two ways to deal with theoretically implausible p(WM) or p(LTM) estimates. In the first approach, we adjusted such values to be plausible (such that a p(WM) corresponding to a WM capacity less than 1 item was adjusted to the equivalent of a capacity of one item, and negative p(LTM) was adjusted to 0). In the second approach, we excluded all participants with one such value in any set size. In this experiment, this affected three participants: one with a low *p*(WM) value and two with low *p*(LTM) values. We report results using both the adjusted values and the values for the second approach (excluding these three participants).

Using these model-based estimates of *p*(WM) and *p*(LTM), we also calculated the ratio of WM to LTM transfer at each set size. Comparisons of model-based estimates of recognition memory – *p*(WM), *p*(LTM), and *p*(LTM)/*p*(WM) – were based on Bayesian *t*-tests and ANOVA models implemented in the BayesFactor package in R (Morey & Rouder, 2015). When reporting Bayes Factors (BFs), we rely on the terminology proposed by van Doorn et al. (2021), in which a BF between 1 and 3 is considered *inconclusive* or *weak*, between 3 and 10 is considered *moderate*, and between 10 and 100 is considered *strong*. We refer to BFs >100 as *decisive* (Wetzels & Wagenmakers, 2012). However, these categorical verbal labels are subjective and should not be interpreted as definitive cut-off points (Tendeiro & Kiers, 2019; van Doorn et al., 2021). We report BFs for the effects hypothesis as BF_{10}, and those for the null hypothesis as BF_{01}, where $BF_{01} = (BF_{10})^{-1}$.

Finally, in addition to these primary analyses examining the effects of WM set size on recognition accuracy and the transfer of information from WM to LTM, we also examined whether primacy and/or recency effects in WM transferred to LTM in the most extreme set size condition (SS6). We focused on SS6 for these analyses because the six-item sequence afforded a longer list for comparisons of primacy and recency items to middle-position items, relative to the list lengths of SS2 and SS4 sequences.

Results

The Effect of Set Size on WM and LTM

WM Accuracy. WM accuracy (*p(*WM)) across set sizes is presented in Figure 2A. First, we explored the effect of set size on WM accuracy, using Hierarchical Bayesian logistic regression models. This analysis detected credible evidence for a set size effect with decreasing accuracy as set size increased (η=-0.40; SE=0.03, 95% CI [-0.46, -0.35]). The BF in favor of the model including set size was 3.14×10^{51} over a model not including this factor, indicating decisive evidence that the set size manipulation influenced WM accuracy (SS2: *M* = 0.95, *SD* = 0.23; SS4: *M* = 0.87, *SD* = 0.33; and SS6: *M* = 0.80, *SD* = 0.40). In addition to this atheoretical analytical approach, we performed an exploratory analysis on *p*(WM), to match the *p*(LTM) analysis reported below. We tested whether performance in the WM task varied as a function of WM set size using a Bayesian *t*-test, comparing the lowest set size (SS2) with the highest (SS6), using the BayesFactor package in R (Morey & Rouder, 2015). We found decisive evidence ($BF_{10} = 2.3 \times 10^7$) that WM retention (*p*(WM)) was better for items presented in lower set size sequences (SS2: *M* = 0.95, *SD* = 0.07) than for items presented at higher set size sequences (SS6: *M* = 0.84, *SD* = 0.19), when theoretically implausible values of *p*(WM) were adjusted. Similar results were found when data from the three participants with theoretically implausible values were excluded ($BF_{10} = 5.9 \times 10^6$). Results replicated with exploratory *d*-prime ($BF_{10} = 5.5 \times 10^{16}$) and *a*-prime ($BF_{10} = 5.4 \times 10^{10}$) measures (see Supplement for details).

***p*(LTM) Results.** Estimates of *p*(LTM) – the proportion of items from each encoding set size that were retained by the time of LTM testing – are depicted in Figure 2. Two separate pre-registered analyses addressed whether successful encoding of items in WM influenced subsequent LTM representations. First, we tested whether performance in the LTM task varied as a function of WM set size using a Bayesian *t*-test, comparing the lowest set size (SS2) with the highest (SS6). We found decisive evidence ($BF_{10} = 4.9 \times 10^5$) that LTM retention (*p*(LTM)) was better for items presented in lower set size sequences (SS2: *M* = 0.51, *SD* = 0.23) than for items presented at higher set size sequences (SS6: *M* = 0.42, *SD* = 0.23, see also SS4: *M* = 0.45, *SD* = 0.22), when theoretically implausible values of *p*(LTM) were adjusted. Similar patterns were found when data from the three participants with theoretically implausible values were excluded ($BF_{10} = 7.5 \times 10^4$). The second pre-registered analysis included set sizes of two, four, and six items as a continuous numeric variable and used the *generalTestBF* function. This analysis also revealed decisive evidence for a set size effect both when adjusting negative values ($BF_{10} = 6.0 \times 10^6$) and excluding them ($BF_{10} = 7.9 \times 10^5$). Results replicated with exploratory *d*-prime ($BF_{10} = 9.4 \times 10^8$) and *a*-prime ($BF_{10} = 6.3 \times 10^5$) measures (see Supplement for detailed values). These results replicate the findings of Forsberg et al. (2021a) and suggest that WM load during memory encoding constrained long-term learning, even when items were presented sequentially.

***p*(WM) and *p*(LTM) for only the final two items of each set.** By combining the set size manipulation with sequential presentation, two potential confounds were introduced. First, the average time between memory item and test was longer for the larger set sizes, because items earlier in the set were presented further from the test, temporally. Second, items presented in the final position may enjoy potential recency effects. At the shortest list length (SS2), the last item was tested on 50% of target-present test trials, whilst the last item contributed fewer target-present test trials at SS6, as items from the six different serial positions were probed. Therefore, in an exploratory analysis, we assessed the set size effect, only considering the final two items from each set size (i.e., both items for SS2, items 3 and 4 for SS4, and items 5 and 6 for SS6). For adjusted *p*(WM), evidence against a set size difference between SS2 and SS6 was found

(BF_{01} = 7.6; SS2 [*M* = 0.95, *SD* = .07], SS4 [*M* = .93, *SD* = .14], and SS6 [*M* = .95, *SD* = .16]), suggesting that set size effects were at least partially driven by these factors. However, for adjusted *p*(LTM), evidence for a difference between SS2 and SS6 was observed (BF_{10} = 25.7; SS2 [*M* = 0.51, *SD* = .23], SS4 [*M* = .45, *SD* = .22], and SS6 [*M* = .42, *SD* = .23]), suggesting that the LTM set size effect cannot be explained by a larger proportion of final-item trials for the larger WM set sizes.

WM Set Size Effects on the Ratio of Items Transferred from WM to LTM. Next, we tested the ratio of WM to LTM transfer, i.e., how many of the items encoded into WM made it into LTM, for each WM set size, using a similar method to Forsberg et al. (2021a). For each participant and WM set size, we assessed the proportion of items that were in memory at the time of WM testing (*p*(WM)) and LTM testing (*p*(LTM)), and using these values, we obtained a ratio of LTM to WM item presence (see Supplementary Material, Section 2 for further details). We tested whether this ratio differed across set sizes and found inconclusive evidence for an effect (BF_{10} = 1.5 for adjusted values, BF_{01} = 1.4 when negative values were excluded). Moderate evidence against a set size effect was found when using exploratory *d*-prime (BF_{01} = 5.6) and *a*-prime (BF_{01} = 5.5) measures (see Supplement for details). Thus, it is unclear whether the probability that items that were successfully encoded in WM were subsequently remembered in the LTM task differed between smaller and larger set sizes. The ratio values (using the adjusted approach) for items originally presented at SS2 were (*M* = 0.53, *SD* = 0.23), for SS4 (*M* = 0.49, *SD* = 0.21), and SS6 (*M* = 0.47, *SD* = 0.24).

Relation Between WM Capacity and LTM Performance

Next, we assessed correlations between an individual's overall *k* (WM capacity) and average LTM performance with a Bayesian correlation test. WM capacity for each set size was estimated for each participant in a hierarchical Bayesian working memory model, based on the "reverse-Pashler" formula (see Cowan et al., 2013; Rhodes et al., 2018). See the Supplementary Materials, Section 3 for further details. Average LTM performance was quantified as the average number of items a participant held in LTM, for each set size (p(LTM) × Set Size), such that average LTM performance equaled the mean of the following three values: ($p(\text{LTM}_{\text{SS2}}) \times 2$), ($p(\text{LTM}_{\text{SS4}}) \times 4$), and ($p(\text{LTM}_{\text{SS6}}) \times 6$). We found decisive evidence for a correlation ($rho = 0.43$, $BF_{10} = 1.6 \times 10^3$, adjusted; $rho = 0.42$, $BF_{10} = 1.0 \times 10^3$, excluded), indicating that participants with better WM capacity also performed better in the LTM test. Exploratory correlations between WM and LTM *d*-prime ($rho = 0.52$, $BF_{10} = 4.9 \times 10^{15}$) and *a*-prime ($rho = 0.50$, $BF_{10} = 2.2 \times 10^{14}$) measures suggested similar patterns.

Primacy and Recency Effects

Finally, we tested the presence of primacy and recency effects by comparing memory for the first serial position item to that of the two middle items (*primacy effect*), and of the last serial position item to the two middle items (*recency effect*), in the SS6 trials. Due to the limited trials per participant per position (1 or 2 trials per serial position in the WM test), we used accuracy data for these analyses instead of computing p(WM) and p(LTM). Following our pre-registration, we ran these analyses using two different approaches. First, in our more 'lenient' scoring approach, we scored a response as correct if it was correct, regardless of the participant's confidence level. Second, in a 'stricter' scoring procedure, all trials marked as 'guesses' were scored as incorrect. We report the outcome of both these analyses below. In the WM data, we found weak (BF01 = 2.85 with lenient scoring) and moderate (BF01 = 6.6 with strict scoring) evidence against a primacy effect, but decisive evidence for a recency effect ($BF_{10} = 7.15 \times 10^4$

with lenient scoring; $BF_{10} = 2.2 \times 10^5$ with strict scoring). In the LTM data, we found inconclusive evidence for a primacy effect ($BF_{10} = 1.00$ with lenient scoring; $BF_{10} = 1.02$ with strict scoring), but moderate (under strict scoring, $BF_{10} = 9.9$) or strong (under lenient scoring, $BF_{10} = 51.4$) evidence for a recency effect (see Figure 3).

