The List Length Effect in Short-Term Memory

Molly MacMillan

A Thesis submitted to the School of Graduate Studies
In partial fulfillment of the
Requirements for the Master of Science in Experimental Psychology
Department of Psychology
Memorial University of Newfoundland
2021

Acknowledgements

I would like to thank my supervisors, Kathleen Hourihan and Ian Neath for lending their time, guidance and expertise to this project. I am grateful for their feedback and patience, which have been invaluable over the last three years. I would also like to thank my committee member Dr. Martin Day for his valuable feedback. This research was supported by NSERC Discovery Grants awarded to Dr. Kathleen Hourihan and Dr. Ian Neath.

Abstract

In free recall, the list length effect (LLE) refers to the finding that the proportion of correctly recalled items decreases as set-size increases but at the same time the total number of recalled items continues to increase with set-size (Murdock, 1962). Oberauer et al. (2018) proposed that a decrease in memory accuracy as a function of increasing set-size was fundamental to conceptualizations of short-term and working memory. Evidence of a LLE in short-term/working memory would contradict this benchmark. Beaman (2006) observed a LLE in serial recall whereas Unsworth and Engle (2006) observed no such effect in either serial recall or complex span. In the current research, we sought to reconcile these conflicting results. Six experiments were conducted, examining the relationship between the number and proportion of words recalled in both serial recall and complex span. No LLE was observed in either task. Instead, the proportion of words recalled decreased as a function of list length, while the number of words recalled initially increased, before either reaching a plateau or decreasing. The results suggest that recall accuracy decreases as a function of increasing list length due to increased interference and decreased positional and temporal distinctiveness in longer lists.

General Summary

One general principle suggested as a defining feature of short-term memory is that recall performance decreases as the number of items an individual is asked to hold in memory increases (Oberauer et al., 2018). Regardless of whether memory is measured by the proportion or number of items recalled, Unsworth and Engle (2006) found that performance decreased as list length increased. In contrast, Beaman (2006) found that the number of items recalled increased with list length. The current research aimed to reconcile these conflicting results and assess the validity of the proposed short-term memory principle. The results were consistent with the proposed principle. Both the number and proportion of words recalled decreased with list length. This pattern is attributed to a loss of item distinctive in longer lists, making it more difficult to retrieve an item from memory. This explanation supports the view that the same mechanisms underlie forgetting over both the short- and long-term.

Table of Contents

The List Length Effect in Short-Term Memory

There is a substantive body of literature that focuses on the study of short-term memory – the temporary storage of small amounts of information held over the span of a few seconds. Within this literature, there are well-established empirical findings that occur consistently across paradigms and stimulus types. In 2018, Oberauer et al. catalogued and rated these critical findings to establish and prioritize benchmarks against which models and theories of short-term memory could be evaluated. The first of these benchmarks asserts that the accuracy of recall in short-term memory decreases as set-size increases. While a significant number of short-term memory studies appear to support this statement, a study examining the list length effect (LLE) in short-term memory suggests that it may be too simplistic (Beaman, 2006).

The LLE refers to the phenomenon that, while the proportion of correctly recalled items declines as set sizes increases, the total number of recalled items continues to increase with set-size (Murdock, 1968). This effect is most commonly explained using models that assume separate short-term and long-term memory systems, and the prevailing perspective concludes that it is a long-term memory effect. As a result, the LLE has predominantly been studied in tasks, such as free recall, purported to assess long-term memory. However, two studies have investigated the LLE's existence in tasks used to assess short-term memory. In both serial recall and operation span, Unsworth and Engle (2006) found that increasing set-size resulted in a decrease in both the proportion and the total number of items recalled. By contrast, in serial recall, Beaman (2006) found that, although the proportion of recalled items decreased, the absolute number of correctly recalled items continued to increase as a function of increasing set-size.

This latter study is inconsistent with Oberauer et al.'s set-size-accuracy benchmark and poses a problem for the assumption that the LLE is a long-term memory phenomenon. If Beaman's (2006) findings are replicable, this would have significant implications for models and theories founded on the basis of a distinct, limited capacity short-term memory system. Moreover, evidence of the LLE in a short-term memory task would favour the perspective that greater integration is required between accounts of memory over the short- and long-term. Consequently, the overall aim of the current research was to reconcile the discrepancies between Beaman (2006) and Unsworth and Engle (2006) in terms of their opposing implications for the set-size-accuracy benchmark proposed by Oberauer et al. (2018).

In the following section, I will review the theoretical accounts that propose a dissociation between the systems underlying short-term memory and long-term memory. I will then contrast these accounts with those that propose a unitary memory system governed by general principles. The models and theories that have emerged from these perspectives offer varying predictions for the effects of list length on recall in tasks that ostensibly measure short-term memory. As such, whether the current research confirms or refutes the existence of a LLE in short-term memory tasks has significant implications for the strength of these models. Next, I will compare immediate free recall and immediate serial recall, highlighting the similarities and differences between the two tasks. Although the immediate free recall and immediate serial recall follow remarkably similar procedures, they are generally approached from different theoretical perspectives. Evidence of the same effect in both tasks might suggest greater theoretical integration is required. I will then outline the evidence supporting Oberauer et al.'s set-size accuracy benchmark. Finally, I will present the current research and the hypothesis and predictions.

Multi-store accounts of memory

In 1958, Broadbent introduced a novel framework that characterized memory as a series of stores through which information flows during processing. In this model, he posited separate memory systems for information held over the short- and long-term. In particular, Broadbent suggested that information is filtered from a pre-attentive sensory store into a limited capacity short-term store and finally into a long-term store. He believed that information held in the short-term store was subject to rapid decay and could only be maintained through active rehearsal. In contrast, he posited that information in the long-term store was more permanent. This purported dissociation between short-term and long-term memory was highly influential, and in the following years, several related models were introduced delineating multi-store accounts of memory (Atkinson & Shiffrin, 1968; Glanzer, 1972; Waugh & Norman, 1965). Eventually termed *the modal model* (Murdock, 1974), these accounts became the dominant perspective through which memory was investigated and understood. The significance of the modal model and its enduring influence on the field of cognitive psychology would be difficult to overstate (Baddeley, Hitch & Allen, 2019).

Serial Position Effects. Most variants of the modal model identified rehearsal as the mechanism that allowed information to pass from short-term memory into long-term memory (Engle & Oransky, 1999). Studies examining the serial position curve (SPC) in free recall were often cited to support this proposed role of rehearsal. Free recall typically yields a U-shaped SPC, showing a higher proportion of items recalled from both the start and end of a list. These distinct features of the curve are referred to as the primacy and recency effects, respectively. The primacy effect is typically attributed to the greater rehearsal of early list items resulting in a record of those items in long-term memory. In contrast, the recency effect is believed to reflect residual traces of the last few list items in short-term memory. Glanzer and Cunitz (1966) provided evidence for this dissociation by demonstrated that extending the retention interval resulted in a reduced recency effect but did not impact the primacy effect. These results supported the view that short-term memory traces, believed to be susceptible to decay, underlie the recency effect but not the primacy effect.

The dual-store account of free recall was used to support the claim that rehearsal drives information transfer from short-term memory into long-term memory. Welch and Burnett (1924) demonstrated that the primacy effect, typically attributed to long-term memory processes, could be eliminated by inhibiting rehearsal. They presented participants with a sequence of unpronounceable trigrams and instructed them to focus their attention exclusively on the most recently presented item. No primacy effect was observed on the subsequent recall test. In two additional experiments, participants were shown pronounceable trigrams and were instructed to remember as many items as possible. The primacy effect re-emerged. These results were interpreted as evidence for the essential role of rehearsal in transferring information into long-term memory. This conclusion was later supported by studies using overt rehearsal, an experimental method that requires participants to speak any rehearsals aloud during the study phase. As anticipated, early list items received more rehearsals during study than later list items (Rundus & Atkinson, 1970; Tan & Ward, 2000). Moreover, incidental learning studies (Glenberg et al., 1980; Seamon & Murray, 1976) have shown that the primacy effect is significantly reduced when participants are not motivated to rehearse list items. Collectively, the evidence from these studies supports the claim that rehearsal is the mechanism of information transfer between memory stores. These studies also favour accounts purporting that the primacy and recency effects result from the operation of separate memory systems.

However, Bjork and Whitten (1974) demonstrated that the recency effect cannot always be attributed to short-term memory processes. They adapted the standard free recall task to create the continual distractor paradigm. In this paradigm a distractor-filled inter-presentation interval is added before and after the presentation of each list item, and a distractor-filled retention interval is added between the list presentation and the recall test. Using this procedure, Bjork and Whitten tested participants' recall for lists of eight word-pairs. Participants completed either 0 or 12 seconds of arithmetic calculations during the inter-presentation interval, and either 0 or 30 seconds of arithmetic calculations during the retention interval. On lists with no inter-presentation interval, a recency effect occurred when recall was immediate, but was eliminated with the addition of a 30-second retention interval. This pattern is consistent with the view that the recency effect reflects retrieval of information not yet decayed from short-term memory. However, contrary to this view, when a 12 second inter-presentation interval was added, a recency effect occurred both with and without the 30 second retention interval.

This latter finding is not predicted by dual-store accounts of free recall and suggests that the recency effect cannot be solely attributed to the easy access of recency items stored in short-term memory. Instead, Bjork and Whitten (1974) concluded that the recency effect arises from long-term memory retrieval processes. They suggested that within an ordered series, later list items benefit from greater positional distinctiveness relative to middle list items. Thus the recency effect occurs because order information is a more effective retrieval cue for later list items. The authors also posited that the size of the recency effect could be predicted by the ratio of the temporal separation of list items and the temporal delay between the presentation episode and the recall test. The authors believed that these two factors were significant because they determine the discriminability of memory traces. Glenberg et al. (1983) confirmed this relationship by demonstrating that the size of the recency effect, measured by the slope of the line of best fit over the last three serial positions, is linearly related to the logarithmic ratio of the inter-presentation interval to the retention interval. This relationship is known as the ratio rule and is believed to be characteristic of free recall. The ratio rule is problematic for the modal model of memory. Instead of offering evidence for distinct short- and long-term memory systems, the ratio rule suggests that the recency effect is timescale invariant (Greene, 1986).

Forgetting. Evidence in favour of multiple, distinct memory systems also emerged from studies demonstrating qualitative differences in forgetting. Researchers attributed information loss in short-term memory to limitations in either temporal persistence (Brown, 1958; Conrad, 1957; Conrad & Hille, 1958) or capacity (Miller, 1956; Waugh & Norman, 1965). In contrast, while decay and displacement were the proposed mechanism of forgetting in short-term memory, loss of information in long-term memory was attributed to temporary disruption of retrieval resulting from proactive and retroactive interference (Melton & Irwin, 1940; Postman et al., 1968). Decay-based short-term memory accounts were typically tested using rehearsal-preventing tasks (J. Brown, 1958; Peterson & Peterson, 1959). In the Brown-Peterson paradigm, participants are shown a consonant trigram, followed immediately by a three-digit number. Participants are asked to count backwards by threes from the given number for a set amount of time. They are then asked to recall the consonant trigram. The counting task is designed to prevent rehearsal during the retention interval while not causing stimulus-consistent interference. When rehearsal was inhibited in such a way, recall from short-term memory declined rapidly as a function of the length of retention interval (Brown, 1958; Peterson & Peterson, 1959). As such, the Brown-Peterson paradigm has frequently been cited as evidence that decay is the primary mechanism of forgetting in short-term memory.