Discussion

The results of Experiment 1 are aligned with past research observing a general set size (or 'list-length') effect, using various materials and paradigms (e.g. Oberauer et al., 2004; Sternberg, 1966). The results also replicated the findings of Forsberg et al. (2021a) and are aligned with a growing body of research suggesting that WM load during memory encoding constrains subsequent LTM retrieval of information (e.g., Čepukaitytė et al., 2023; Forsberg et al. 2021a; 2022a; 2022b; Fukuda & Vogel, 2019). Critically, this suggests that the results of a WM-to-LTM bottleneck in those prior studies using simultaneous presentation of items at encoding were not driven by a shared perceptual encoding bottleneck. This alternative account of the prior results cannot be reconciled with the fact that we obtained similar results under sequential presentation, which equated the concurrent perceptual load across encoding set sizes. However, interpretations of memory experiment results depend on assumptions related to the notion of serial vs. parallel or concurrent processing, as famously noted by, for example, Townsend (1990). In the context of our study, it is difficult to determine the extent to which sequential presentation of memory items promoted more serial processing. While items were displayed in a serial way, the extent to which participants allocated resources to specific items (i.e., maintaining all items, or focusing on rehearsing the earliest ones), especially during the 2000 ms delay, cannot be determined based on the results of this study alone. The results of the follow up analysis including only the two final items from each set size found no evidence for a set size effect on immediate WM performance, while clear evidence for a LTM effect was observed. The built-up WM load did not appear to influence immediate recall of the final two items from each set, which suggests that they were processed just before the test. However, this processing appeared insufficient to protect items from the higher set sizes from the decay introduced in the interval prior to the LTM test. Our encoding time for each item was 250 ms. We chose this rather brief timing to encourage participants to fully consolidate the LTM representation from the contents of WM, rather than forming a LTM representation from extended exposure to individual items, which may be less limited by the hypothesized WM limitations. Our brief encoding time was likely sufficient for participants to extract at least a gist representation of each item, given that rapid LTM gist extraction has been observed under similar (and sometimes faster) timescales for both verbal (e.g., Draine & Greenwald, 1998) and visual (e.g., Potter, 1976; Tatler et al., 2003; Thorpe et al., 1996) memoranda. Nevertheless, future studies should explore whether the set size effect may be reduced – or disappear entirely – if more time was given at encoding. Indeed, the amount of time available to process each item in simple span WM tasks predicts subsequent LTM for those items (Souza & Oberauer, 2017), though it remains unknown whether increasing the WM processing time per item would be sufficient to offset the encoding set size effect on subsequent LTM.

Moreover, we found strong correlations between individual differences in WM capacity and subsequent long-term retention of information. This supports suggestions that WM capacity limitations play a key role in the long-term learning of information (Cowan, 2023; Gathercole et al., 2006; Forsberg et al., 2021b). Our results highlight that to support learning, it is important to consider the amount of information that needs to be retained in WM within a given learning episode, in addition to the amount of concurrent information within a particular display.

Previous studies using a similar approach have indicated strong evidence for a set size effect on the LTM/WM ratio when items were presented simultaneously at encoding (Forsberg et al., 2021a; 2022a; 2022b), such that the transfer of information from WM to LTM appeared higher at smaller compared to larger set sizes. However, previous studies found weak evidence against a set size effect on the LTM/WM ratio when using sequential presentations of sub-capacity set sizes (i.e., two, three, or four items; Forsberg et al., 2023). In the present study, we found inconclusive effects of set size (with set sizes of two, four, or six items presented sequentially) on the LTM/WM ratio, and moderate evidence against a set size effect using exploratory d-prime and a-prime measures. Thus, additional research is needed to clarify whether set size influences the informational transfer from WM to LTM under sequential presentation conditions. Finally, further research is also needed to understand how WM load constrains longer-term retention of complex, bound information (Bartsch et al., 2019; Forsberg et al., 2023) as well as material that is more complex and relevant for learning in real-world settings.

Our second research question concerned whether serial position effects in WM transferred to LTM. We observed recency effects in both the WM and LTM tests of SS6 sequences, aligned with past literature (e.g., Monsell. 1978). This finding suggests that some processes that underlie the WM recency effect are also beneficial for LTM encoding. The WM recency benefit may be at least partly driven by deep maintenance processes (e.g., actively holding final list-position items in the focus of attention and employing elaborative strategies on those active representations during the brief delay prior to the WM test), rather than by shallow processes (e.g., verbal rehearsal). Such an explanation is consistent with the levels of processing framework (Craik & Lockhart, 1972), which predicts that deeper but not shallower WM maintenance processes should enhance subsequent LTM. However, the evidence regarding primacy effects in both WM and LTM were statistically inclusive, potentially reflecting the relative sparseness of the data for each serial position in Experiment 1. Therefore, we conducted an additional experiment to follow up on these inconclusive results.

Experiment 2

The purpose of this experiment was to follow up on the statistically inconclusive WM and LTM primacy effects in Experiment 1, by increasing the sample size and the number of SS6 trials, which allow the assessment of the entire serial position range.

Method

Transparency and Openness

The methods and analyses reported below were pre-registered on the Open Science Framework at [https://doi.org/10.17605/OSF.IO/5BGJ2][3], except analyses labeled as 'exploratory'. Data, analysis code, and study materials are available at [https://osf.io/y3qxf/?view_only=cdd71f1f49e54011bb549145edc87ebe]. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study below.

Participants

[3] Initially, we planned to use set size 5. However, after recruiting the planned 10 pilot participants, we noticed that performance in the WM task was close to ceiling levels (accuracy for 'same' trials at each serial position: 1 = 0.90, 2 = 0.96, 3 = 0.93, 4 = 0.91, 5 = 0.97). This might limit our ability to detect primacy and recency effects in the WM task, thus limiting our ability to meaningfully compare these patterns to primacy and recency effects in LTM task – which was the purpose of this follow-up experiment. Therefore, we amended our pre-registration to use set size 6 instead.

Sample Size Determination. As in Experiment 1, sample size determination was based on a BFDA (Schönbrodt & Stefan, 2018), using a sequential design with a pre-specified, maximum *N* of 120. We followed our pre-specified plan: First, 10 participants completed the study, and we looked at the data to ensure it was saved properly. Then, we recruited 70 more participants (total $N = 80$, as in Experiment 1). If, at this minimum sample size, the evidence for (or against) any of the main effects (memory type or serial position) or interactions (memory type × serial position) for either of our two pre-registered analyses were inconclusive (BF < 3 either for or against the hypothesis), we planned to add 40 participants (maximum $N = 120$). Based on this criterion, an additional 40 participants were added, and thus, the final sample size was $N = 120$.

Recruitment and Inclusion Criteria. Recruitment and inclusion criteria were similar to those in Experiment 1, with one additional exclusion criterion: participants who had completed Experiment 1 were not invited to participate in Experiment 2. Next, the performance-based exclusion criterion in Experiment 1 (<.55 accuracy in WM SS2 trials), was instead applied to SS6 trials, because this was the easiest – and only – set size condition in Experiment 2. One participant was excluded and replaced due to an average accuracy of less than .55 in the WM test and for taking long breaks in the LTM test.

Demographics. The mean age of the participants was 22.9 years ($SD = 3.2$, range 18–30 years; 68.8% female, 27.5% male, 2.5% other, and 1.3% ‘prefer not to say’, and 6.25% Asian, 1.25% Black or African American, 6.25% More than one Race, 82.5% White or European, 2.25% Other, 1.25% Prefer not to say). On average, participants completed the experiment in 20.0 minutes ($SD = 4.6$).

Procedure

The general procedure was similar to that in Experiment 1, with one key exception. Set size was consistent across all trials, such that all WM trials included six items, and each participant completed 60 WM trials. These 60 trials included 30 *different* trials (in which the probe item was a novel item), and 30 *same* trials (in which the probe was an item previously presented in the memory array). The 30 *same* trials featured an equal number of items (5 trials each) drawn from each of the six serial positions. After the WM phase, participants completed the arithmetic distraction task described in Experiment 1. During this one-minute distractor period, participants attempted 14.4 (SD = 5.4, range 2 – 29) problems, and the average accuracy rate was 77.4% (SD = 13.0). Finally, each participant completed 180 LTM trials, consisting of 90 same trials (15 trials for each of the six WM serial positions), and 90 different trials, in which a new, unstudied item was presented. As in Experiment 1, none of the items in the LTM tests had also appeared as probes in the WM tests.

Results

Primacy Effects: Not in WM, but in LTM

We tested whether there were primacy effects by comparing recognition accuracy in WM and LTM for items from the first studied position relative to items from the third and fourth studied positions in a sequence. Similar to in Experiment 1, we used accuracy data for these pre-registered primacy and recency analyses. We specified a 2 (Memory Test: WM versus LTM) × 2 (Serial Position: First versus Middle) Bayesian ANOVA model in the BayesFactor package (Morey & Rouder, 2015), with memory accuracy (regardless of confidence level expressed) as the dependent variable. We observed decisive evidence for a difference in memory test type ($BF_{10} = 2.0 \times 10^{35}$), as WM accuracy was higher than LTM accuracy. There was inconclusive e vidence against an overall primacy effect ($BF_{01} = 1.8$), but there was moderate evidence for an interaction ($BF_{10} = 4.4$). To follow up on this interaction, we performed two exploratory Bayesian *t*-tests, which indicated moderate evidence *against* a WM primacy effect ($BF_{01} = 8.4$), and decisive evidence for an LTM primacy effect ($BF_{10} = 256.1$; see Figure 4). See Supplement (Section 1) for serial position patterns for *p*(WM), p(LTM), *d*-prime, and *a*-prime.

Recency Effects: In both WM and LTM, but Stronger in WM

Next, we used a similar approach to explore recency effects (comparing memory for the middle two items vs. the final two serial position items). We observed decisive evidence for a difference in memory test type ($BF_{10} = 3.8 \times 10^{69}$), decisive evidence for an overall recency effect ($BF_{10} = 3.5 \times 10^{10}$), and strong evidence for an interaction ($BF_{10} = 20.97$). Exploratory follow-up *t*-tests found decisive evidence for a WM recency effect ($BF_{10} = 2.9 \times 10^{9}$) and strong evidence for a LTM recency effect ($BF_{10} = 57.5$). The interaction indicated that the recency effect was more substantial in the WM test than in the LTM test (see Figure 4).

WM to LTM Loss of Information by Position

We estimated the amount of information that was lost between the WM and LTM test for each serial position, by subtracting the average *p*(LTM) from *p*(WM), for each participant and each serial position. We used the same approach to negative *p*(WM) and *p*(LTM) values as outlined for Experiment 1 previously, resulting in an adjustment of 15 and 29 *p*(WM) and *p*(LTM) values, respectively. An exploratory ANOVA suggested a continuous increase in informational loss from serial position one through six ($BF_{10} = 2.9 \times 10^{7}$; see Figure 5, Panel C). Evidence for serial position effects were also observed using exploratory *d*-prime ($BF_{10} = 2.8 \times 10^{11}$) and *a*-prime ($BF_{10} = 1.2 \times 10^{3}$) measures (see Supplement for details).

WM/LTM Ratio Transfer by Position

To follow up on this finding, we estimated the LTM/WM ratio for each participant, for each serial position (using a similar approach to that outlined for set size ratios in Experiment 1). An exploratory ANOVA suggested that the WM to LTM information transfer occurred at different rates from different serial positions ($BF_{10} = 4.1 \times 10^3$). Transfer to LTM appeared most effective for the first item in the sequence followed by the second item, while transfer at positions three through six appeared consistent (see Figure 5, Panel D). Moderate evidence for serial position effects was also observed using exploratory *d*-prime ($BF_{10} = 7.7$) and *a*-prime ($BF_{10} = 4.7$) measures (see Supplement for details).

Discussion

In Experiment 2, we aimed to test our pre-registered hypotheses regarding the transfer of WM primacy and recency effects to LTM. We assumed that if these effects did transfer, this would suggest that the mechanisms allowing the WM primacy and/or recency boost relied on a deep, rather than a shallow, process. In Experiment 2, we observed contrasting evidence regarding primacy effects, with moderate evidence against primacy effects in WM but decisive evidence for these effects in LTM. Although recency effects were found in both WM and LTM, they were more prominent in the WM tests. We will return to the theoretical implications of these unexpected findings in the general discussion.

Our WM findings mirror a recent study using a partial-report visuospatial WM test in which no primacy effects, but strong recency effects, were observed (see Exp. 3, McAteer et al., 2023). Similar patterns have been found for item memory for unfamiliar faces and nonwords (e.g., Ward et al., 2005). Observing evidence for both primacy and recency effects in LTM also fits with previous findings (Ebbinghaus, 1885/1913). However, our study provides novel insights into the informational transfer between WM and LTM at different serial positions, and we discuss the implications of these results below.

Specifically, we observed a differential informational transfer from WM to LTM based on the serial position of an item in the study sequence. This differential informational transfer was detected both when considering the ratio of transfer from WM to LTM (that is, out of the information participants could hold in WM at different serial positions, how much was transferred to LTM), and the loss of information (quantified by subtracting LTM from WM proportional accuracy at each serial position; see Figure 3). For example, almost twice as much information was lost between the WM and LTM phases from items presented last, compared to items presented first in the sequence (Figure 5, Panel C). In other words, across the course of a learning sequence of six items, more information was lost from the earlier list items than the later list items by the time of the WM test (strong recency effect in WM, but no primacy effect, see Figure 5A). However, the later-presented items in the sequence exhibited accelerated loss beyond the WM test, by the time of LTM testing. While overall LTM was better for items either presented first or last in the sequence, relative to middle-position items (see Figure 5B), the patterns of information loss across serial positions appeared linear (see Figure 5C). This suggests that a first-presented item that was successfully held in WM was more likely to also be recalled in LTM, compared to a last-presented item that was also successfully held in WM.

If STM/WM and LTM were one system (e.g., Brown et al., 2007; McElree, 2006; Nairne, 2002; Surprenant & Neath, 2009), relying on identical mechanisms, one might expect the shape of the LTM accuracy serial position curve to mirror the WM accuracy curve, with consistent LTM forgetting across all six serial positions. Instead, we observed strong evidence for an LTM primacy effect which was not present in the WM test, which suggests that the cognitive processes that supported the LTM primacy effect (i.e., better performance for the first item in a sequence, compared to the middle items) were not beneficial during the WM test. This indicates that the processes that underpinned relative differences in LTM recall between the first and the middle items did not create a similar pattern of differential recall in WM. Moreover, while recency effects were observed both in WM and LTM, the WM recency effects were more substantial than the LTM recency effects. Thus, processes that supported WM recency benefits seemed to transfer to an LTM benefit only partially. We discuss the theoretical implications of these results in the general discussion below.

General Discussion

The two studies reported here suggested that (1) WM capacity plays a key role in what can be encoded into LTM, irrespective of perceptual encoding bottlenecks, and (2) serial positions associated with more successful WM encoding were not associated with better LTM, suggesting that WM and LTM success depended, at least partially, on different mechanisms. Hence, our findings support the notion that the overall amount of information that can be encoded into LTM depends on what can be maintained in a limited WM capacity store at the time of encoding, but also that distinct mechanisms determine which items within a studied sequence are likely to be successfully recalled at WM and LTM test, respectively.

WM as an Encoding Bottleneck for LTM Retention

The results of Experiment 1 add to a growing body of research suggesting that WM load at encoding constrains subsequent LTM retrieval of information (e.g., Čepukaitytė et al., 2023; Forsberg et al. 2021a; 2022a; 2022b; Fukuda & Vogel, 2019). The results highlight that WM capacity limitations play a key role in long-term learning of information, beyond perceptual bottlenecks. Indeed, because each item was presented individually in the present study, the only difference between set size conditions was the number of items that participants needed to keep in WM. Moreover, we found strong correlational evidence for a relationship between WM capacity and subsequent long-term retention of information. Combined, these findings contribute to evidence seeking to understand the link between WM capacity and learning. Specifically, if the WM system acts as an encoding bottleneck for longer-term retention of information, this may partially explain the well-established link between individual differences in WM capacity and educational outcomes (Gathercole et al., 2004). Moreover, the findings underscore the crucial role of WM capacity in long-term learning (for a discussion, see Forsberg et al., 2021b). According to this account, the education-related advantages of individuals with higher WM capacity may stem partly from their ability to retain comparatively more information in WM at once, which promotes better LTM transfer, while also being closely related to fluid intelligence (Conway & Kovacs, 2013).

Our results suggest that to support learning, it is important to consider the amount of information that needs to be retained in WM within a learning episode in addition to how much information is presented at once. Our findings also highlight the importance of considering individual differences in how much information can be retained. However, discrepancies in primacy and recency patterns in Experiment 2 suggested differences in the transfer of information between WM and LTM depending on where in the sequence information was presented, which we discuss next.

Primacy and Recency Effects in WM and LTM: Shallow vs. Deep Processing Mechanisms

In our pre-registered hypotheses, we made predictions regarding whether WM primacy and recency effects would *transfer* to the LTM test – if they did, it would be an indication that the mechanism allowing the WM primacy/recency boost relied on a deep, rather than a shallow, process (see Craik & Lockhart, 1972; Shiffrin & Atkinson, 1969). Based on past literature, we expected shallow processes to help in the short-term and deeper processes to help both in the shorter and longer term (but for evidence that depth of processing may exert minimal influence in WM tests, see Rose et al., 2010). However, some evidence suggests that shallow level processes, such as maintenance rehearsal, may affect long-term recognition memory for item information (see Greene, 1987, for a review).

Traditional conceptualizations of elaborative rehearsal include encoding the material based on its meaning or its emotional tone (Craik & Lockhart, 1972). Typically, deep elaboration of the semantic meaning of an item during its encoding benefits long-term retention for that item (Craik & Lockhart, 1972; Craik & Tulving, 1975; Roediger et al., 2002). Although some studies have shown that deeper processing also benefits WM performance (Loaiza et al., 2011), others have found no benefit of deep elaboration on WM tests, at least for verbal stimuli encoded intentionally (Bartsch et al., 2019; Mazuryk & Lockhart, 1974; Rose et al., 2010). Rose and Craik (2012) attributed the lack of a depth of processing effect on WM performance for short word lists to participants' tendency to actively refresh items throughout the study phase to keep the items accessible for immediate testing purposes. When participants encoded items incidentally in supra-capacity conditions, the classic levels of processing effect was obtained in immediate free recall tests (Rose & Craik, 2012). Thus, participants probably rely more often on shallow (e.g., rehearsal) rather than deep processes to strengthen WM performance under intentional learning conditions, like ours. However, at least some of the time, participants probably engaged in more elaborate rehearsal of some of the items in our procedure, benefitting LTM for those items. Indeed, in the studies by Rose et al. (2010) and Rose and Craik (2012), although deep processing (i.e., elaborating on items' meanings) conferred little advantage in immediate WM tests, it enhanced LTM recall performance compared to shallow processing (i.e., focusing on the visual or phonological features of the items).

In our procedure we used distinct, common visual objects, presented quickly (250 ms per item). This was likely sufficient time for participants to encode the meaning of each item and thus assign a semantic label to each object (for evidence of rapid gist extraction of visual stimuli, see Potter, 1976; Tatler et al., 2003; Thorpe et al., 1996). Although verbal rehearsal of these common objects was possible, the limited presentation time may have made elaborating on these item meanings challenging, and participants may have prioritized the perceptual features of the visual stimuli, a typically shallow level of processing (e.g., Craik & Lockhart, 1972).

However, Plancher and Barrouillet (2020) suggested that in a WM context, *articulatory rehearsal* constitutes more shallow rehearsal of information, whereas *attentional refreshing* is similar to elaborative encoding processes (but see evidence against this claim; Bartsch et al., 2018; 2019; Loaiza & Camos, 2018). Articulatory rehearsal is the verbal repeating of memoranda either overtly or covertly (Baddeley, 2012; Elliott et al. 2021), known to be beneficial in WM tasks (e.g., Forsberg et al., 2020). Attentional refreshing is a maintenance mechanism thought to reactivate information in WM using attention (Barrouillet et al. 2009; Camos et al. 2018; Valentini & Vergauwe, 2023), such as by thinking of an item again. While additional opportunities for attentional refreshing appear associated with improved WM (see Camos et al., 2018) and long-term retention (e.g., Camos & Portrat, 2015; Jarjat et al., 2018; Souza & Oberauer, 2017), verbal rehearsal appeared less beneficial for LTM (e.g., Loaiza & McCabe, 2012, see also Camos & Portrat; 2015).

Additionally, vocalizing list items may result in poorer recall of items at the beginning but better recall at the end of a list (Dauphinee et al., 2024). However, Neath et al. (1993) displayed images of snowflakes – which are presumably difficult to verbally rehearse – and tested memory at varying retention intervals (from 0 to 10 seconds). In their shortest retention interval (0 seconds), a strong recency effect was observed, but no primacy effect. At their longest retention interval (10 seconds), there was a weak primacy effect but no recency effect, further highlighting the complexity of the interactions between retention duration and memory maintenance mechanisms.

Finally, we note that recent research which explored whether the depth-of-encoding effect varied across different retention intervals (0 to 18 seconds), found that depth-of-encoding effect may occur during the initial encoding of items, but without differential forgetting across different retention intervals (Lawrence et al., 2024). To conclude, while our methodological approach and results do not allow us to distinguish between the specific verbal, attentional, or semantic mechanisms that were used, they provide evidence for a difference in the efficiency of such mechanisms for immediate (WM) and delayed (LTM) recall (cf., Rose et al., 2010; Rose & Craik, 2012). Future studies may assess fine-grained mechanisms more directly for example by preventing verbal rehearsal using articulatory suppression. Next, we consider the differential WM and LTM primacy and recency effects through this theoretical lens.

Recency Effects Appear in WM and LTM, but are More Prominent in WM. In Experiment 2, we observed a comparatively stronger recency effect in the WM tests, compared to the LTM tests. That is, although the final list-position items were better remembered than middle list-position items in both the WM and LTM tests, the relative difference was greater in the WM test. In other words, the recency effect transferred partially from WM to LTM. This pattern may suggest that this effect was driven by at least two separate processes in WM; one that was 'deep' and helped LTM encoding, and one that was shallower (e.g., verbal rehearsal). Moreover, our design included a longer (2000 ms) gap after the final WM item, to allow comparison with previous studies. As such, participants had more time to process this particular item, while also presumably rehearsing the previous items, and this final item might have enjoyed a perceptual benefit, as it was not followed by another item. This longer gap may contribute to the performance boost for the final item observed in the study, in a way that is not directly attributable to the recency effect.