However, Keppel and Underwood (1962) challenged this assumption, arguing that proactive interference could account for forgetting in this paradigm. They observed that performance was consistently equivalent on the first trial of the Brown-Peterson task regardless of the length of the retention interval. Furthermore, they noted that the rates of forgetting increased as the number of trials increased. This finding cannot be accounted for under decay-based explanations of forgetting. Moreover, later studies using the same paradigm demonstrated that changing the to-be-remembered stimulus type on the third trial resulted in increased recall performance on that trial (Wickens et al., 1963). Again, this pattern is difficult to account for under decay-based explanations of forgetting but is consistent with a release from a build-up of proactive interference. These findings led Waugh and Norman (1965) to compare decay and interference-based accounts directly. The researchers used a probed recall task in which participants were presented with lists of digits, either at a rate of one or four items per second. At test, participants were probed with a single digit and were asked to recall the immediately preceding digit in the list. The position of probed item was manipulated. Recall was affected by the number of items subsequent to the target, not by the amount of time that passed between the presentation of the target and the test. The authors, therefore, concluded that forgetting in short-term memory occurred as a result of displacement and not decay.

On the basis of this conclusion, Waugh and Norman (1965) proposed a capacity/displacement model of immediate memory. They posited that memory is supported by two separate systems, primary memory and secondary memory. Under their model, perceptual information first enters primary memory. However, once the capacity of this system is met, newer items may displace older items. Once an item is displaced from primary memory, it can no longer be accessed and is permanently forgotten. Rehearsal can be used to prevent displacement from primary memory, and rehearsed information may be transferred to secondary. secondary was characterized as an unlimited and permanent store. Although Waugh and Norman's model was able to account for many findings, interference via displacement was not unanimously accepted as the source of forgetting in short-term memory. Some researchers continued to support time-based decay as the mechanism of forgetting (Baddeley et al., 1975; Barrouillet et al., 2004; McKeown & Mercer, 2012), others argued that displacement and decay might cause forgetting in a summative manner (Reitman, 1974), or alternatively, that both mechanisms lead to forgetting, but operate differentially depending on the context (Ricker et al., 2016). In 2021, Miller identified four families of accounts purporting to explain memory failure. They contrasted theories of forgetting based on interference, decay, displacement and inadequate retrieval cues. Based on a review of the evidence supporting each of these accounts, they concluded that the dominant cause of forgetting is associative interference coupled with inadequate retrieval cues.

Coding. In addition to the mechanisms of forgetting, multi-store theorists also proposed a dissociation between how information is coded in short-term memory and long-term memory. Although stimuli were presented visually, Conrad and Hull (1964) found that recall was poorer for lists composed of phonetically similar rather than dissimilar letters. They also found that participants made acoustic rather than visual confusion errors. Based on these results, the authors concluded that information is coded phonetically in short-term memory. Baddeley (1966) extended these results by presenting participants with either phonologically or semantically similar and dissimilar lists. In serial recall, a task targeting short-term memory, phonological similarity was detrimental to recall performance while semantic similarity was not. Baddeley interpreted these results as suggesting that information in short-term memory was stored phonetically and not semantically. In a related study, using a serial probe task, Kintsch and

Buschke (1969) found a phonemic similarity effect in recall of the last few list items and a semantic similarity effect in recall of the first list items. As the primacy portion of the serial recall curve is thought to reflect long-term memory processes, while the recency portion reflects short-term memory processes, the researchers concluded that information was coded semantically in long-term memory and phonetically in short-term memory.

Using concurrent articulation[1], Murray (1967) directly examined the necessity of a phonetic code in short-term memory. He visually presented participants with lists of letters for recall. In the concurrent articulation condition, participants were required to repeat the word “the” aloud during the list presentation. The procedure aimed to inhibit subvocal rehearsal. While participants tended to make acoustic confusion errors in the control condition, this was not the case in the concurrent articulation condition. These results suggested that, although a phonetic code may be favoured in short-term memory, information can also be stored in alternative forms within the memory system. Different memory tasks require the storage of different stimulus representations, and as such, the preferred coding format is likely context-dependent (Engle & Oransky, 1999).

The Atkinson and Shiffrin model. Atkinson and Shiffrin (1968) proposed a more flexible model that accounts for both the coding and forgetting differences ascribed to the proposed memory systems. In keeping with the modal model's previous incarnations, they suggested that environmental information is filtered through a sensory register before being transferred into a short-term store and then passing into a long-term store. Unlike previous

---

[1] This methodological procedure is generally referred to as *articulatory suppression*. The term concurrent articulation (Kleiman, 1975) was used here because it describes the manipulation, rather than the hypothetical consequence of the manipulation.

models, they posited the existence of multiple modality-specific sensory registers. The short-term store was characterized as a limited capacity buffer that temporarily holds information, whereas the long-term store was characterized as an unlimited and permanent store. In addition to these structural features, Atkinson and Shiffrin focused on the functional components of memory – the processes that can be actively controlled to accommodate the memory demands of a particular context.

Within the short-term store, Atkinson and Shiffrin emphasized the control an individual can exert over storage, rehearsal, and retrieval processes. As a result, their model makes task-specific predictions rather than general predictions about memory performance. They proposed that different rehearsal methods can be used to maintain information in the short-term store or transfer information to the long-term store. For example, rote rehearsal can be employed to offset decay. In this case, a limited capacity rehearsal buffer containing a fixed number of slots is established. Control operations are then used to determine what information is rehearsed within this buffer. Similarly, different coding processes can determine which aspects of a stimulus are retained, and different retrieval strategies may be used to access different information.

Raaijmakers and Shiffrin later proposed a computational model based on many of the tenets proposed in Atkinson and Shiffrin’s theoretical framework (Raaijmakers & Shiffrin, 1981; Raaijmakers & Shiffrin, 1980; Raaijmakers, 1979). The SAM (search of associative memory) model posits the existence of a limited capacity short-term store and an unlimited long-term store. According to SAM, new information is temporarily held in the short-term store, during which time it can be easily and accurately accessed. An item will remain in this store until capacity is reached, at which point new items randomly displace the older items. The long-term store holds associative information: relationships between context and item information and relationships between the context and information associated with pairs of items. The framework suggests that the amount of associative context-item information that enters the long-term store is a function of the amount of time an item spends in the short-term store. Similarly, the amount of associative item-item information that enters the long-term store is a function of the amount of time two items spend in the short-term store simultaneously. Retrieval from the long-term store was conceptualized as a two-phase, cue-dependent and highly fallible process. During the search phase, cues held within the short-term store are used to sample memory traces from the long-term store based probabilistically on the strength of association between the cues and the memory traces. During the subsequent recovery phase, the sampled cues are accessed and evaluated. Under this model, forgetting occurs when the relative strength of associations between a set of cues and the target item is weakened. This weakening may occur as a result of cue overload or context change between storage and retrieval.

Working Memory

One important contribution of the Atkinson and Shiffrin model was its characterization of the short-term store as a system responsible for both processing and storage. Rather than a passive store, the authors envisioned the short-term store as a system that supports learning and other cognitive activities. While this model was highly influential, it had difficulty accounting for some of the neuropsychological evidence. Patients with short-term memory deficits were often unimpaired in other cognitive tasks ascribed to the short-term store (Baddeley, 1992). These dissociations led Baddeley and Hitch (1974) to investigate concurrent performance on functions attributed to the short-term store. They asked participants to remember lists of digits while simultaneously completing other cognitive tasks. The researchers reasoned that if both tasks were dependent on the same resources, increasing the memory load in the digit span task would decrease performance on the concurrent task. As the list length in the digit span task was increased, performance on reasoning, learning and comprehension tasks declined. However, the drop in performance was less than would be expected if the two tasks depended on the same pool of resources (Baddeley, 1992). Baddeley and Hitch, therefore, concluded that the short-term store was not a unitary system but rather was composed of multiple subsystems responsible for different operations.

Baddeley's model of working memory. In a highly influential paper, Baddeley and Hitch (1974) posited a more dynamic and complex short-term store highlighting the role of attention in information processing. They proposed a three-component working memory (WM) system composed of a central executive, an articulatory loop and a visuospatial sketchpad. Under this model, the central executive is an attentionally limited subsystem responsible for information processing. It is supported by two subordinate systems: the phonological loop and the visuospatial sketchpad. The phonological loop maintains information via subvocal rehearsal and translates visual information into a phonetic code. In contrast, the visuospatial sketchpad maintains and manipulates visual and spatial information. In a subsequent paper, Baddeley (2000) proposed a fourth component of WM. The episodic buffer was conceptualized as a storage system capable of maintaining around four units of information. Baddeley posited that this subsystem was responsible for integrating information from different code sources and integrating information held in WM with inputs from perception and long-term memory.

In a series of subsequent papers, Baddeley and his colleagues further elaborated on this multicomponent model (Baddeley, 1992; Baddeley & Hitch, 1994; Baddeley & Logie, 1999). The framework that emerged remains the most influential approach to WM, defining it as a limited capacity store that temporarily holds information relevant to ongoing cognitive processes.

Baddeley (1992) suggested that memory storage is only one of the WM system's multiple functions. In his view, the system’s main role is to coordinate the resources required to perform cognitive tasks. The central executive oversees this process by selecting and manipulating input from the subsidiary slave systems. Meanwhile, the phonological loop performs the functions that most earlier models of immediate memory attributed to short-term memory. It maintains information via subvocal rehearsal and translates visual information into a phonetic code. Like previous models of immediate memory, Baddeley’s model suggests that information decays unless maintained through rehearsal.

Embedded Processes. Cowan (1999, 2005) proposed a different approach to working memory, focusing on the role of attentional processes. In the embedded processes model, he characterized WM as a set of long-term memory processes that hold information in a highly accessible state. Under this model, a subset of information in long-term memory is activated, and a subset of that information falls within the focus of attention. Cowan posited that attention is directed either through automatic responses to the environment or through voluntary control. Information within the focus of attention can be easily accessed and manipulated to perform cognitive functions. However, activation decays over time gradually unless reinstated. Similar to the rehearsal buffer proposed in Baddeley’s model, the embedded processes model also suggests that the focus of attention is limited to four units of information. However, Cowan believed that when semantic information is activated by attention, the corresponding sensory information is activated automatically. As a result, interference may occur between activated information with similar sensory features. Thus, Cowan’s model suggests that WM is limited by both decay and interference.

Alternative models. In 1998, Henson proposed the Start-End model to explain serial order recall. According to this model, list items are encoded as individual tokens in short-term memory. These tokens are composed of both item and positional information. During retrieval, the positional information is used as a cue to access the associated item information. The model assumes that position is coded as a function of an item's spatial or temporal proximity to both the start and end of a list. The start marker is more salient for earlier list items, while the end marker is more salient for later list items. As a result, the relative strength of the two markers can be used to identify each list position. The model suggests that tokens are not ordered in short-term memory, but rather that order is derived from the recall process. During retrieval, the positional code is reinstated and compared against the tokens held in short-term memory. The degree to which the positional cue corresponds with a token in short-term memory determines the likelihood that the token should become activated and successfully retrieved.

The ACT-R model, developed by Anderson (1983), characterizes WM as a subset of information available for processing that can currently be accessed either from long-term memory or from the outside world. Under this model, both the encoding of internal events and the processing of external stimuli result in WM activation. Similarly, the Connectionist/Control Architecture posited by Schneider and Detweiler (1987) proposes that WM consists of a subset of information activated above a specific threshold. This model differs in the assertion that WM activation may occur in parallel across multiple processing systems located in specific, specialized brain regions.