While our study cannot clarify the mechanisms which supported the WM recency effect, factors like not being followed by another item, and having a longer break (2000 ms) before the next attention-demanding display, are likely to play a role. However, we note that the fifth item in the sequence was comparatively better remembered than middle sequence items (see Figures 3 and 4), which suggests that the WM recency effect cannot be fully explained by the extended gap after the final item. Overall, our findings suggest that some of the factors that support WM recency are less beneficial for subsequent LTM retrieval, and future research exploring factors like post-stimulus masking may help dissociate the mechanisms that support WM and LTM respectively. Moreover, shallow maintenance like rehearsal has been found to boost LTM recognition in some circumstances (e.g., Naveh-Benjamin & Jonides, 1984), which could have been the case here for the recency effect. Finally, we note that different strategies between participants, and even trials are possible. Indeed, similar performance patterns may be underpinned by different underlying processes or strategic approaches. More fine-grained experimental manipulations are needed to explore this further (see Chooi & Logie, 2020; Overkott et al., 2022; Valentini & Vergauwe, 2023). Moreover, foreknowledge and expectations about the memory test may influence recall (Dames & Popov, 2023) and serial position curves in some contexts (see Guitard & Saint-Aubin, 2022; Guitard et al., 2020).

Nevertheless, the fact that we detected a recency effect in LTM tests at all is worth highlighting. Classic inferences about serial position effects have relied primarily on verbal free recall tasks (Craik, 1970; Murdock, 1962), in which recency effects are attenuated by delay (for a review, see Kahana, 2017). However, in delayed free recall tasks, participants often initiate the recall period with primacy items (Healey et al., 2014), and strong temporal (Kahana, 1996) and semantic (Howard & Kahana, 2002) clustering effects in recall dynamics (cf., Kahana, 2020) likely restrict participants' probability of transitioning to recency items. This likelihood of transitioning to end-of-list items in delayed free recall decreases as output interference from earlier recalls increases (Cowan et al., 2002). Accordingly, the very nature of the free recall task may have masked why numerous prior studies have failed to find recency effects in LTM tests. In recognition procedures, by contrast, recency effects do occasionally show up in LTM tests, with enhanced memory for end-of-list relative to middle-of-list items, at least under incidental encoding conditions (Jiang & Cowan, 2020). Our results replicate these earlier recognition results. Although participants were aware that their memory was being tested in our procedure (as they completed a WM recognition test at the end of each study trial), they may have been encoding items for the LTM test more "incidentally," given that they were not explicitly informed about the LTM tests prior to the study phase. Nevertheless, participants in our procedure at least attempted to intentionally encode the items at some point (i.e., for WM tests), and as such, our procedures may reflect more of a mixture of intentional versus incidental encoding than the more fully incidental encoding procedures of Jiang and Cowan (2020).

Primacy Effects in LTM but not in WM. In Experiment 1, evidence regarding a LTM primacy effect was inconclusive, likely due to insufficient trials and participants. However, Experiment 2 demonstrated evidence *against* a primacy effect in WM (i.e., the first item in the list was not better remembered than the middle items), but evidence *for* a primacy effect in the LTM test, in which the first-presented item in the WM phase was on average better retained than middle list-position items (see Figure 3). To understand this intriguing pattern, we consider two possible explanations. First, the very first item in the sequence may have been encoded using a particular strategic or automatic process, that was not beneficial for WM recall – when compared to the middle-sequence items – but that was highly beneficial in supporting long-term retention for such items. However, although various established theoretical mechanisms may account for this LTM primacy effect (such as attentional refreshing), it seems surprising that such a mechanism would not *also* help in the WM test (see Camos et al., 2018; Cowan, 1992; Vergauwe, & Langerock, 2017; Raye et al., 2007).

Alternatively, it might be that some processes that support WM success for the *middle list-position items* are more beneficial for WM than for LTM retrieval. To quantify both primacy and recency effects, we used memory for the middle items as the crucial contrast. Thus, it is possible that a shallow maintenance process allowed participants to maintain the middle list-position items quite well in the WM test but failed to result in deep LTM encoding. For example, neural evidence suggests that active maintenance of a stimulus representation may not be necessary for its short-term retention (Lewis-Peacock et al., 2012). This position is also consistent with transfer-appropriate processing theory (Morris et al., 1977). According to this theory, the degree to which encoding or maintenance processing depth matters depends on the extent to which the subsequent memory test requires deep processing. Shallow encoding and/or maintenance processes like rehearsal can yield strong WM performance but are not suitable for long-term retention.

A Dissociation between WM and LTM Processes. Memory for an item may be characterized by both its *storage strength* and its *retrieval strength*, the latter representing the current ease of access to the item, given the current cues (see Bjork & Bjork, 2020). Although our WM and LTM tests were identical and required participants to respond to one probe only, in the WM test, participants were comparing the probe item to however many items they were able to hold in WM from the recently presented sequence of six items. In the LTM test, they compared the probe item to *all* the items that they saw during the WM phase. Thus, the differences in serial memory patterns between our WM and LTM phases may be caused by differences in ease of access during the memory test, for items presented at different serial positions during the WM phase.

A case can be made that retrieval conditions differ between the WM and the LTM tests, as they may impose different degrees of distraction (e.g., Bjork & Whitten, 1974). Either way, these results indicate a dissociation between *processes* that support WM and LTM retrieval at different serial positions, and as such, appear to support accounts that view WM and LTM as at least partially separate systems, in which memory performance is underpinned by at least partially separate processes (cf., Rose & Craik, 2012). As we tentatively afford yet another life to STM, we acknowledge that various arguments may be made against this interpretation (see Surprenant & Neath, 2009; "*The nine lives of Short-Term Memory*").

WM-to-LTM informational transfer: The role of time

Finally, we discuss the informational transfer across all six serial positions. At first glance, LTM performance was superior for the first and final list-position items (compared to all middle list-position items). This finding may fit with either theories that emphasize that the length of time an item spends in WM or those that emphasize that the number of times an item is refreshed, will support LTM encoding. For example, Craik (1970) assumed a model in which to-be-remembered items enter a limited capacity short-term store and remain there until they are replaced by new, incoming items, and an item's strength of registration in the long-term store is a direct function of how long it stayed in the short-term store. Looking at the information loss between WM and LTM in our study, we observe a clear linear increase, such that the amount of information that is lost between WM and LTM increased as serial positions increased (Figure 3, panel D). In other words, out of the information that was successfully held in WM, more were transferred to LTM from the earlier, compared to the later, positions. However, it is possible (assuming an average WM capacity of around four items) that the first item was overwritten in the WM test (due to overcrowding or interference caused by subsequent items at the time of response), but still spent significant time being actively maintained in the short-term memory store, which resulted in this LTM advantage.

Practical Implications

While the result of Experiment 1 reaffirm that WM capacity is a limiting factor in overall, subsequent LTM success, Experiment 2 supports accounts that view WM and LTM as at least partially separate systems, in which performance is underpinned by at least partially separate mechanisms. Indeed, the differential WM and LTM primacy and recency patterns have several interesting practical implications. First, they suggest that items that are presented in a comparatively less favorable position for immediate recall (first item) may be particularly well recalled at a later test**.** Although it is well established that information tends to decay from memory over time (Rivera-Lares et al., 2022), our results suggest that sometimes, items at positions associated with relatively mediocre levels of initial recall may enjoy boosted recall when tested after a longer delay. In other words, information presented in positions associated with more memory challenges a few seconds after being presented may be better remembered later on (c.f. the idea of 'desirable difficulties' in learning, Bjork & Bjork, 2020). Indeed, the LTM primacy effect – and the relatively small informational loss between WM and LTM at this first position – might be a result of the participant's attentional efforts to protect the initial item from the interference caused by the five following items.

Similarly, our findings illustrate that the ratio of transfer from WM to LTM was highest for the first item in a sequence and decreased for the subsequent items. However, in terms of the absolute LTM performance, while both initial and final items were better remembered than middle items, the final item also enjoyed superior WM recall, which likely confers benefits in a learning context, since the learner would be better able to use that information immediately. While these research findings should be replicated to include more ecologically valid, and informationally complex materials, and with a longer time window – with and without distraction – between the learning and the final test phase, they provide an interesting insight into discrepancies between WM and LTM, depending on the serial position of the information during encoding.

References

Adam, K. C., Vogel, E. K., & Awh, E. (2017). Clear evidence for item limits in visual working memory. *Cognitive Psychology*, *97*, 79-97 https://doi.org/10.1016/j.cogpsych.2017.07.001

Atkinson, A. L., Oberauer, K., Allen, R. J., & Souza, A. S. (2022). Why does the probe value effect emerge in working memory? Examining the biased attentional refreshing account. *Psychonomic Bulletin & Review*, *29*(3), 891-900. https://doi.org/10.3758/s13423-022-02056-6

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). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3

Azizian, A., & Polich, J. (2007). Evidence for attentional gradient in the serial position memory curve from event-related potentials. *Journal of Cognitive Neuroscience*, *19*(12), 2071-2081. 10.1162/jocn.2007.19.12.2071

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1-29. https://doi.org/10.1146/annurev-psych-120710-100422

Barrouillet, P., & Camos, V. (2015). *Working memory: Loss and reconstruction*. London, UK: Psychology Press. https://doi.org/10.4324/9781315755854

Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working memory span development: a time-based resource-sharing model account. *Developmental Psychology*, *45*(2), 477. https://doi.org/10.1037/a0014615

Bartsch, L. M., Loaiza, V. M., & Oberauer, K. (2019). Does limited working memory capacity underlie age differences in associative long-term memory? *Psychology and Aging*, *34*(2), 268–281. https://doi.org/10.1037/pag0000317

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. https://doi.org/10.1016/j.neuroimage.2019.06.028

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*, 796-808.

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied research in Memory and Cognition*, *9*(4), 475. https://doi.org/10.1016/j.jarmac.2020.09.003

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, *6*(2), 173–189. https://doi.org/10.1016/0010-0285(74)90009-7

Broadbent, D. E. (1958). Perception and communication. New York: Oxford University Press. https://doi.org/10.1037/10037-000

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

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1-28. 10.18637/jss.v080.i01

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

Camos, V., Johnson, M., Loaiza, V., Portrat, S., Souza, A., & Vergauwe, E. (2018). What is attentional refreshing in working memory?. *Annals of the New York Academy of Sciences*, *1424*(1), 19-32. https://doi.org/10.1111/nyas.13616

Capitani, E., Della Sala, S., Logie, R. H., & Spinnler, H. (1992). Recency, primacy, and memory: Reappraising and standardising the serial position curve. *Cortex*, *28*(3), 315-342. https://doi.org/10.1016/S0010-9452(13)80143-8

Čepukaitytė, G., Thom, J. L., Kallmayer, M., Nobre, A. C., & Zokaei, N. (2023). The relationship between short-and long-term memory is preserved across the age range. *Brain Sciences*, *13*(1), 106. https://doi.org/10.3390/brainsci13010106

Chen, Z., & Cowan, N. (2009). How verbal memory loads consume attention. *Memory & Cognition*, *37*, 829-836. https://doi.org/10.3758/MC.37.6.829

Chooi, W. T., & Logie, R. (2020). Changes in error patterns during n-back training indicate reliance on subvocal rehearsal. *Memory & Cognition*, *48*, 1484-1503. https://doi.org/10.3758/s13421-020-01066-w

Conway, A. R., & Kovacs, K. (2013). Individual differences in intelligence and working memory: A review of latent variable models. *Psychology of Learning and Motivation*, *58*, 233-270. https://doi.org/10.1016/B978-0-12-407237-4.00007-4

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. https://doi.org/10.1037/xlm0000954

Cowan, N., & Guitard, D. (2024). Encoding colors and tones into working memory concurrently: A developmental investigation. *Developmental Science*, e13552. **https://doi.org/10.1111/desc.13552**

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. https://doi.org/10.1037/0033-2909.104.2.163

Cowan, N. (1992). Verbal memory span and the timing of spoken recall. *Journal of Memory and Language*, *31*(5), 668-684. https://doi.org/10.1016/0749-596X(92)90034-U

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87-114. Doi: 10.1017/S0140525X01003922

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, *24*, 1158-1170. https://doi.org/10.3758/s13423-016-1191-6

Cowan, N. (2019). Short-term memory based on activated long-term memory: a review in response to Norris 2017. *Psychol. Bull.* 145(8):822–47. https://doi.org/10.1037/bul0000199

Cowan, N. (2023). Working memory and child development with its windfalls and pitfalls. *Memory in Science for Society*: There Is Nothing As Practical As a Good Theory, 215. 10.1093/oso/9780192849069.003.0009

Cowan, N., AuBuchon, A. M., Gilchrist, A. L., Ricker, T. J., & Saults, J. S. (2011). Age differences in visual working memory capacity: Not based on encoding limitations. *Developmental Science*, *14*(5), 1066-1074. https://doi.org/10.1111/j.1467-7687.2011.01060.x

Cowan, N., Bao, C., Bishop-Chrzanowski, B.M., Costa, A.N., Greene, N.R., Guitard, D., Li, C., Musich, M.L., & Ünal, Z.E. (2024). The relation between attention and memory. *Annual Review of Psychology, 75,* 183-214. https://doi.org/10.1146/annurev-psych-040723-012736

Cowan, N., Blume, C. L., & Saults, J. S. (2013). Attention to attributes and objects in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 731-747. 10.1037/a0029687

Cowan, N., Blume, C. L., & Saults, J. S. (2013). Attention to attributes and objects in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 731-747. https://doi.org/10.1037/a0029687

Cowan, N., Saults, J. S., Elliott, E. M., & Moreno, M. V. (2002). Deconfounding serial recall. *Journal of Memory and Language*, *46*(1), 153–177. https://doi.org/10.1006/jmla.2001.2805

Craik, F. I. (1970). The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*, *9*(2), 143-148. https://doi.org/10.1016/S0022-5371(70)80042-1

Craik, F. I. (2020). Remembering: An activity of mind and brain. *Annual Review of Psychology*, *71*, 1-24. DOI: 10.1146/annurev-psych-010419- 051027

Craik, F.I.M., and Lockhart, R.S. (1972). Level of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*,6 7 1. https://doi.org/10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Dames, H., & Popov, V. (2023). When does intent matter for memory? Bridging perspectives with Craik. *Journal of Experimental Psychology: General*, *152*(11), 3300–3309. https://doi.org/10.1037/xge0001486

Dauphinee, I., Roy, M., Guitard, D., Yearsley, J. M., Poirier, M., & Saint-Aubin, J. (2024). Give me enough time to rehearse: presentation rate modulates the production effect. *Psychonomic Bulletin & Review*, 1-12. https://doi.org/10.3758/s13423-023-02437-5

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: empirical and computational investigations of recency effects. *Psychological Review*, *112*(1), 3. https://doi.org/10.1037/0033-295X.112.1.3

Draine, S. C., & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General, 127*(3), 286-303. https://doi.org/10.1037/0096-3445.127.3.286

Ebbinghaus, H. (1885/1913) Memory: a contribution to experimental psychology. Translated by H.A. Ruger and C.E. Bussenius. Teachers College, Columbia University, New York. (Originally in German, Ueber das gedächtnis: Untersuchen zur experimentellen psychologie). 10.5214/ans.0972.7531.200408

Elliott, E. M., Morey, C. C., AuBuchon, A. M., Cowan, N., Jarrold, C., Adams, E. J., ... & Voracek, M. (2021). Multilab direct replication of Flavell, Beach, and Chinsky (1966): Spontaneous verbal rehearsal in a memory task as a function of age. *Advances in Methods and Practices in Psychological Science*, *4*(2), https://doi.org/10.1177/25152459211018187

Fischler, I., Rundus, D., & Atkinson, R. C. (1970). Effects of overt rehearsal procedures on free recall. *Psychonomic Science*, *19*, 249-250. https://doi.org/10.3758/BF03328801

Forsberg, A. (2021, March 10). The effect of Working Memory Encoding on Subsequent Long-Term Memory Representations: Sequential Presentation. https://doi.org/10.17605/OSF.IO/B6RCD retrieved from https://osf.io/43hsd/?view_only=01456942b0ff4c31baac6bfd96a844aa

Forsberg, A. (2021, July 9). The effect of Working Memory Encoding on Subsequent Long-Term Memory Representations: Sequential Presentation - Experiment 2. https://doi.org/10.17605/OSF.IO/5BGJ2

Forsberg, A. (2024, March 25). Differential Information Transfer and Loss between Working Memory and Long-Term Memory Across Serial Positions. Retrieved from https://osf.io/43hsd/?view_only=01456942b0ff4c31baac6bfd96a844aa

Forsberg, A., Adams, E. J., & Cowan, N. (2021b). The role of working memory in long-term learning: Implications for childhood development. In *Psychology of Learning and Motivation* (Vol. 74, pp. 1-45). Academic Press. https://doi.org/10.1016/bs.plm.2021.02.001

Forsberg, A., Guitard, D., & Cowan, N. (2021a). Working memory limits severely constrain long-term retention. *Psychonomic Bulletin & Review*, *28*, 537-547. https://doi.org/10.3758/s13423-020-01847-z

Forsberg, A., Guitard, D., Adams, E. J., Pattanakul, D., & Cowan, N. (2022a). Children's long-term retention is directly constrained by their working memory capacity limitations. *Developmental Science*, *25*(2), e13164. https://doi.org/10.1111/desc.13164

Forsberg, A., Guitard, D., Adams, E. J., Pattanakul, D., & Cowan, N. (2023). Working memory constrains long-term memory in children and adults: memory of objects and bindings. *Journal of Intelligence*, *11*(5), 94. https://doi.org/10.3390/jintelligence11050094

Forsberg, A., Guitard, D., Greene, N. R., Naveh-Benjamin, M., & Cowan, N. (2022b). The proportion of working memory items recoverable from long-term memory remains fixed despite adult aging. *Psychology and Aging, 37*(7), 777. https://doi.org/10.1037/pag0000703

Forsberg, A., Johnson, W., & Logie, R. H. (2020). Cognitive aging and verbal labeling in continuous visual memory. *Memory & Cognition*. 1196–1213. https://doi.org/10.3758/s13421-020-01043-3

Fukuda, K., & Vogel, E. K. (2019). Visual short-term memory capacity predicts the "bandwidth" of visual long-term memory encoding. *Memory & Cognition, 47*, 1481-1497. https://doi.org/10.3758/s13421-019-00954-0

Gathercole, S. E., Lamont, E. M. I. L. Y., & Alloway, T. P. (2006). Working memory in the classroom. In *Working memory and Education* (pp. 219-240). Academic Press. https://doi.org/10.1016/B978-012554465-8/50010-7

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *18*(1), 1-16. https://doi.org/10.1002/acp.934

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness*, *5*(2), 189-211. https://doi.org/10.1080/19345747.2011.618213.

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, *19*, 847-857. https://doi.org/10.3758/s13423-012-0296-9

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 351-360. https://doi.org/10.1016/S0022-5371(66)80044-0

Glenberg, A. M., & Swanson, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 3–15. https://doi.org/10.1037/0278-7393.12.1.3

Greene, N. R., & Naveh-Benjamin, M. (2022a). Adult age differences in specific and gist associative episodic memory across short- and long-term retention intervals. *Psychology and Aging*, *37*(6), 681–697. https://doi.org/10.1037/pag0000701

Greene, N. R., & Naveh-Benjamin, M. (2022b). Effects of divided attention at encoding on specific and gist representations in working and long-term memory. *Journal of Memory and Language*, *126*, 104340. https://doi.org/10.1016/j.jml.2022.104340

Greene, N. R., & Naveh-Benjamin, M. (2023). Differential attentional costs of encoding specific and gist episodic memory representations. *Journal of Experimental Psychology: General, 152*(11), 3292-3299. https://doi.org/10.1037/xge0001448.

Greene, N. R., Guitard, D., Forsberg, A., Cowan, N., & Naveh-Benjamin, M. (2024). Working memory limitations constrain visual episodic long-term memory at both specific and gist levels of representation. *Memory & Cognition*. Advance online publication. https://doi.org/10.3758/s13421-024-01593-w

Greene, R. L. (1987). Effects of maintenance rehearsal on human memory. *Psychological Bulletin, 102*(3), 403–413. https://doi.org/10.1037/0033-2909.102.3.403

Guitard, D., & Cowan, N. (2023). Attention allocation between item and order information in short-term memory. *Quarterly Journal of Experimental Psychology, 76*(6), 1391-1409. https://doi.org/10.1177/17470218221118451

Guitard, D., & Saint-Aubin, J. (2022). Backward recall and foreknowledge of recall direction: a test of the Encoding–Retrieval Matching Hypothesis. *Memory*, *30*(8), 1057-1072. https://doi.org/10.1080/09658211.2022.2079675

Guitard, D., Saint-Aubin, J., Poirier, M., Miller, L. M., & Tolan, A. (2020). Forward and backward recall: Different visuospatial processes when you know what's coming. *Memory & Cognition*, *48*, 111-126. https://doi.org/10.3758/s13421-019-00966-w

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d′. *Behavior Research Methods, Instruments, & Computers*, *27*, 46-51. https://doi.org/10.3758/BF03203619

Healey, M. K., Crutchley, P., & Kahana, M. J. (2014). Individual differences in memory search and their relation to intelligence. *Journal of Experimental Psychology: General*, *143*(4), 1553–1569. https://doi.org/10.1037/a0036306

Hebb, D. O. (1949). The first stage of perception: growth of the assembly. *The Organization of Behavior*, *4*(60), 78-60. https://doi.org/10.7551/mitpress/4943.003.0006

Hollingworth, H. L. (1913). Characteristic differences between recall and recognition. *The American Journal of Psychology*, *24*(4), 532-544. https://doi.org/10.2307/1413450

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 923–941. https://doi.org/10.1037/0278-7393.25.4.923

Howard, M. W., & Kahana, M. J. (2002). When does semantic similarity help episodic retrieval? *Journal of Memory and Language, 46*(1), 85–98. https://doi.org/10.1006/jmla.2001.2798

Hulme, C., Maughan, S., & Brown, G. D. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*(6), 685-701. https://doi.org/10.1016/0749-596X(91)90032-F

James, W. (1890). The principles of psychology, Vol. 1. Henry Holt and Co. *New York*. https://doi.org/10.1037/10538-000

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

Jeanneret, S., Bartsch, L. M., & Vergauwe, E. (2023). To be or not to be relevant: Comparing short-and long-term consequences across working memory prioritization procedures. *Attention, Perception, & Psychophysics*, 1-13. https://doi.org/10.3758/s13414-023-02706-4

Jiang, Q., & Cowan, N. (2020). Incidental learning of list membership is affected by serial position in the list. *Memory, 28*(5), 669–676. https://doi.org/10.1080/09658211.2020.1761398

Jones, D., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 1008. https://doi.org/10.1037/0278-7393.21.4.1008