Many alternative conceptualizations of WM have been proposed since its inclusion in the Baddeley and Hitch model. Researchers have put forward a diverse range of theories and models, each characterizing the mechanisms of WM differently (Miyake & Shah, 1999). The limited capacity nature of the WM system is central to all conceptualizations. Three main categories of explanation have been proposed to account for this feature (Oberauer et al., 2016). Decay-based explanations propose that representations in WM become degraded over time rapidly unless maintained through rehearsal. These rehearsal processes take time and require the focus of attention. As such, the amount of information that can be maintained before being lost to decay constitutes the system's capacity. In contrast, resource-based explanations propose that the WM system draws on a limited pool of cognitive resources that must be divided between representations and processes. A portion of these available resources must be allocated to an item or process at all times for that item to be maintained or that process to be continued. Lastly, interference-based explanations propose that retrieval of information involves competition amongst representations held in WM. Under these WM models, the amount of interference within WM increases as a function of the number and similarity of representations held in the system. Only the most activated representation is successfully retrieved. Thus, the capacity limitations of the WM system result from an accumulation of inter-item interference.

Unitary accounts of memory

Despite the prevalence of multi-store models, whether human memory is better conceptualized as a single system or multiple systems remains controversial. Since the 1960s, the balance of opinions has generally favoured the position that two or more interacting systems underlie human memory, and the models and theories founded on this belief have been highly influential. Indeed, support for this perspective continues to be strong (Norris, 2017; Plancher & Barrouillet, 2020) However, some theorists oppose the assumption of separate memory systems and highlight invariance in performance across memory tasks and timescales as evidence favouring a unitary memory system. From this perspective, parallel interpretations of the memory phenomenon explained by multi-store models have been proposed and refined without relying on the existence of a distinct short-term memory or WM system.

In 1972, Craik and Lockhart presented an influential model that characterized memory as a by-product of other cognitive processes such as perception, comprehension, categorization and discrimination. According to their model, maintaining information in primary memory requires controlled attention. Information in primary memory can be quickly and accurately accessed but is limited in capacity due to finite attentional resources. Successful retrieval of an item from secondary is dependent on the depth of processing. Craik and Lockhart proposed that early sensory analysis, or shallow processing, of an item requires less attention than deeper, more semantic analysis and that the level of processing an item receives determines the strength of the memory trace in secondary. Therefore, the differential processing demands of a specific task would alter how information was stored in memory. Importantly, Craik and Lockhart's framework suggested that primary memory capacity limitations were due to attentional and not structural constraints.

Similarly, Crowder (1993) argued that individuals might favour different coding methods based on the specific task demands of a given context. He suggested that the distinctions made between short-term and long-term memory may reflect differences in how we code information. Thus the method of coding, and not the underlying memory system, may impact the characteristics of retention and subsequent retrieval. Studies targeting short-term memory generally involve the retention of five to nine items. In contrast, long-term memory studies typically test recall on lists of greater than ten items (Beaman, 2006; Grenfell-Essam & Ward, 2012). Thus, not only do these tasks differ in the purported memory system targeted, but also in the specific task demands. We could, therefore, attribute any qualitative differences in performance on short-term memory and long-term memory tasks to the specific task demands and not the underlying mechanisms. Indeed, Grenfell-Essam and Ward (2012) suggested that immediate serial recall and immediate free recall only differ as a result of the different list lengths typically used to test memory in these tasks.

The Feature model. Nairne developed both a descriptive framework (1988) and a simulation model (1990) explaining immediate memory without supposing separate short-term memory and long-term memory systems. According to Nairne's Feature model, to-be-remembered items are represented in memory as vectors composed of features. Features can either be modality dependent – associated with the physical characteristics of an item, or modality independent – associated with the internal labelling and categorization of an item. The number, value, and type of feature encoded for each item can vary. When an item is presented, a vector of features representing the item is simultaneously stored in primary and secondary memory. The features held in primary memory are then subject to overwriting caused by retroactive interference from subsequently presented items and related internal cognitive operations. At recall, successful retrieval of an item depends on a matching process in which residual trace information in primary memory is compared to a series of vectors sampled from a search set within secondary. This comparison is based on a degraded primary memory trace's relative similarity to each candidate secondary trace within a search set. Similarity is computed by considering the number of mismatching features relative to the number of features compared. The probability of sampling the correct secondary trace depends on the distinctiveness of the intact primary memory features. Thus according to the Feature model, all information is accessed via partially degraded cues in primary memory and is ultimately retrieved from secondary.

Although the Feature model can simulate a wide range of memory effects (Neath & Surprenant, 2003), it falls short in accounting for several findings (Hulme et al., 2006; Norris et al., 2004). For example, Norris et al. (2004) demonstrated that irrelevant sound presented after a to-be-remembered list can decrease memory performance. This retroactive effect of irrelevant sound continued to be observed when rehearsal of the list items was inhibited during the presentation of the irrelevant sound. Under the Feature model, irrelevant sounds disrupts recall by adding noise to features. However, the model predicts that this process will occurs exclusively when irrelevant sound is presented at the same time as the list or during rehearsal of list items. Therefore, the model has difficulty accounting for Norris et al's findings. A second example comes from the word length effect in immediate serial recall. Neath and Nairne (1995) extended the Feature model of explain this effect. They proposed that lists of shorter words are recalled better than lists of longer words because each word is composed of fewer segments. They posited that in order to successfully retrieve a word, each segment must be recalled in the correct order. Longer words, composed of more segments, present more opportunities for ordering errors. This account attributes the word length effect to item characteristics. However, Hulme et al. (2004) demonstrated that the word length effect is eliminated when recall is tested for lists containing both short and long words. Moreover, Hulme et al. (2006) showed that a single short or long word isolated within a list of long or short words is recalled better than the other list items. The isolated words were also better recalled than words from lists of uniform length. These findings suggest that the word length effect does not results from item characteristics as suggested by the Feature model.

SIMPLE. Neath and Brown (2006) demonstrated that primacy and recency effects could be predicted by a local distinctiveness model that does not assume separate underlying short-term memory and long-term memory systems. SIMPLE (Scale Invariant Memory Perceptual Learning; Brown et al., 2007) posits that successful retrieval is a function of an item's discriminability. Under this model, items in memory are understood to occupy a particular position along one or more psychological dimensions. An item's position along a dimension is used as a retrieval cue. The efficacy of a given cue varies as a function of the number of close neighbours an item has along the relevant dimension. In a standard free recall task, list items only vary systematically in terms of their presentation order. Thus each item in free recall occupies a unique position along a temporal dimension. The early and late list items are more discriminable as they have fewer neighbours on one side along the temporal dimension. This discriminability advantage underlies the primacy and recency effects. Importantly, under the SIMPLE model, discriminability is unaffected by scale. Spacing between items along a dimension can be increased or decreased by a constant factor without affecting recall performance. SIMPLE can, therefore, be applied to both short-term memory and long-term memory phenomena without requiring a distinction in how information is processed, stored or retrieved.

Immediate Free Recall and Immediate Serial Recall

Immediate free recall and immediate serial recall are among the most common tasks used to study human memory. In both tasks, research participants are shown sequences of individually presented items and are tested on their ability to recall those items. In immediate free recall, participants can output recalled items in any order, while in immediate serial recall, they are required to output items in forward serial order. Studies of immediate free recall typically involve lists of 10 to 20 items, whereas studies of immediate serial recall typically involve lists of five to seven items. Performance on both tasks has been used as evidence for a distinct limited capacity short-term memory store. However, despite the congruent theoretical conclusions derived from both tasks and the significant overlap between their methodologies, performance on the two tasks is most commonly explained using separate models that purport distinct underlying mechanisms (Ward et al., 2010).

immediate free recall is often described as a dual store task whereby retrieval of pre-recency items reflects long-term memory processes, and retrieval of recency items reflects short-term memory processes. In contrast, retrieval in immediate serial recall is generally attributed exclusively to short-term memory. In fact, short-term memory capacity is often approximated using memory span – the list length at which an individual can recall all items in the correct serial order on 50% of trials. Although performance on both tasks is commonly cited as evidence for a limited capacity short-term memory store, different theoretical accounts are generally used to explain the recency effect in immediate free recall and immediate serial recall. Baddeley and Hitch (1974) demonstrated that maintaining a six-digit sequence while concurrently performing an immediate free recall task did not decrease the size of the recency effect. They concluded that immediate free recall and immediate serial recall were drawing on separate short-term memory resources and were supported by different mechanisms. Bhatarah, Ward and Tan (2006) confirmed and extended this finding by showing that the recency effect for 16-word lists in immediate free recall was unaffected when the presentation of each list word was alternated with the presentation and serial recall of six-digit lists. The authors concurred with the earlier conclusion drawn by Baddeley and Hitch, asserting that a single short-term memory system could not be supporting both immediate serial recall and recency in immediate free recall. Converging evidence for this conclusion was found in studies demonstrating dissociations between the variables influencing immediate serial recall and recency in immediate free recall

(Cortis, 2015). As a result, the majority of immediate memory theories have focused on accounting for either immediate serial recall or immediate free recall, but not both (Cortis, 2015; Ward et al., 2010).

In 2002, Ward conducted an immediate free recall experiment using overt rehearsal. Participants were tested on their recall of lists of 10, 20 and 30 words in length. The overt rehearsal procedure allowed the researchers to plot a traditional serial position curve, in which recall performance for each list position was plotted according to the order of presentation, and a functional serial position curve in which performance for each position was plotted by the order each words' last rehearsal. The researchers found a LLE: the proportion of words recalled decreased, and the number of words recalled increased as list length increased. The experiment also showed that early list items received more rehearsals than later list items. The functional serial position curve showed an extended recency effect and no primacy effect. This suggests that recency affects recall throughout the whole list and not just in the short-term memory proportion of the curve as posited by dual-store accounts of FR. Moreover, no effect of proactive interference was observed. According to dual-store accounts, items from the primacy and asymptote portion of the serial position curve are retrieved from long-term memory and should be subject to interference effects. However, Ward demonstrated that regardless of whether the lists were 10, 20 or 30 words in length, recall of the last ten rehearsal items was the same.

Ward et al. (2010) noted that this theoretical divide was surprising given the methodological similarities between the two tasks. They suggested that previous comparisons between immediate free recall and immediate serial recall were inaccurate due to the different list lengths typically used in each task. Instead, the authors argued that similarities between the two tasks emerge when participants are tested under identical conditions, using the same list lengths. To support this hypothesis, Ward et al. (2010) tested participants using immediate free recall and immediate serial recall on lists ranging from one to 15 words. The results indicated that, regardless of the task, participants tended to initiate recall with the first list item when tested using short lists, and this tendency decreased as list length increased. Similarly, in both tasks, the tendency for forward order recall was greatest for short lists and decreased as a function of increasing list length. These output order patterns also translated to differences in the SPCs. A greater primacy effect was observed when participants initiated recall with the first list item, and a greater recency effect was observed when participants initiated recall with one of the last list items. Thus in both tasks, the shape typically associated with an immediate serial recall-SPC occurred for shorter list lengths, and the shape associated with an immediate free recall-SPC was found for longer list lengths. As anticipated, these results suggested a need for greater theoretical integration between accounts of immediate free recall and immediate serial recall.

Moreover, Bhatarah, Ward and Tan (2008) had previously demonstrated that the same encoding strategies were employed in immediate free recall and immediate serial recall. Using both within- and between-subject designs, they manipulated test expectancy by informing participants whether they would be tested using a free recall or serial recall procedure before (pre-cued) or after (post-cued) the to-be-recalled lists were presented. Although recall differed between the immediate free recall and immediate serial recall conditions, there was no difference between recall in the pre-cued or post-cued conditions. These results suggest that the different SPCs associated with immediate free recall and immediate serial recall reflect retrieval rather than encoding differences (Bhatarah, Ward, Smith & Hayes, 2009). Additional research has shown that variables such as presentation rate, word length, concurrent articulation, phonological similarity and knowledge of list length have similar influences on immediate free recall and immediate serial recall (Bhatarah et al., 2009; Grenfell-Essam & Ward, 2012; Spurgeon et al., 2014, 2015).

The Set Size-Accuracy Benchmark

Oberauer et al. (2018) proposed a collection of robust memory phenomena as benchmarks against which a researcher could assess the explanatory strength of any theory or model of working and short-term memory. The first benchmark identified in this paper concerns the effect of set size on retrieval accuracy. Empirical evidence suggests that across stimulus-type and memory paradigm, the accuracy of retrieval from WM decreases as a function of the number of items a participant is asked to retain. Thus, as set-size increases, the accuracy of recall decreases. Oberauer et al. classified the ability of a model or theory to explain this effect as a high priority benchmark. They justified this classification by suggesting that the effect of set size on accuracy provides evidence for the capacity limitations that are central to the conceptualization of WM.

Evidence from Spatial Memory. Evidence from a series of studies investigating spatial WM appears to support the set size-accuracy principle identified by Oberauer et al. (2018). Shah and Miyake (1996) demonstrated that memory performance decreased as a function of set size in three spatial orientation tasks. In all three tasks, participants were tested on their ability to remember the orientation arrows individually presented in sets ranging from two to six. In a simple span task, the arrows were individually presented in an uninterrupted sequence, while in two complex span tasks, the presentation of each arrow was alternated with a verification or rotation task, respectively. In the verification task, participants assessed the validity of brief statements between the presentation of each arrow and in the rotation task, participants determined whether a rotated letter was depicted normally or as a mirrored-image. Across all three tasks, as set size increased, recall performance deteriorated. This occurred more rapidly in the complex span tasks in which participants were required to both maintain and process information than in the simple span task, which did not require processing. This aligns with the view that WM resources are distributed between processing and maintaining functions.

Similarly, Oberauer and Kliegl (2006) used a visual memory updating task to estimate WM capacity. In this task, participants saw a set of frames positioned to form a circle on the visual display. Each frame contained a dot located in one of nine possible locations on an invisible three-by-three grid. When the dots disappeared, individual arrows were presented sequentially in each frame, with the presentation of each arrow moving in a serial clockwise direction. The arrows pointed in different directions and indicated to participants the direction in which to mentally shift the position of the dot held in each frame. After eight mental shifts, participants were asked to indicate the new location of the dot in each frame. Oberauer and Kliegl varied the number of frames presented in each trial so that the set size ranged from one to four. The proportion of correct responses decreased as a function of increasing set size.

The same pattern was found by Cortis, Dent, Kennett, and Ward (2015) when participants were tested on memory for spatial location across a wider range of set sizes. In this study, participants were shown a set of rectangles individually presented in different positions across a screen. Participants were asked to recall the location of the rectangles in a free recall task, and list length was manipulated between trials such that recall was tested for eleven different set sizes (one to eight, 10, 12, and 15). As predicted by the set size-accuracy principle, the proportion of locations correctly recalled decreased as set size increased. Using the same list lengths, Cortis et al. (2015) tested tactile spatial memory by tapping different sequences across participants’ faces and testing their recall. Again, the accuracy of responses decreased as the sequence lengths increased.

In a fourth study, Woods, Wyma, Herron and Yund (2016) piloted a computerized spatial span test (C-SST) developed to assess visuospatial WM span. In this task, participants are shown ten red squares randomly located on a grid of 50 possible locations. On each trial, the position of the squares is varied, and a white cursor indicated a sequence from among the squares. Participants were then tested on their serial recall for the sequences presented. Wood et al. (2016) tested participants using list lengths of three to eight and found that the proportion of correct responses decreased monotonically as list length increased. Collectively, evidence from these four spatial WM studies seems to support the validity of the first benchmark outlined by Oberauer et al. (2018). In simple span, complex span, memory updating, free recall, and serial recall tasks, the proportion of correctly recalled items decreased as set size increased.

Evidence from Visual Memory. In visual working memory, Luck and Vogel (1997) provided similar evidence supporting the assertion that the accuracy of WM performance declines as set size increases. Participants were shown a sample array of coloured squares, then, after a brief delay, a second test array was presented. The test array was either identical to the first display or differed in terms of a single feature. Participants made a change-detection decision indicating whether the sample and test arrays were identical or different. Set size was manipulated such that participants were shown arrays ranging in size from one to 12 squares. Performance was consistently high for set sizes one to three but decreased monotonically as set size increased from four to 12. In a follow-up experiment, the researchers used a partial reporting procedure whereby participants were cued to make a change detection judgment about a single specific feature within the test array. Again, performance was nearly perfect for arrays containing one to three squares and then declined as set size increased from four to 12. These results also held true when participants were shown arrays containing bars in which the variable feature was orientation instead of colour.

A more recent study, conducted by Wilken and Ma (2004), further extended the examination of set size effects in visual memory. Again, participants were asked to make change detection judgments between two briefly presented arrays. In three separate experiments, participants were tested on their ability to detect colour change, change in spatial frequency, and change in orientation. Across all three experiments, the hit rate monotonically decreased, and the false alarm rate monotonically increased as a function of increasing set size. In a second set of experiments, recall for colour, spatial frequency, and orientation was again tested, this time using a continuous reproduction paradigm. In each task, arrays containing two, four, six or eight stimuli were presented. In the colour recall task, participants were shown arrays of coloured circles, while in the spatial frequency and orientation tasks, participants were shown arrays of Gabor patches. A brief delay followed the presentation of each sample array. Then participants were cued to recall the target feature of a single stimulus from the sample array. In the colour recall task, participants attempted to match the colour of the probed circle as closely as possible on a colour-wheel. In the spatial frequency and orientation tasks, participants attempted to match either the target feature of a probed Gabor patch by using the arrow keys to manipulate the property on a probe stimulus. Across all three tasks, as set size increased, responses varied more significantly from the target response. Thus, in all experiments reported by both Luck and Vogel (1997) and Wilken and Ma (2004), visual WM accuracy seems to decline as a function of increasing set size. In both change detection and continuous reproduction tasks, the set size-accuracy principle outline by Oberauer et al. (2018) is supported.

Evidence from Verbal Memory. Evidence in support of this principle also exists in studies of verbal WM. In a free recall task, Grenfell-Essam, Ward and Tan (2013) tested participants on their recall of lists between two and 15 words in length. Across three experiments, the researchers manipulated the speed of stimulus presentation and use of concurrent articulation and overt rehearsal techniques. Nevertheless, the proportion of words recalled consistently decreased as a function of increasing list length. Likewise, in the fourth experiment, Cortis, Dent, Kennett and Ward (2015) found that regardless of concurrent articulation, the proportion of items correctly recalled decreased as list length increased. Again using free recall, Cortis Mack et al. (2017) piloted a new technique for conducting list learning tasks with large inter-stimulus presentation intervals without requiring participants to spend extended periods in the lab. The researchers used an iPhone app to present stimuli at a rate of one item per hour. In the second of three experiments, over the course of 50 days, participants saw 10 lists each of two, four, six, eight and ten words in length. For each list, an hour after the last word was presented, participants completed a free recall test. The results of this experiment indicated a standard LLE. The mean proportion of words recalled decreased as list length increased, yet the mean number of words recalled increased with list length. Thus the list length effect in free recall appears to be timescale invariant.

Similarly, in a free recall task, Grenfell-Essam and Ward (2012) found that the proportion of words recalled correctly decreased as list length increased. This pattern held true both when participants knew the list length prior to each trial and when the list lengths were unknown to participants at encoding. Likewise, in an immediate serial recall task, the researchers also showed that the proportion of correctly recalled words decreased as a function of increasing set size in both known and unknown list length conditions. These results are consistent with an earlier serial recall study which demonstrated a decrease in the proportion of correctly recalled digits, letters, and words corresponding to increases in the number of items participants were asked to recall (Crannell & Parrish, 1957).

Bunting, Cowan and Saults (2006) used a running span task in which participants were presented with lists of 12, 14, 16, 18 and 20 digits and were asked to remember the last seven digits. This task requires participants to continuously update the information they are holding in memory as the number of presented items increases beyond seven. Updating tasks are thought to more effectively measure both storage and processing components of WM (Wilhelm et al., 2013). Across trials, participants were prompted to recall a specific number ranging from one to seven of the last list items. Overall, and for each serial position, the proportion of correctly recalled items decreased as list length increased. McElree and Dosher (1989) found the same pattern of results using a probed recognition task in which participants were tested on recall for lists of three to six words. In each trial, participants were shown a list of words sequentially presented on a screen. After the list presentation, participants were given a probe word and were asked to make an old-new recognition judgment as rapidly and as accurately as possible. The proportion of incorrect recognition judgments was highest for lists of six words and decreased systematically as list length decreased.

In an n-back task, participants are shown a series of items and are asked to indicate whether each item matches the one presented n positions back in the series. To accurately respond, participants must continuously update the information they hold in memory and the larger the assigned n-value, the more items participants must retain. Jonides et al. (1997) measured brain activation using Positron Emission Tomography (PET) while participants completed zero-, one-, two- and three-back tasks on a list of 45 letters. They found greater activation of the brain as the n-value increased and a corresponding decrease in performance. Thus, as the number of letters participants had to maintain in memory increased, the accuracy of performance decreased. Similarly, Verhaeghen and Basak (2005) individually presented participants with digits positioned in rows n-items long. After the first row, participants indicated whether each subsequently presented digit matched the one presented at the same position in the row above. Set size was manipulated by varying the n-value from one to five across trials. As in the n-back task, the number of items a participant had to hold in memory increased as the n-value increased. The proportion of correct responses in each task decreased as the n-value increased. Collectively, evidence from verbal memory studies supports the set size-accuracy principle across free recall, serial recall, running span, recognition, and n-back paradigms.

The set-size accuracy relationship proposed by Oberauer et al. (2018) is also supported by evidence from verbal memory studies involving numerical stimuli. Using a standard free recall procedure, Katz (1968) tested nine participants on their recall of 12 lists each of one, two and three double-digit numbers and on 36 lists each of four, five, six and seven double-digit numbers. Performance was near perfect on list lengths one to three, However, the proportion of numbers correctly recalled decreased as list length increased from four to seven. Similarly, Oberauer and Kliegl (2001) used a numerical updating task in which digits were initially presented inside individual rectangles forming a circular display. After the initial presentation, the digits disappeared and were replaced by a series of eight individually presented arithmetic operations, each displayed in a different rectangle that had previously contained a digit. Participants were asked to update the number corresponding to each rectangle by applying the presented operation to the number held in memory. After all eight operations had been presented, participants were asked to enter the new numbers in each position. Across multiple trials, participants were initially presented with and updated one, two, three, four, five, and six digits. In both young and older adults, recall performance declined as set size increased.