Kahana, M. J. (1996). Associate retrieval processes in free recall. *Memory & Cognition*, *24*(1), 103–109. https://doi.org/10.3758/BF03197276

Kahana, M. J. (2012). *Foundations of Human Memory*. New York: Oxford University Press

Kahana, M. J. (2020). Computational models of memory search. *Annual Review of Psychology, 71*, 107–138. https://doi.org/10.1146/annurev-psych-010418-103358

Kahana, M. J. (2017). Memory search. In J. H. Byrne (Ed.), *Learning and memory: A comprehensive reference* (Second Edition, vol. 2, pp. 181–200). Academic Press. http://dx.doi.org/10.1016/B978-0-12-809324-5.21038-9

Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd edition). Academic Press.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4

Kuhn, J. R., Lohnas, L. J., & Kahana, M. J. (2018). A spacing account of negative recency in final free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(8), 1180. https://doi.org/10.1037/xlm0000491

Lawrence, C. O., Guitard, D., & Cowan, N. (2024). Short-term retention of words as a function of encoding depth. *Memory & Cognition*, 1-19. https://doi.org/10.3758/s13421-024-01546-3

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience*, *24*(1), 61-79. https://doi.org/10.1162/jocn_a_00140

Loaiza, V. M., & Camos, V. (2018). The role of semantic representations in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44*(6), 863–881. https://doi.org/10.1037/xlm0000475

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*, 191-203. https://doi.org/10.3758/s13421-011-0148-2

Loaiza, V. M., McCabe, D. P., Youngblood, J. L., Rose, N. S., & Myerson, J. (2011). The influence of levels of processing on recall from working memory and delayed recall tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(5), 1258–1263. https://doi.org/10.1037/a0023923

Loaiza, V. M., Oftinger, A. L., & Camos, V. (2023). How does working memory promote traces in episodic memory?. *Journal of Cognition*, *6*(1). https://doi.org/10.5334/joc.245

Logie, R., Camos, V., & Cowan, N. (Eds.). (2020). Working memory: The state of the science. 10.1093/oso/9780198842286.001.0001

Makowski, D. (2018). The psycho package: An efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software*, *3*(22), 470. 10.21105/joss.00470

Malmberg, K. J., Raaijmakers, J. G., & Shiffrin, R. M. (2019). 50 years of research sparked by Atkinson and Shiffrin (1968). *Memory & Cognition*, *47*, 561-574. https://doi.org/10.3758/s13421-019-00896-7

Maskarinec, A. S., & Brown, S. C. (1974). Positive and negative recency effects in free recall learning. *Journal of Verbal Learning and Verbal Behavior*, *13*(3), 328-334. https://doi.org/10.1016/S0022-5371(74)80070-8

Mazuryk, G.F., & Lockhart, R.S. (1974). Negative recency and levels of processing in free recall. *Canadian Journal of Psychology\/revue Canadienne De Psychologie*, *28*, 114-123. https://doi.org/10.1037/H0081971

McAteer, S. M., Ablott, E., McGregor, A., & Smith, D. T. (2023). Dynamic resource allocation in spatial working memory during full and partial report tasks. *Journal of Vision*, *23*(2), 10-10. https://doi.org/10.1167/jov.23.2.10

McElree, B. (2006). Accessing recent events. *Psychology of Learning and Motivation*, *46*, 155-200. https://doi.org/10.1016/S0079-7421(06)46005-9

Monsell, S. (1978). Recency, immediate recognition memory, and reaction time. *Cognitive Psychology*, *10*(4), 465-501. https://doi.org/10.1016/0010-0285(78)90008-7

Morey, C. C., Morey, R. D., Van Der Reijden, M., & Holweg, M. (2013). Asymmetric cross-domain interference between two working memory tasks: Implications for models of working memory. *Journal of Memory and Language*, *69*(3), 324-348. https://doi.org/10.1016/j.jml.2013.04.004

Morey, R. D. (2011). A Bayesian hierarchical model for the measurement of working memory capacity. *Journal of Mathematical Psychology, 55*(1), 8-24. https://doi.org/10.1016/j.jmp.2010.08.008

Morey, R. D., & Rouder, J. N. (2015). BayesFactor: Computation of Bayes factors for common designs (R package version 0.9.11-1) [Computer software]. Retrieved from http://CRAN.R-project.org/package=BayesFactor

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. Journal of Verbal Learning & Verbal Behavior, 16(5), 519–533. https://doi.org/10.1016/S0022-5371(77)80016-9

Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482. https://doi.org/10.1037/h0045106

Nairne, J.S. (2002).Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53*,53–81. https://doi.org/10.1146/annurev.psych.53.100901.135131

Naveh-Benjamin, M., & Jonides, J. (1984). Maintenance rehearsal: A two-component analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(3), 369. https://doi.org/10.1037/0278-7393.10.3.369

Neath, I. (1993). Distinctiveness and serial position effects in recognition. *Memory & Cognition*, *21*(5), 689-698. https://doi.org/10.3758/BF03197199

Norris, D. (2017). Short-term memory and long-term memory are still different. *Psychological Bulletin*, *143*(9), 992. https://doi.org/10.1037/bul0000108

Oberauer, K., Lange, E., & Engle, R. W. (2004). Working memory capacity and resistance to interference. *Journal of Memory and Language*, *51*(1), 80-96. https://doi.org/10.1016/j.jml.2004.03.003

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, *19,* 779-819. https://doi.org/10.3758/s13423-012-0272-4

Overkott, C., & Souza, A. S. (2022). Verbal descriptions improve visual working memory but have limited impact on visual long-term memory. *Journal of Experimental Psychology: General*, *151*(2), 321. https://doi.org/10.1037/xge0001084

Öztekin, I., Davachi, L., & McElree, B. (2010). Are representations in working memory distinct from representations in long-term memory? Neural evidence in support of a single store. *Psychological Science, 21*(8), 1123-1133. https://doi.org/10.1177/0956797610376651

Page, M., & Norris, D. (1998). The primacy model: a new model of immediate serial recall. *Psychological Review*, *105*(4), 761. https://doi.org/10.1037/0033-295X.105.4.761-781

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70,* 153-163. https://doi.org/10.1016/j.jesp.2017.01.006

Plancher, G., & Barrouillet, P. (2020). On some of the main criticisms of the modal model: Reappraisal from a TBRS perspective. *Memory & Cognition, 48*, 455-468. https://doi.org/10.3758/s13421-019-00982-w

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124, p. 125).

Postman, L., & Phillips, L. W. (1965). Short-term Temporal Changes in Free Recall. *Quarterly Journal of Experimental Psychology*, *17*(2), 132-138. https://doi.org/10.1080/17470216508416422

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory, 2*(5), 509-522. https://doi.org/10.1037/0278-7393.2.5.509

R Core Team. (2020). R: A language and environment for statistical computing R version 3.6.3 (2020-02-29) [Computer software]. Vienna, Austria: Author. Retrieved from http://www.R-project.org/

Ranganath, C., & Blumenfeld, R. S. (2005). Doubts about double dissociations between short- and long-term memory. *Trends in Cognitive Sciences, 9*(8), 374-380. https://doi.org/10.1016/j.tics.2005.06.009

Raye, C. L., Johnson, M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. (2007). Refreshing: A minimal executive function. *Cortex, 43*(1), 135-145. https://doi.org/10.1016/S0010-9452(08)70451-9

Rhodes, S., Cowan, N., Hardman, K. O., & Logie, R. H. (2018). Informed guessing in change detection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(7), 1023. https://doi.org/10.1037/xlm0000495

Ricker, T. J., Nieuwenstein, M. R., Bayliss, D. M., & Barrouillet, P. (2018). Working memory consolidation: insights from studies on attention and working memory. *Annals of the New York Academy of Sciences*, *1424*(1), 8-18. https://doi.org/10.1111/nyas.13633

Ricker, T.J., & Hardman, K.O. (2017). The nature of short-term consolidation in visual working memory. *Journal of Experimental Psychology: General*, *146*(11):1551-1573. doi: 10.1037/xge0000346. https://doi.org/10.1037/xge0000346

Ricker, T.J., & Vergauwe, E. (2022). Boundary conditions for observing cognitive load effects in visual working memory. *Memory & Cognition, 50*, 1169–1185. https://doi.org/10.3758/s13421-022-01320-3

Rivera-Lares, K., Logie, R., Baddeley, A., & Della Sala, S. (2022). Rate of forgetting is independent of initial degree of learning. *Memory & Cognition*, *50*(8), 1706-1718. https://doi.org/10.3758/s13421-021-01271-1

Roediger, H. L., Gallo, D. A., & Geraci, L. (2002). Processing approaches to cognition: the impetus from the levels-of-processing framework. *Memory*, *10*(5-6), 319–332. https://doi.org/10.1080/09658210224000144

Rose, N. S., & Craik, F. I. M. (2012). A processing approach to the working memory/long-term memory distinction: Evidence from the levels-of-processing span task.*Journal of*

Experimental Psychology: Learning, Memory, and Cognition, 38(4), 1019–1029. https://doi.org/10.1037/a0026976

Rose, N. S., Myerson, J., Roediger, H. L. III, & Hale, S. (2010). Similarities and differences between working memory and long-term memory: Evidence from the levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 471–483. https://doi.org/10.1037/a0018405

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*(2), 301-308. https://doi.org/10.3758/s13423-014-0595-4

Rouder, J. N., Morey, R. D., Speckman, P. L., & Pratte, M. S. (2007). Detecting chance: A solution to the null sensitivity problem in subliminal priming. *Psychonomic Bulletin and Review, 14*, 597-605. https://doi.org/10.3758/BF03196808

Rundus, D., & Atkinson, R. C. (1970). Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, *9*(1), 99-105. https://doi.org/10.1016/S0022-5371(70)80015-9

Saint-Aubin, J., & Poirier, M. (2000). Immediate serial recall of words and nonwords: Tests of the retrieval-based hypothesis. *Psychonomic Bulletin & Review*, *7*(2), 332-340. https://doi.org/10.3758/BF03212990

Schönbrodt, F. D., & Stefan, A. M. (2018). BFDA: An R package for Bayes factor design analysis (version 0.4. 0). https://doi.org/10.3758/s13423-017-1230-y

Scoville, W.B. and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* 20,11–212 10.1136/jnnp.20.1.11

Sederberg, P. B., Gauthier, L. V., Terushkin, V., Miller, J. F., Barnathan, J. A., & Kahana, M. J. (2006). Oscillatory correlates of the primacy effect in episodic memory. *NeuroImage*, *32*(3), 1422-1431. https://doi.org/10.1016/j.neuroimage.2006.04.223

Shallice, T. and Warrington, E.K. (1970) Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology. 22*, 261–273. https://doi.org/10.1080/00335557043000203

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review*, *76*(2), 179. https://doi.org/10.1037/h0027277

Shimi, A., & Logie, R. H. (2019). Feature binding in short-term memory and long-term learning. *Quarterly Journal of Experimental Psychology*, *72*(6), 1387-1400. https://doi.org/10.1177/1747021818807718

Smyth, M. M., & Scholey, K. A. (1996). Serial order in spatial immediate memory. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 159-177. https://doi.org/10.1080/713755615

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

Sternberg, S. (1966). High-speed scanning in human memory. *science, 153*(3736), 652-654. https://doi.org/10.1126/science.153.3736.652

Stigler, S. M. (1978). Some forgotten work on memory. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(1), 1-4. **https://doi.org/10.1037/0278-7393.4.1.1**

Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*(4), 1096-1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24-31. https://doi.org/10.1177/0098628316677643

Su., Y.-S., & Yajima, M. (2015). R2jags: Using R to Run 'JAGS' [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=R2jags (R package version 0.5-7).