The List Length Effect

The majority of research operationalizes recall accuracy as the proportion of correct responses in a recall test. However, it is also possible to assess recall accuracy by considering the absolute number of correct responses made by a participant. The list length effect (LLE) is a well-known phenomenon in immediate free recall and refers to the finding that, while the proportion of correctly recalled items decreases as set sizes increases, the total number of recalled items continues to increase with set size (Murdock, 1968). Two studies, conducted by Murdock (1962) and Roberts (1972), demonstrate a LLE in immediate free recall. Murdock (1962) tested participants' memory for lists of 10, 15 and 20 words. He found an increase in the total number of words recalled and a corresponding decrease in the proportion of words recalled as list length increased. Roberts (1972) included a greater range of list lengths, testing free recall for lists of ten, 20, 30 and 40 words, presented both visually and verbally. This second study, therefore, provides more insight into the relationship between list length and number of words recalled. Again, the results show a monotonic increase in the number of words recalled as list length increased regardless of the mode or rate of word presentation. Simultaneously, the proportion of correctly recalled words monotonically decreased as list length increased.

Unsworth and Engle (2007) also demonstrated an initial LLE in immediate serial recall. In a study conducted by Unsworth and Engle (2006), 235 participants were tested on their recall of lists of either words or letters ranging in length from two to seven items. In both simple and complex span tasks, the proportion of correctly recalled items decreased as list length increased. At the same time, the mean number of recalled words increased across the simple span list lengths of two to five before subsequently decreasing between list lengths of five and six and again between list lengths of six and seven. In the complex span task, the mean number of recalled words increased between list lengths two and three before reaching a plateau and subsequently decreasing between list lengths four and five. Similarly, Adam, Mance, Fukuda and Vogel (2015) showed evidence of the LLE in visual memory. They used both change-detection and whole-report tasks to measure recall for visually presented arrays composed of coloured squares. In both tasks, participants made judgments about arrays that contained either two, three, four, five, or six coloured squares. In the change detection task, participants were first presented with the sample array, then, after a brief retention interval, a probe square. Participants indicated whether the probe square matched the colour of the square that had been in the corresponding position in the sample array. Alternatively, in the whole-report task, at recall, participants selected the colour of each square seen in the sample array from a choice of nine colours. For both tasks, the mean number of correct responses initially increased with set size before reaching a stable plateau in performance that was subsequently maintained. Simultaneously, the proportion of correct responses in both tasks decreased as a function of increasing set size. These results would be anticipated under fixed-capacity models of WM, as a plateau or decrease in the number of items recalled would be expected once working memory capacity is reached.

In contrast, using the same paradigm and stimulus type as Unsworth and Engle (2006), Beaman found a continued increase in the mean number of words recalled across lists lengths five to nine in a simple span serial recall task. Beaman (2006) also demonstrated that these results are predicted by the feature model proposed by Nairne (1990). Moreover, by reanalyzing the results of a short-term serial order study conducted by Drewnowski (1980), he found that participants in this earlier study had recalled an increasing number of words as list length increased from four to eight. Thus the prediction of the feature model - that the mean number of words recalled from a list will increase as a function of list length - is seemingly supported by the empirical data collected by Beaman (2006) and by Drewnowski (1980). This continued increase in the absolute number of items recalled is inconsistent with the view of WM as a fixed, limited capacity store. It is, however, possible to explain the presence of a LLE by suggesting that the same processes are underlying both free and serial recall despite the different time scales involved in these tasks. Beaman (2006) proposed that the LLE results from a balance between the interference effects caused by overloading and the overall increase in the amount of possible information available to be recalled. He suggested that his results could be accommodated by both unitary models of memory and models proposing interference as a mechanism of forgetting in both short-term memory and long-term memory.

Present Research

Beaman (2006)'s results are congruent with the Feature model's prediction of a constant increase in the number of items recalled as a function of list length. However, these results conflict with those of Unsworth and Engle (2006), who found that the number of words recalled initially increased with list length, but eventually decreased when list length increased further. Beaman's results also conflict with the set-size-accuracy benchmark outlined by Oberauer et al. (2018), which predicts that memory performance always decreases as a function of the number of items an individual is asked to retain. The aim of the current research was to reconcile these discrepancies. Additionally, if the LLE was demonstrated in multiple short-term memory paradigms, this would provide evidence against a limited capacity model of working memory. Across six experiments, using immediate free recall, immediate serial recall and operation span, we tested for the presence of a LLE in short-term memory. One procedural difference between

Beaman (2006) and Unsworth and Engle (2006) was participants' prior knowledge of list length during encoding. In Beaman (2006), participants were informed of the number of items in each list prior to its presentation, while participants in Unsworth and Engle (2006) were not given this information. Knowledge of list length may accommodate more effective mnemonic strategies by allowing participants to structure information strategically during encoding (Bunting et al., 2006; Grenfell-Essam & Ward, 2012). When participants are unaware of list length, they may alter their chosen encoding strategy during the presentation of longer lists. Moreover, Pollack, Johnson and Knaff (1959) demonstrated participants recalled a greater number of items when they had prior knowledge of list length. They also found that variables, such as presentation rate and item grouping differentially influenced recall for lists of known and unknown lengths. Similarly, Crowder (1968) found that the primacy effect is significantly reduced when the list length is unknown. In the current research, recall was tested both for known and unknown list lengths in order to gauge whether this variation in methodology underlay the contradictory results. I also extended the number of list lengths included in the experiments to provide a more precise understanding of the effects of list length on recall accuracy.

In the first two experiments participants' memory performance was tested using serial recall for lists of three to eight words. In Experiment 1, participants were not informed of the list lengths prior to encoding, while in Experiment 2, they were. Similarly, Experiments 3 to 6 used operation span to measure participants' recall across a range of list lengths. In Experiments 3 and 4, participants were tested with known and unknown list lengths, respectively. Lists in these two experiments ranged in length from two to six words long. Experiment 5 followed the same procedure as Experiment 3 but included different word stimuli. Lastly, in Experiment 6 recall was tested for lists of 3 to 8 words. In all experiments, performance was measured both in terms of the number and proportion of words recalled.

Predictions and Hypothesis. The dominant explanation of the LLE suggests that it is a long-term memory phenomenon. Therefore, it was not anticipated that a LLE would occur in either the immediate serial recall or operation span experiments. A decline in the number and proportion of words recalled as a function of increasing list length was anticipated in all experiments. These results would provide evidence against the presence of a LLE in short-term memory and support the results of Unsworth and Engle (2006) over those of Beaman (2006). By confirming the results presented by Unsworth and Engle, the current research would also support the set-size accuracy benchmark proposed by Oberauer et al. (2018). It was also anticipated that overall performance would be higher in those experiments in which participants had prior knowledge of list length before each trial. Informing participants of the length of each list allows them to apply more effective encoding strategies, which should enhance memory performance. Knowledge of list length was not, however, expected to impact the overall pattern of results and was not expected to impact whether a LLE occurred.

General Methods

Justification of sample size

Serial recall experiments

Simulation-based a priori power analyses were conducted using the R package Superpower developed by Lakens and Caldwell (2019). For Experiments 1 and 2, a sample size of 30 was determined to be sufficient to detect the main within-subject effects with 90% power ($\alpha = .05$) for a repeated measures ANOVA. The power analyses were performed using the

Greenhouse-Geisser correction. The mean number of recalled words for each list length were estimated to be 2.9, 3.8, 4.3, 4.7, 5.1, and 5.2 for list length 3, 4, 5, 6, 7, and 8, respectively. These estimations were based on Unsworth and Engle (2006; as cited in Unsworth & Engle, 2007) and Beaman (2006). Standard deviations were estimated to be 0.25, 0.50, 0.75, 1.00, 1.25, and 1.5 for list length 3, 4, 5, 6, 7 and 8, respectively. The Pearson correlations between repeated measures were assumed to be 0.5 between list lengths that differ by one word in length, 0.25 between list lengths that differed by two or three words in length, and 0 between list lengths that differed by four or five words in length. These correlation estimates were conservative to reduce the possibility of overestimating power. Power was estimated based on 10 000 simulations.

Operation span experiments

An a priori power analyses was conducted for Experiments 3 to 5 using the R package Superpower (Lakens & Caldwell, 2019). The same procedure was repeated for Experiment 6. For all experiments, a sample size of 30 was determined to be sufficient to detect the main within-subject effects with 90% power (α = .05) for a repeated measures ANOVA. The power analysis was performed using the Greenhouse-Geisser correction. The mean number of recalled words for each list length was estimated based on Unsworth and Engle (2006; as cited in Unsworth & Engle, 2007). For Experiments 3 to 5, the estimated means for each list length were 1.80, 2.30, 2.30, 3.00, and 1.90 for list lengths 2, 3, 4, 5, and 6, respectively. For Experiments 3 to 5, standard deviations were estimated to range from 0.20 to 1.0, increasing incrementally by 0.20 as a function of list lengths. Standard deviations were estimated to be 0.25, 0.45, 0.65, 0.85 and 1.05, for list length 2, 3, 4, 5, and 6, respectively. For Experiment 6, standard deviations ranged from 0.40 to 1.40, increasing incrementally by 0.20 with list lengths. The standard deviations were estimated to be 0.40, 0.60, 0.80, 1.00, 1.20, and 1.40 for list lengths 3, 4, , 5, 6, 7 and 8, respectively. The Pearson correlations between repeated measures were assumed to be 0.5 between list lengths that differ by one word in length, 0.25 between list lengths that differed by two or three words in length, and 0 between list lengths that differed in length by four or five words. Power was estimated based on 10 000 simulations.

Stimuli

The stimuli were one syllable words selected to avoid any unusual or emotional language. The word stimuli used in Experiment 1, 2, 5 and 6 are included in Appendix B. The word stimuli used in Experiment 3 are included in Appendix C. The mathematic operations used in Experiment 3, 4 and 5, are included in Appendix D and the mathematic operations used in Experiment 6 are included in Appendix E.

Experiment 1: Serial Recall with Unknown List Lengths

Experiment 1 was adapted from Beaman's (2006) experiment. Beaman tested recall performance on lists of five, seven and nine words in length. In the current experiment, the number of list lengths examined was extended to include lists ranging from three to eight words in length. The longest list length used was set at eight based on performance levels observed in other experiments. Similarly, while Beaman's participants were informed of list length before each trial, in the current experiment list length was not known to participants prior to list presentation.

Method

Participants

The sample was 30 volunteers from ProlificAC. Each was paid £8 per hour, pro-rated. The following inclusion criteria were used for this and all subsequent experiments: (1) Native speaker of English; (2) approval rating of at least 90% on prior submissions at ProlificAC; and (3) age between 19 and 39. The mean age of participants was 29.43 (*SD* = 5.73, range 20-39), and 15 self-identified as female and 15 self-identified as male.

Stimuli

The word stimuli consist of 959 one-syllable words selected to avoid any unusual or emotional language. These stimuli are listed in Appendix B

Design

The experiment followed a within-subject design with list length as a six-level within-subject factor (list lengths: three, four, five, six, seven, eight). The dependent variables were proportion and number of correctly recalled words.

Procedure

Participants provided consent, responded to a brief demographics questionnaire, then received instructions for the experimental task. The task was programmed using Java Script by Ian Neath. The experiment followed a standard serial recall procedure. On each trial, participants were shown a series of individually presented words. Each word was displayed in the center of the screen in 28 point Helvetica for one second. After each list was fully presented, participants were asked to recall as many of the words as possible in the presentation order. They entered their responses by typing the words into a grid box centered on the screen. If participants could not recall a word corresponding to a particular position, they had the option to guess or skip that position. Each word was individually prompted so that participants had to enter list words in the order of presentation. Recall was self-paced and had no time limit. Participants completed 36 trials: six trials each of list lengths three to eight. List order was randomized and different for all participants. Randomization ensured that participants did not know the length of a particular list in advance of the trial.