Surprenant, A.M., &Neath, I. (2009). The nine lives of short-term memory. In A. Thorn & M. Page (Eds.), Interactions between short-term and long-term memory in the verbal domain (pp.16–43). Hove, UK: Psychology Press. https://doi.org/10.4324/9780203938966

Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review*, *15*(3), 535-542. https://doi.org/10.3758/PBR.15.3.535

Tatler, B. W., Gilchrist, I. D., & Rusted, J. (2003). The time course of abstract visual representation. *Perception, 32*(5), 579-592. https://doi.org/10.1068/p3396

Tendeiro, J. N., & Kiers, H. A. (2019). A review of issues about null hypothesis Bayesian testing. *Psychological Methods*, *24*(6), 774. https://doi.org/10.1037/met0000221

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520-522. https://doi.org/10.1038/381520a0

Townsend, J. T. (1971). A note on the identifiability of parallel and serial processes. *Perception & Psychophysics*, *10*(3), 161-163. https://doi.org/10.3758/BF03205778

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science, 1*(1), 46-54. https://doi.org/10.1111/j.1467-9280.1990.tb00067.x

Townsend, J. T., & Wenger, M. J. (2004). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review*, *11*(3), 391-418. https://doi.org/10.3758/BF03196588

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

Valentini, B., & Vergauwe, E. (2022). Stuck on the Last: The Last-Presented Benefit as an Index of Attentional Refreshing in Adolescents. *Journal of Intelligence*, *11*(1), 4. https://doi.org/10.3390/jintelligence11010004

Van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., ... & Wagenmakers, E. J. (2021). The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychonomic Bulletin & Review*, *28*, 813-826. https://doi.org/10.3758/s13423-020-01798-5

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

Vergauwe, E., Camos, V., & Barrouillet, P. (2014). The impact of storage on processing: How is information maintained in working memory?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1072. https://doi.org/10.1037/a0035779

Ward, G., Avons, S. E., & Melling, L. (2005). Serial position curves in short-term memory: Functional equivalence across modalities. *Memory, 13*(3-4), 308-317. https://doi.org/10.1080/09658210344000279

Wetzels, R., & Wagenmakers, E. J. (2012). A default Bayesian hypothesis test for correlations and partial correlations. *Psychonomic Bulletin & Review*, *19*, 1057-1064. https://doi.org/10.3758/s13423-012-0295-x

Williams, J. R., Robinson, M. M., Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2022). You cannot “count” how many items people remember in visual working memory: The importance of signal detection–based measures for understanding change detection performance. *Journal of Experimental Psychology: Human Perception and Performance, 48*(12), 1390. https://doi.org/10.1037/xhp0001055

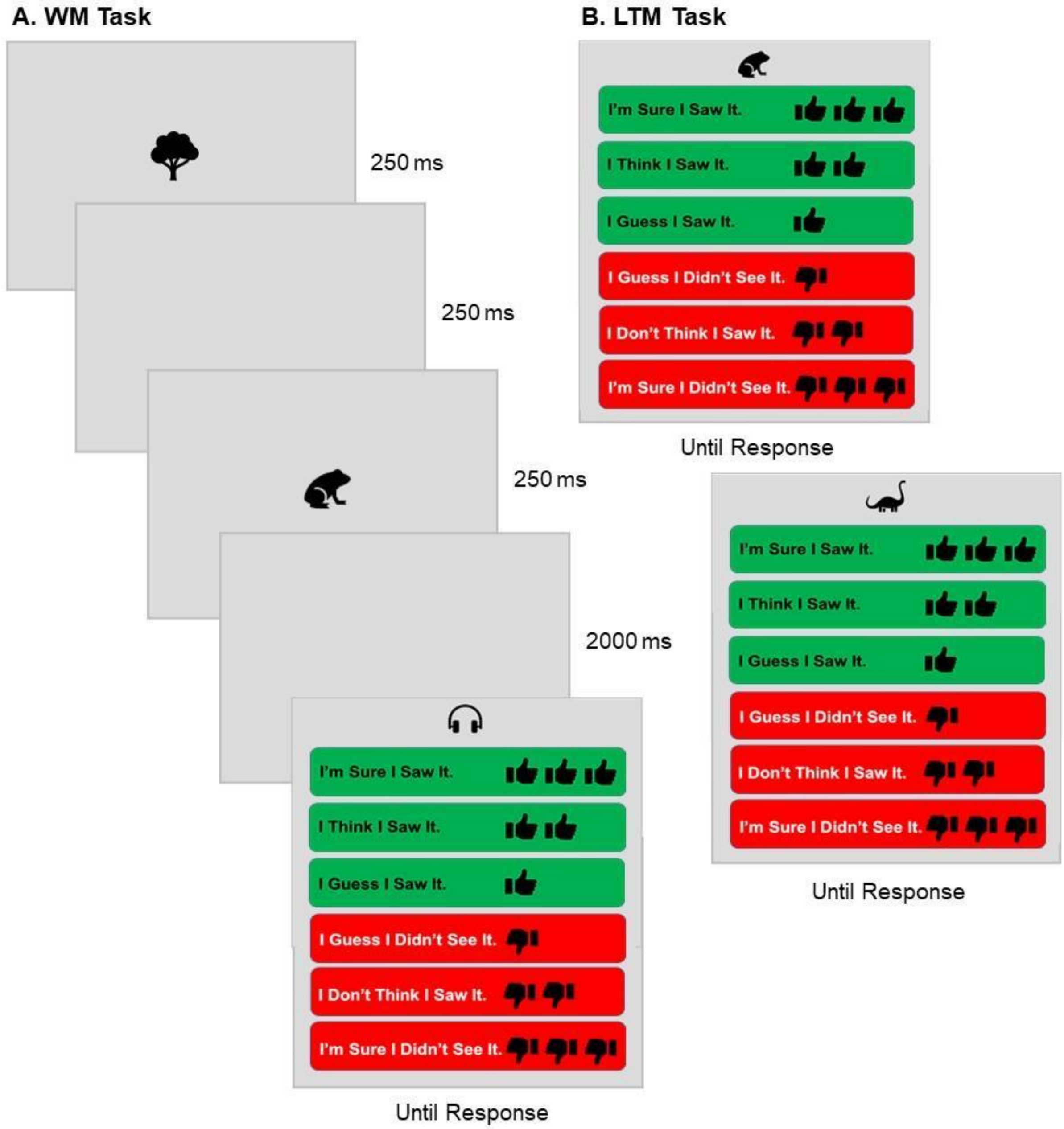

Figure 1. Outline of some typical trials. **Panel A**, Working Memory (WM) Task trial, at set size 2. **Panel B**, two trials in the Long-Term Memory (LTM) Task. The memory array set size in the WM task varied between 2, 4, or 6 items, and each item was presented for 250 ms. During the WM response phase, participants indicated whether the probe item was the same as – or different from – an item in the array, by clicking on the relevant option. In the LTM task, participants

were asked to respond to whether items had been studied in the WM task. ‘Until Response” in the figure indicates that participants had 10 minutes to provide their response.

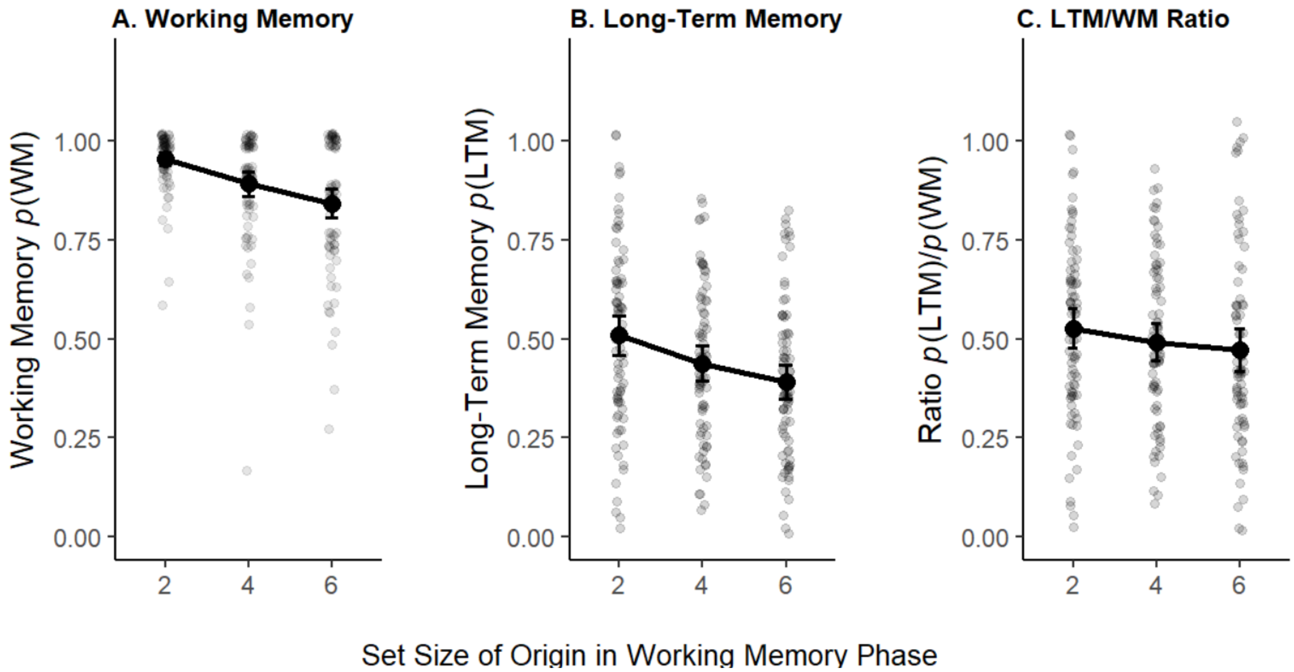

Figure 2. Data from Experiment 1. **Panel A.** Working Memory. **Panel B.** Long-Term Memory. **Panel C.** LTM / WM Ratios, by WM set size. The large circles triangles show the average for each set size. Smaller circle outlines show individual subject averages. Error bars represent Standard Errors.