Results

Serial Recall Scoring

Responses were analyzed using serial recall scoring: recall of a list was considered correct only when participants recalled all the list words in their correct serial positions. Figure 1 shows the mean number and proportion of words recalled as a function of list length.

Two one-way repeated measures analyses of variance (ANOVAs) were conducted to verify the presence of the LLE. In both ANOVAs, the assumption of sphericity was violated, and the Greenhouse-Geisser correction was applied. The first ANOVA revealed a significant main effect of list length on the proportion of words recalled, $F(3.20, 92.69) = 131.46$, $p < .001$, $\eta_p^2 = 0.819$, $MSE = 0.019$. Post-hoc analysis showed that all pairwise comparisons were significant (*all ps < .05*). The mean proportion of recalled words decreased as a function of list length (see Table 1). Planned contrasts indicated a significant linear effect, $t\ (145) = -25.45$, $p < .001$. Thus, the first component of the LLE was present: as list length increased, the proportion of correctly recalled items decreased.

The second analysis revealed a significant main effect of list length on the number of words recalled, $F(2.54,73.67) = 5.01$, $p=.005$, $\eta_p^2=.147$, $MSE = 0.991$. Post-hoc analysis showed that the number of words recalled from lists of three words was significantly lower than for both list lengths four and five. The number of words recalled for list length five was also significantly greater than for list length eight. No other pairwise comparisons were significant ($\alpha = .05$). As list length increased, the proportion of correctly recalled words initially increased from list length three to list length five before decreasing from list length five to list length eight (see Table 1). Planned contrasts indicated a significant quadratic effect, *t* (145) = -4.19, *p* < .001.

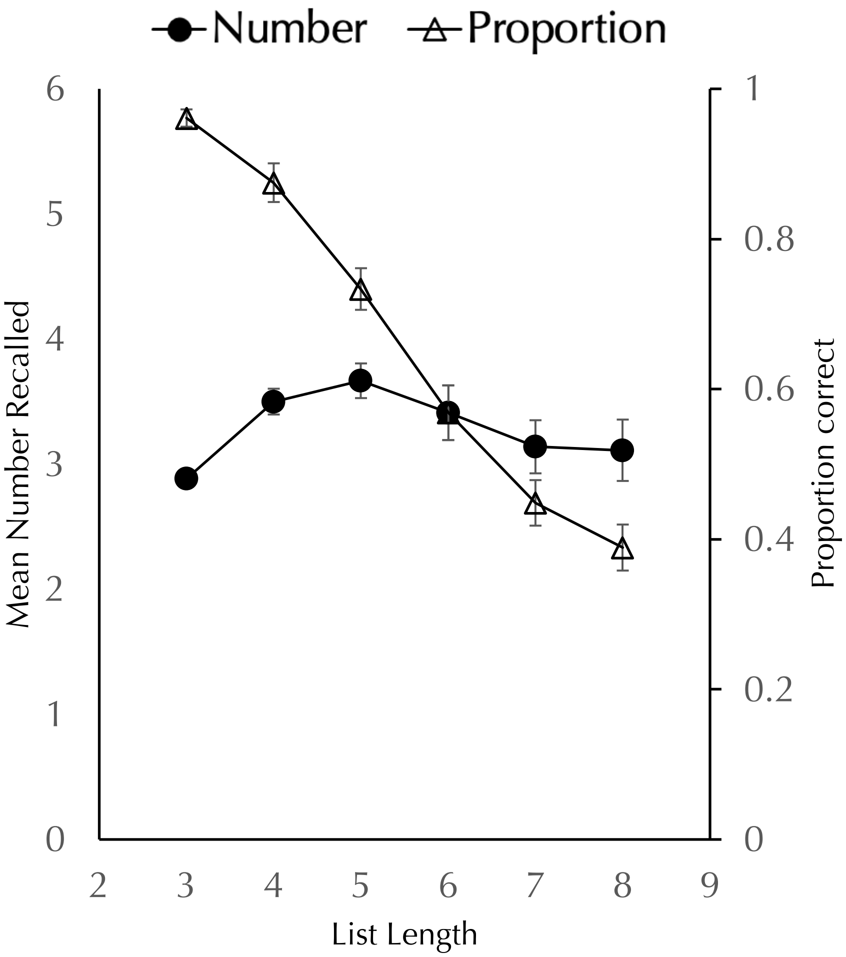

Figure 1. Mean proportion and mean number of correctly recalled words in Experiment 1 – immediate serial recall with unknown list lengths. Error bars represent the standard error of the mean.

The quadratic trend in the number of words recalled indicates that as list length increased, participants initially recalled more words. However, as list length increased further, participants recalled fewer words. This trend was not observed by Beaman (2006), but is consistent with Unsworth and Engle (2006). Moreover, the eventual decrease in the number of words recalled as a function of increasing list lengths is anticipated under the set-size accuracy benchmark. Thus, although the first component of the LLE was present, the second component was not: the absolute number of items recalled did not increase as list length increased.

Table 1

Mean and standard deviation scores for the number and proportion of correct responses in Experiment 1 with both serial recall and free recall scoring.

	List length	Three	Four	Five	Six	Seven	Eight
Serial recall scoring	number	2.88	3.50	3.67	3.44	3.14	3.11
	correct (SD)	(0.19)	(0.56)	(0.75)	(1.18)	(1.14)	(1.32)
	proportion	0.96	0.88	0.73	0.570	0.45	0.39
	correct (SD)	(0.06)	(0.14)	(0.15)	(0.20)	(0.16)	(0.17)
Free recall scoring	number	2.88	3.61	4.03	4.13	4.10	4.23
	correct (SD)	(0.19)	(0.43)	(0.48)	(0.81)	(0.84)	(1.22)
	proportion	0.96	0.90	0.81	0.69	0.59	0.53
	correct (SD)	(0.06)	(0.11)	(0.10)	(0.14)	(0.12)	(0.15)

Free Recall Scoring

Crowder (1969) compared performance on an immediate free recall task and an immediate serial recall task scored using a free recall criterion. Despite the method of scoring being the same in both tasks, participants showed better recall under immediate free recall than immediate serial recall instructions. Crowder concluded that participants adopt different acquisition strategies based on the task instructions they are given. Thus, results of an immediate serial recall task scored using a free recall criterion do not necessarily provide an accurate estimate of immediate free recall performance. As such, the following analysis was included for completeness but showed to interpreted with caution.

Responses were re-analyzed using free recall scoring: recall of list words were considered correct regardless of the order in which they were recalled. A repeated-measures ANOVA with the Greenhouse-Geisser correction revealed a significant main effect of list length on the proportion of words recalled, $F(2.52, 72.96) = 117.99$, $p < .001$, $\eta_p^2 = 0.803$, $MSE = 0.015$. Post-hoc analysis showed that all pairwise comparisons were significant (*all ps < .05*). As list length increased, the proportion of words recalled decreased (see Table 1). Planned contrasts indicated a significant linear effect, $t\,(145) = -24.16$, $p < .001$. As with serial recall scoring, the first component of the list length effect was present.

A second repeated-measures ANOVA with the Greenhouse-Geisser correction revealed a significant main effect of list length on the number of words recalled, $F(2.01, 58.16) = 23.60$, $p<.001$, $\eta_p^2=.449$, $MSE = 0.829$. Post-hoc analysis showed that the number of words recalled was significantly lower for list length three than for all other list lengths (*all ps < .05*). Similarly, the number of words recalled for list length four was significantly lower than for list lengths five to eight. No other comparisons were significant ($\alpha = .05$). As list length increased the absolute number of words recalled initially increased from list length three to five, before reaching a plateau between list lengths five and eight (see Table 1). Planned contrasts indicated a significant linear effect, $t\,(145) = 9.42$, $p < .001$, and a significant quadratic effect, $t\,(145) = -4.97$, $p < .001$. As with serial recall scoring, although the first component of the list length effect was present, the second component was not.

Experiment 2: Serial Recall with Known List Lengths

The results of Experiment 1 were consistent with the predictions of the set-size accuracy benchmark (Oberauer et al., 2018) and with the results of Unsworth and Engle (2006). The results were not consistent with Beaman (2006) and did not show evidence of the LLE in serial recall. However, one key procedural difference between Unsworth and Engle (2006) and Beaman (2006) was whether participants had prior knowledge of list length before each trial. In Unsworth and Engle (2006) list length was unknown, while in Beaman (2006) the list presentation order was sequential, therefore, list length was known to participants before each trial. This difference may account for the discrepancy in results between the two studies. The second experiment investigates whether prior knowledge of list length impacts the number or proportion of words recalled in serial recall across multiple list lengths. While in Experiment 1 list length was unknown, in Experiment 2 participants were given prior knowledge of list length before each trial.

Method

Participants

Thirty different volunteers from ProlificAC participated in the second experiment. Each was paid £8 per hour, pro-rated. The following inclusion criteria were used for this and all subsequent experiments: (1) Native speaker of English; (2) approval rating of at least 90% on prior submissions at ProlificAC; and (3) age between 19 and 39. The mean age of participants was 27.27 *SD* = 5.06, range 20-39), and 16 self-identified as female and 14 self-identified as male.

Stimuli

The stimuli were the same as those used in Experiment 1.

Design

The experiment followed a within-subject design with list length as a six-level within-subject factor (list lengths: three, four, five, six, seven, eight). The dependent variables were proportion and number of correctly recalled words.

Procedure

Experiment 2 followed the same procedure outlined in Experiment 1, with the exception that the order of list presentation was not randomized. Instead, list order was blocked, such that all six lists of the same length were presented sequentially, and the overall presentation order of lists was from shortest to longest. Participants were informed of the list length of each trial prior to its presentation.

Results

Serial Recall Scoring

Responses were analyzed using serial recall scoring. Recall of a list was considered correct only when participants recalled all the list words in their correct serial positions. Figure 2 shows the mean number and proportion of words recalled as a function of list length.

As in Experiment 1, two analyses were conducted to verify the presence of the two components of the LLE. In both ANOVAs, the Greenhouse-Geisser correction was applied as the assumption of sphericity was violated. A repeated measures ANOVA revealed a significant main effect of list length on the proportion of words recalled $F(3.70, 107.38) = 127.21$, $p < .001$, $\eta_p^2 = .814$, $MSE = 0.016$. Post-hoc analysis showed that all pairwise comparisons were significant (*all ps < .05*). The mean proportion of words recalled decreased as a function of list length (see Table 2). Planned contrasts indicated a significant linear effect, t (145) = -25.15, $p <$

.001. Thus, the first component of the LLE was present: as list length increased, the proportion of correctly recalled items decreased.

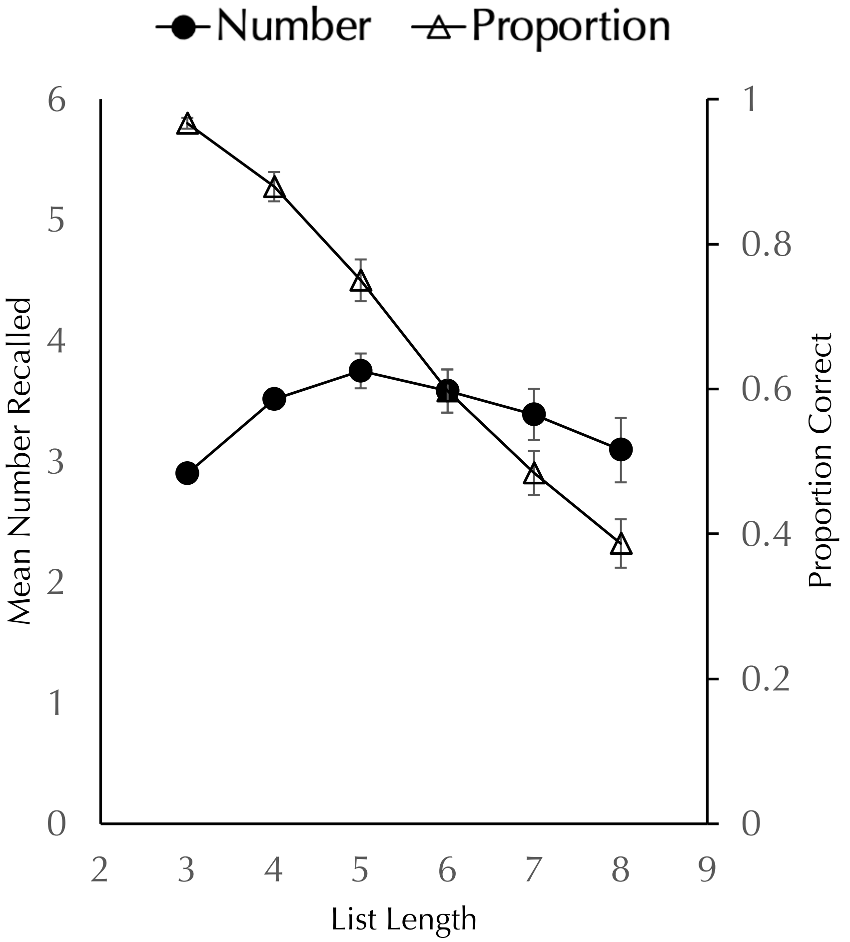

Figure 2. Mean proportion and mean number of correctly recalled words in Experiment 2 – immediate serial recall with known list lengths. Error bars represent the standard error of the mean.

A second repeated measures ANOVA revealed a significant main effect of list length on the number of words recalled, *F*(2.78, 80.44) = 5.73, p<.001, η_p^2=.165, *MSE* = 0.961. Post-hoc analysis revealed that the mean number of words recalled for list length three was significantly lower than the number recalled for list lengths four, five and six. The number of words recalled from list length five was also significantly greater than the number of words recalled from list lengths eight. No other pairwise comparisons were significant (*p* > .05). As list length increased, the number of correctly recalled words increased from list length three to five, before decreasing from list length five to eight (see Table 2). Planned contrasts indicated a significant quadratic effect, *t* (145) = -5.13, *p* < .001. Thus, although the first component of the LLE was present, the second component was not: the absolute number of recalled items did not continue to increase with increasing list length.

Table 2

Mean and standard deviation scores for the number and proportion of correct responses in Experiment 2 with both serial recall and free recall scoring.

	List length	Three	Four	Five	Six	Seven	Eight
Serial recall scoring	number	2.90	3.52	3.75	3.58	3.39	3.09
	correct (SD)	(0.12)	(0.43)	(0.78)	(0.96)	(1.14)	(1.44)
	proportion	0.97	0.88	0.75	0.60	0.48	0.39
	correct (SD)	(0.04)	(0.11)	(0.16)	(0.16)	(0.16)	(0.18)
Free recall scoring	number	2.90	3.51	3.76	3.61	3.46	3.12
	correct (SD)	(0.12)	(0.51)	(0.76)	(0.97)	(1.13)	(1.42)
	proportion	0.97	0.88	0.75	0.61	0.49	0.39
	correct (SD)	(0.04)	(0.126)	(0.15)	(0.16)	(0.16)	(0.18)

Free Recall Scoring

Responses were re-analyzed using free recall scoring. In this case, recall of list words were considered correct regardless of the order in which participants entered the words. A repeated-measures ANOVA revealed a significant main effect of list length on the proportion of words recalled, $F(5, 145) = 123.48$, $p < .001$, $\eta_p^2 = .810$, $MSE = 0.012$. Post-hoc analysis showed that all pairwise comparisons were significant (*all ps < .05*). As with serial recall scoring, as list length increased, the number of words recalled decreased (see Table 2). Planned contrasts indicated a significant linear effect, $t(145) = -24.78$, $p < .001$. The first component of the list length effect was present.

The second repeated-measures ANOVA revealed a significant main effect of list length on the number of words recalled, $F(2.86, 82.84) = 5.86$, $p = .001$, $\eta_p^2 = .268$, $MSE = 0.922$. Post-hoc analysis showed that the number of words recalled was significantly lower for list length three than for list lengths four, five, six and seven (*all ps < .05*). Similarly, the number of words recalled for list length five was significantly greater than for list lengths eight. No other comparisons were significant ($p > .05$). As list length increased, the absolute number of words recalled initially increased from list length three to five, before decreasing from list length five to eight (see Table 2). Planned contrasts indicated a significant quadratic effect, $t(145) = -5.19$, $p < .001$. Consistent with analysis performed using serial recall scoring, although the first component of the list length effect was present, the second component was not. The absolute number of items recalled did not continue to increase as a function of list length.

Experiment 3: OSPAN with Unknown List Lengths

The results of both Experiments 1 and 2 support the set-size accuracy principle outlined by Oberauer et al (2018) and are consistent with Unsworth and Engle (2006) and inconsistent Beaman (2006). Regardless of whether list length is known or unknown, the proportion of words recalled decreased as a function of list length. The number of words recalled initially increased between list lengths three and five, before reaching a plateau at list length five. Thus, as set-size increased, memory performance decreased.

Experiment 3 was conducted to determine whether results consistent with those of the Experiment 1 and 2 would be observed using another short-term memory task. In Experiment 3, recall was tested using operation span. Should the result be consistent with those obtained in the serial recall experiments, this would strengthen the evidence that memory performance in short-term memory decreases as set-size increases, as stated in the set-size accuracy principle. In Experiment 3, participants did not have prior knowledge of list length before each trial.

Method

Participants

Thirty different volunteers from ProlificAC participated in the third experiment. Each was paid £8 per hour, pro-rated. The following inclusion criteria were used for this and all subsequent experiments: (1) Native speaker of English; (2) approval rating of at least 90% on prior submissions at ProlificAC; and (3) age between 19 and 39. The mean age of participants was 28.933 (*SD* = 6.079, range 19-39), 20 self-identified as female and 10 self-identified as male.

Stimuli

The word stimuli consisted of 66 nouns and 66 math problems taken from Conway and Engle (1996). The word stimuli are included in Appendix C and the math problems are listed in Appendix D

Design

The experiment followed a within-subject design with list length as a five-level within-subject factor (list lengths: two, three, four, five, and six). The dependent variables were proportion and number of correctly recalled words.

Experimental Task

The operation word span (OSPAN) task was adapted from Turner and Engle (1989). The OSPAN task requires participants to hold a list of words in memory while completing a series simple of arithmetic operations (e.g., 5 + 7 = 12). Participants were presented with a series of words alternated with a single arithmetic operation. For each operation they were asked to verify whether or not the sum is correct. On a subsequent memory test, participants were asked to recall the presented words in the order of presentation. Participants were asked to maintain a high level of accuracy in their responses to the arithmetic operations in order to discourage participants from promoting recall by limiting the amount of attention devoted to the operations.

Procedure

Participants provided consent and responded to a brief demographics questionnaire. They were then reminded of the task instructions, before beginning the OSPAN task. The task was programmed using Java Script by Ian Neath. On each trial, participants were shown a list of words alternated with arithmetic sums. Each word was displayed in the center of the screen for one second. Immediately after each list, participants were asked to recall as many of the presented words as possible in the order of presentation. They entered their responses by typing the words into a grid box centered on the screen. Each word was individually prompted so that participants had to enter list words in the order of presentation. Recall was self-paced and there was no time limit. Participants had the option to skip a word if they could not recall it. The first three trials consisted of practice lists, each of two words in length. Following the practice trials, participants completed an additional 15 experimental trials: three trials each of list lengths two to six. List presentation order was randomized and different for each participant. Randomization ensured that participants did not know the length of a particular list in advance of the trial.

Results

Results from the operation span task were analyzed including data from all participants (N = 30), and including only data from participants who scored 85% or above on the math questions (N = 17). Eliminating data from participants with lower math performance did not change the significance or pattern of the results, therefore, the analysis reported below includes all participants. [2]

Figure 3 shows the mean number and proportion of words recalled as a function of list length. Two separate ANOVAs were conducted to verify the presence of the two components of the LLE. In both ANOVAs, the Greenhouse-Geisser correction was applied as the assumption of sphericity was violated. A repeated measures ANOVA revealed a significant main effect of list length on the proportion of words recalled, $F(2.67, 77.30) = 15.29$, $p < .001$, $\eta_p^2 = 0.345$, $MSE = 0.024$. Post-hoc analysis showed that the mean proportion of words recalled for list length six was significantly lower than the mean proportion recalled for all other list lengths. Similarly, the

---

[2] ANOVA results when the analysis was limited to participants who scored 85% or above on the math questions showed a significant main effect of list length on the proportion of words recalled, $F(2.08, 33.27) = 8.02$, $p = .001$, $\eta_p^2 = 0.334$, $MSE = 0.01$, and a significant main effect of list length on the number of words recalled, $F(1.50, 24.02) = 153.21$, $p < .001$, $\eta_p^2 = 0.905$, $MSE = 0.47$

mean proportion of words recalled for list length five was significantly less than for list lengths three and two. No other pairwise comparisons were significant ($\alpha = .05$). Planned contrasts indicated a significant linear effect, *t* (116) = -7.62, $p < .001$. The mean proportion of words recalled for each list length are shown in Table 3. Thus, the first component of the LLE was present: as list length increased, the proportion of correctly recalled words decreased.

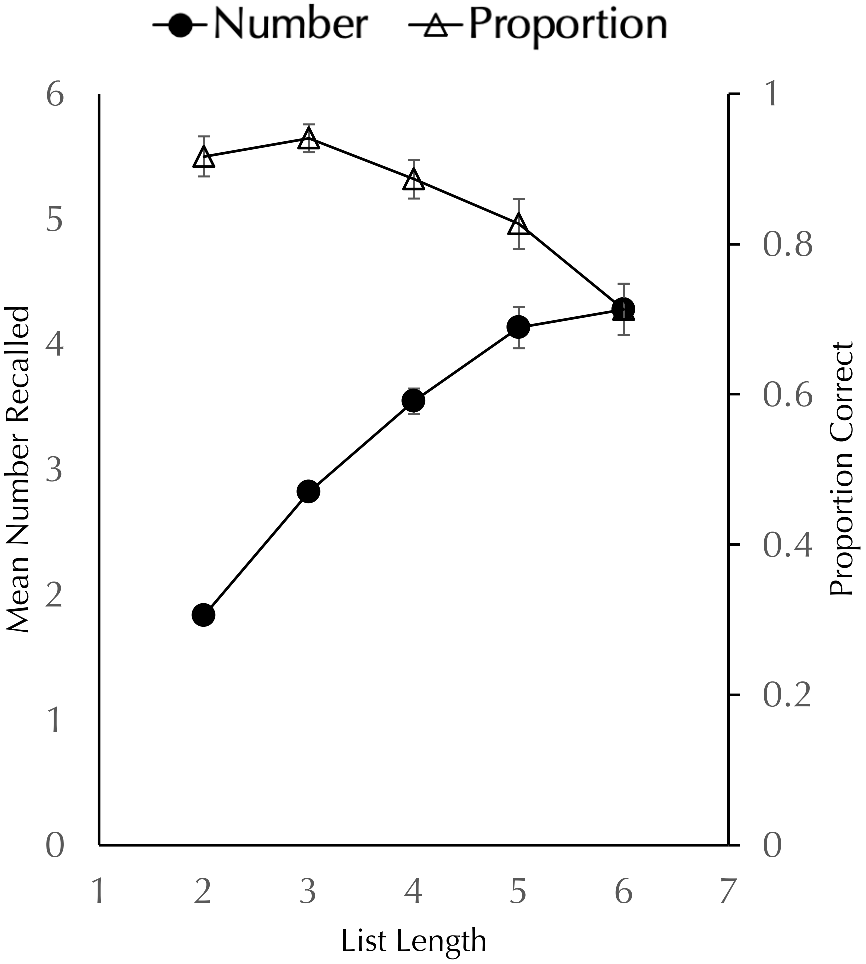

Figure 3. Mean proportion and mean number of correctly recalled words in Experiment 3 – OSPAN with unknown list lengths. Error bars represent the standard error of the mean.