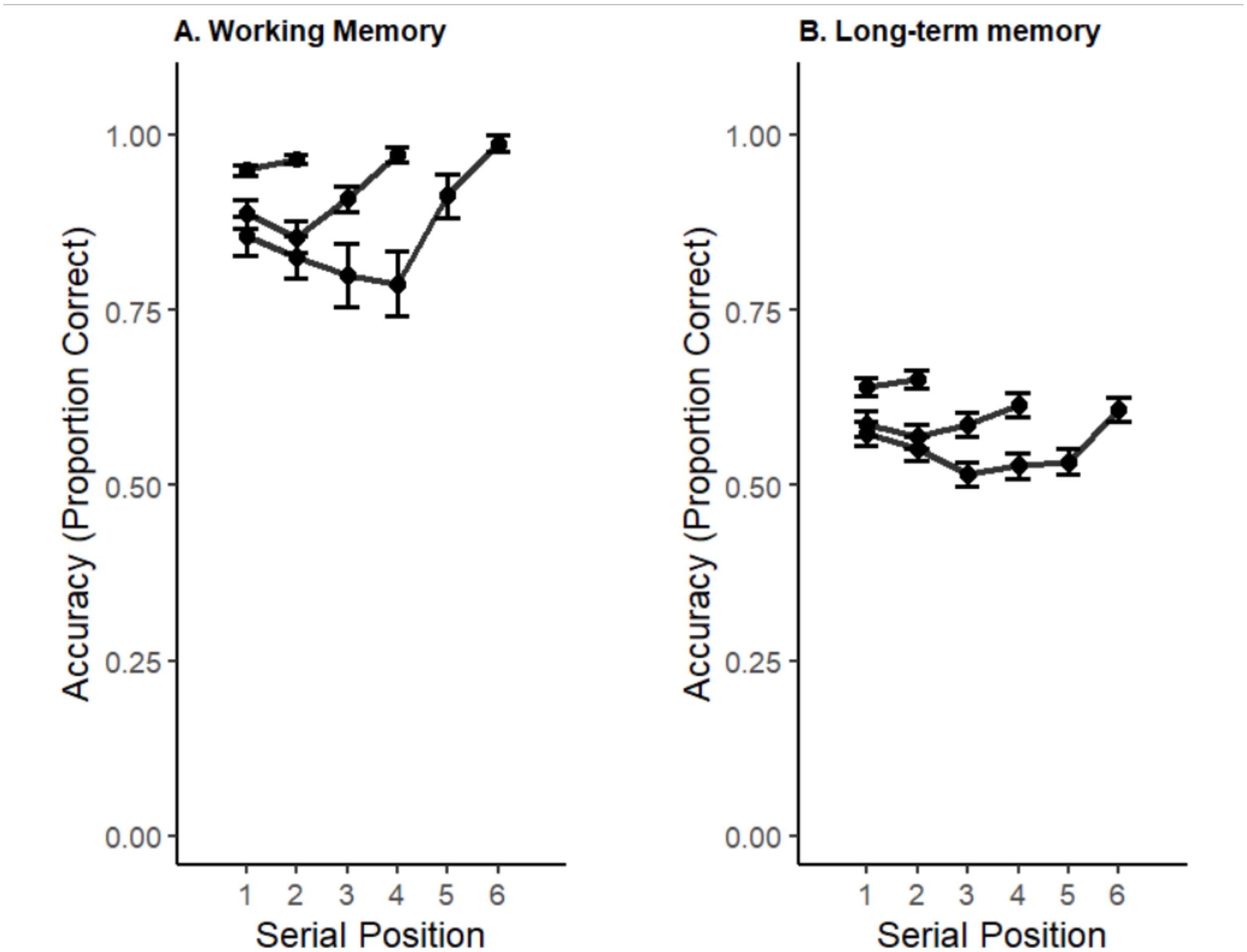

Figure 3. Data from Experiment 1, depicting serial Position effects at Set Sizes 2, 4 and 6. **Panel A.** Working Memory. **Panel B.** Long-Term Memory. . Error bars represent Standard Errors.

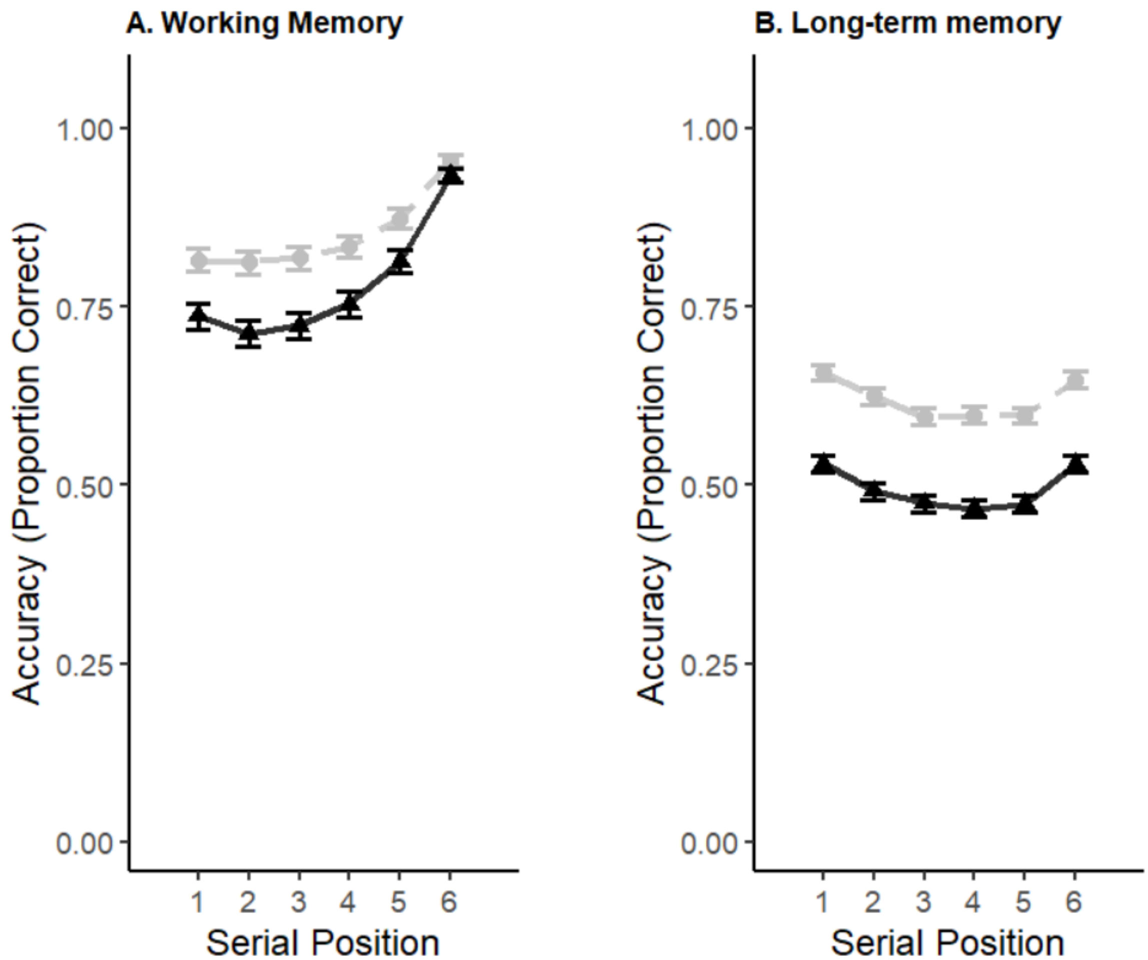

Figure 4. Data from Experiment 2, depicting serial Position effects at Set Size 6. **Panel A.** Working Memory. **Panel B.** Long-Term Memory. Black triangles represent 'strict' scoring, in which all trials marked as guesses are scored as incorrect. Error bars represent Standard Errors.

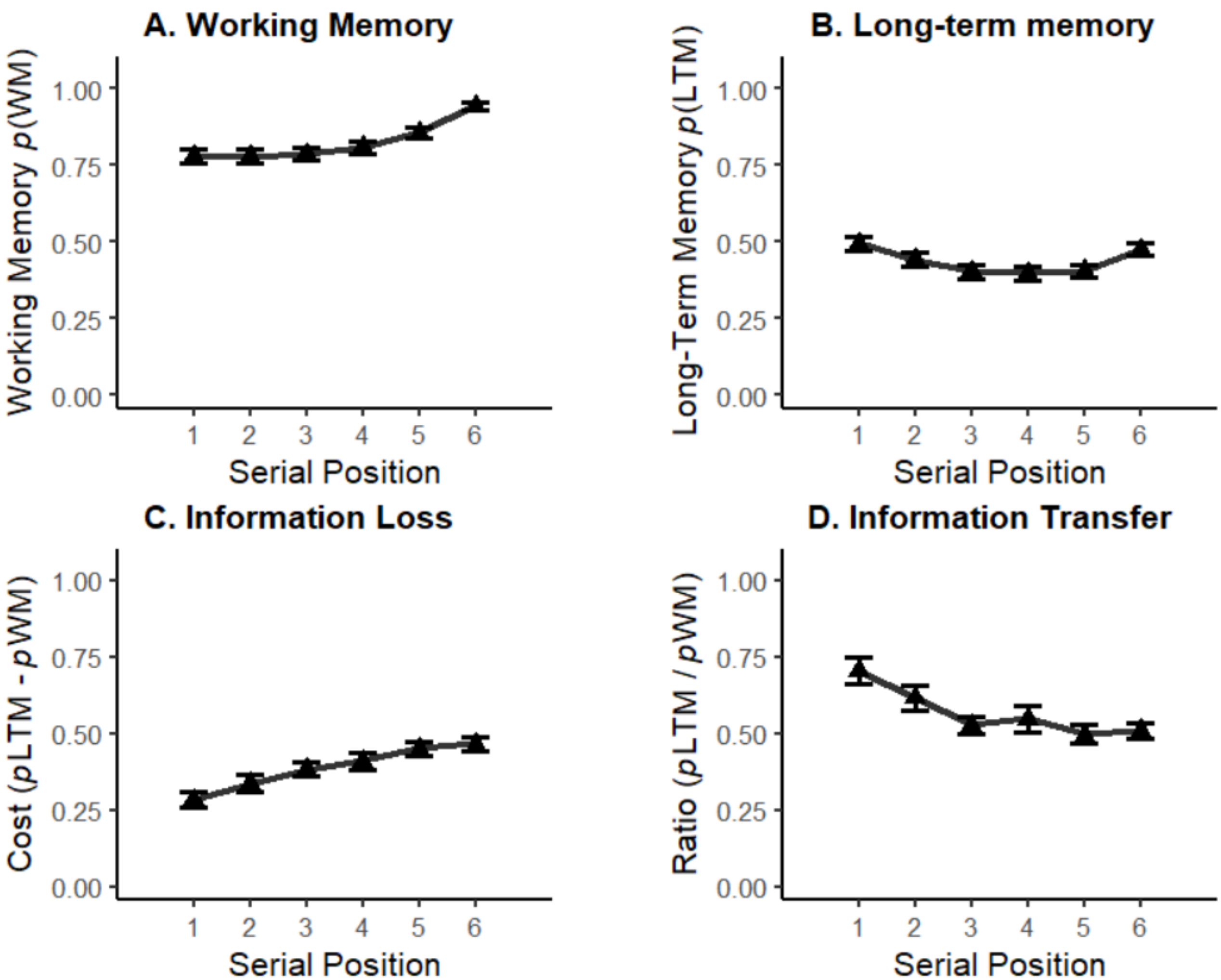

Figure 5. Data from Experiment 2, depicting serial Position effects. **Panel A.** Working Memory; *p*(WM). **Panel B.** Long-Term Memory, *p*(LTM). **Panel C**. Information Loss (i.e., the difference between *p*(WM) and *p*(LTM). **Panel D.** Transfer Ratio between *p*(WM) and *p*(LTM). Adjusted *p*(WM) and *p*(LTM) values are represented in the figure. Error bars represent Standard Errors.