The second repeated measures ANOVA revealed a significant main effect of list length on the number of words recalled, *F*(2.07,60.06) = 67.51, $p<.001$, $\eta_p^2 = 0.700$, *MSE* = 0.906. Post-hoc analysis revealed that there was so significant difference between list length five and list length six, however all pairwise comparisons were significant (*all ps < .05*). The mean number of words recalled for each list length is shown in Table 3. Planned contrasts indicated a significant linear effect, t (116) = 16.18, p < .001 and a significant quadratic effect, t(116) = -2.80, p = .006. As list length increased, the mean number of correctly recalled words increased until list length five. From this experiment it is not clear whether the number of words recalled would continue to increase with list lengths greater than six words in length.

Table 3

Mean and standard deviation scores for the number and proportion of correct responses in Experiment 3.

List length	Two	Three	Four	Five	Six
number correct	1.83	2.82	3.54	4.13	4.28
(SD)	(0.29)	(0.30)	(0.55)	(0.89)	(1.11)
proportion	0.92	0.94	0.89	0.83	0.71
correct (SD)	(0.14)	(0.10)	(0.14)	(0.18)	(0.19)

Experiment 4: OSPAN with Known List Lengths

Experiment 4 is a conceptual replication of Unsworth and Engle’s (2006) complex span experiment. However, operation span was used instead of reading span to minimize the possibility of stimulus consistent interference. Unsworth and Engle found that the number of words recalled reached a plateau between list length three and five. By contrast, in Experiment 3 of the current research, the number of words recalled increased monotonically from list length two to five. It is possible that this difference may reflect participants' prior knowledge of list length. In Unsworth and Engle (2006), participants knew the list length of each upcoming trial, while in Experiment 3 they did not. To examine how this procedural difference may influence memory performance, in Experiment 4 participants were tested using operation span with known list lengths.

Method

Participants

Thirty different volunteers from ProlificAC participated in the fourth experiment. Each was paid £8 per hour, pro-rated. The following inclusion criteria were used for this and all subsequent experiments: (1) Native speaker of English; (2) approval rating of at least 90% on prior submissions at ProlificAC; and (3) age between 19 and 39. The mean age of participants was 29.433 (*SD* = 5.847, range 19-39), 20 self-identified as female and 10 self-identified as male.

Stimuli

The stimuli were the same as those used in Experiment 3.

Design

The experiment followed a within-subject design with list length as a six-level within-subject factor (list lengths: two, three, four, five and six). The dependent variables were proportion and number of correctly recalled words.

Procedure

Experiment 4 followed the same procedure outlined in Experiment 3, with the exception that the list lengths were blocked such that all three lists of the same length were presented sequentially and the overall presentation order of the lists was from shortest to longest. In this experiment participants were informed of the length of the upcoming list before each trial.

Results

Results from the operation span task were analyzed twice: once including data from all participants (N=30), and once with only data from participants who scored 85% or above on the math questions (N=16). Eliminating data from participants with lower math performance did not change the significance or pattern of the results, therefore, the analysis reported below includes all participants.[3]

Figure 4 shows the mean number and proportion of words recalled as a function of list length. As in the previous experiments, two separate analyses were conducted to verify the presence of the two components of the LLE. In both ANOVAs, the Greenhouse-Geisser correction was applied as the assumption of sphericity was violated. A repeated-measures ANOVA revealed a significant main effect of list length on the proportion of words recalled, $F(3.18, 92.18) = 18.78$, $p < .001$, $\eta_p^2 = 0.39$, $MSE = 0.017$. Post-hoc analysis showed that the mean proportion of words recalled for list length six was significantly lower than the mean proportion recalled for all other list lengths. Similarly, a significantly greater proportion of words were recalled from list length two and three than from list length five. No other pairwise comparisons were significant ($\alpha = .05$). The proportion of words recalled decreased as a function of list length (see Table 4). Planned contrasts indicated a significant linear effect, $t\ (116) = -7.86$,

---

[3] ANOVA results when the analysis was limited to participants who scored 85% or above on the math questions showed a significant main effect of list length on the proportion of words recalled, $F(4, 60) = 5.620$, $p < .001$, $\eta_p^2 = 0.27$, $MSE = 0.004$, and a significant main effect of list length on the number of words recalled, $F(1.91, 28.61) = 297.32$, $p < .001$, $\eta_p^2 = 0.\ 27$, $MSE = 0.21$

$p < .001$ and a significant quadratic effect, t(116) = -3.57, $p < .001$. The first component of the LLE was present; as list length increased, the proportion of correctly recalled words decreased.

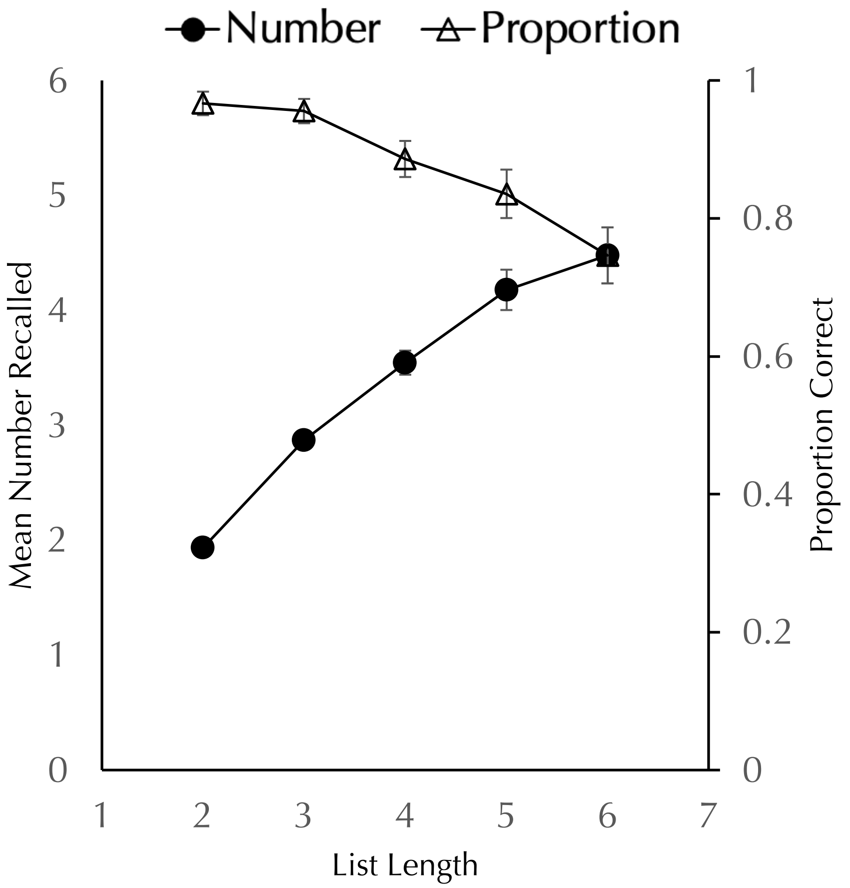

Figure 4. Mean proportion and mean number of correctly recalled words in Experiment 4 – OSPAN with known list lengths. Error bars represent the standard error of the mean.

The second analysis revealed a significant main effect of list length on the number of words recalled, $F(2.27, 65.88) = 93.34$, $p<.001$, $\eta_p^2=0.763$, $MSE = 0.578$. Post-hoc analysis revealed that there was no significant difference between the number of words recalled from list lengths five and six, however all pairwise comparisons were significant (*all ps < .05*). The mean number of words recalled for each list length is shown in Table 4. Planned contrasts indicated a significant linear effect, $t\,(116) = 18.74$, $p < .001$ and a significant quadratic effect, t(116) = -

4.66, $p < .001$. As list length increased, the proportion of correctly recalled words increased until list length five. It is, therefore, not clear whether the number of words recalled would continue to increase as a function of list length if more list lengths were included.

Table 4

Mean and standard deviation scores for the number and proportion of correct responses in Experiment 4.

List length	Two	Three	Four	Five	Six
number correct	1.93	2.87	3.54	4.18	4.48
(SD)	(0.18)	(0.29)	(0.56)	(0.95)	(1.31)
proportion	0.97	0.96	0.89	0.84	0.75
correct (SD)	(0.14)	(0.10)	(0.14)	(0.19)	(0.22)

Experiment 5: OSPAN with Unknown List Lengths

Experiment 5 was a replication of Experiment 3 conducted using different words as stimuli to ensure that the result of Experiment 3 were not an artefact of the particular stimulus-set used in that experiment.

Method

Participants

Thirty different volunteers from ProlificAC participated in the fifth experiment. Each was paid £8 per hour, pro-rated. The following inclusion criteria were used for this and all subsequent experiments: (1) Native speaker of English; (2) approval rating of at least 90% on prior submissions at ProlificAC; and (3) age between 19 and 39. The mean age of participants was 30 (*SD* = 5.79, range 19-39), 18 self-identified as female and 12 self-identified as male.

Stimuli

The word stimuli were the same as those used in Experiment 1 and 2. The math questions used in the Operation Span task are listed in Appendix D.

Design

The experiment followed a within-subject design with list length as a five-level within-subject factor (list lengths: two, three, four, five, and six). The dependent variables were proportion and number of correctly recalled words.

Procedure

Experiment 5 was a replication of Experiment 3 and followed the same procedures.

Results

The analyses of the operation span task were conducted twice: once including all participant data (N=30), and once considering only data from participants who scored 85% or above on the math question (N=18). Eliminating data from participants with lower math performance did not change the results, therefore, the analysis reported below includes all participants.[4]

Figure 5 shows the mean number and proportion of words recalled as a function of list length. A repeated-measures ANOVA revealed a significant main effect of list length on the

---

[4] ANOVA results when the analysis was limited to participants who scored 85% or above on the math questions showed a significant main effect of list length on the proportion of words recalled, $F(4, 68) = 4.23$, $p = .004$, $\eta_p^2 = 0.20$, $MSE = 0.007$, and a significant main effect of list length on the number of words recalled, $F(2.62, 44.44) = 212.15$, $p < .001$, $\eta_p^2 = 0.\ 93$, $MSE = 0.23$.

proportion of words recalled, *F*(4, 116) = 14.54, $p < .001$, $\eta_p^2 = 0.334$, *MSE* = 0.025. Post-hoc analysis showed that the mean proportion of words recalled for list length two and three were significantly higher than the mean proportion recalled for list lengths four, five and six. No other pairwise comparisons were significant ($\alpha = .05$). Planned contrasts indicated a significant linear effect, t (116) = -7.38, $p < .001$. The mean proportion of words recalled for each list length are shown in Table 5. The first component of the LLE was present: as list length increased, the proportion of correctly recalled words decreased.

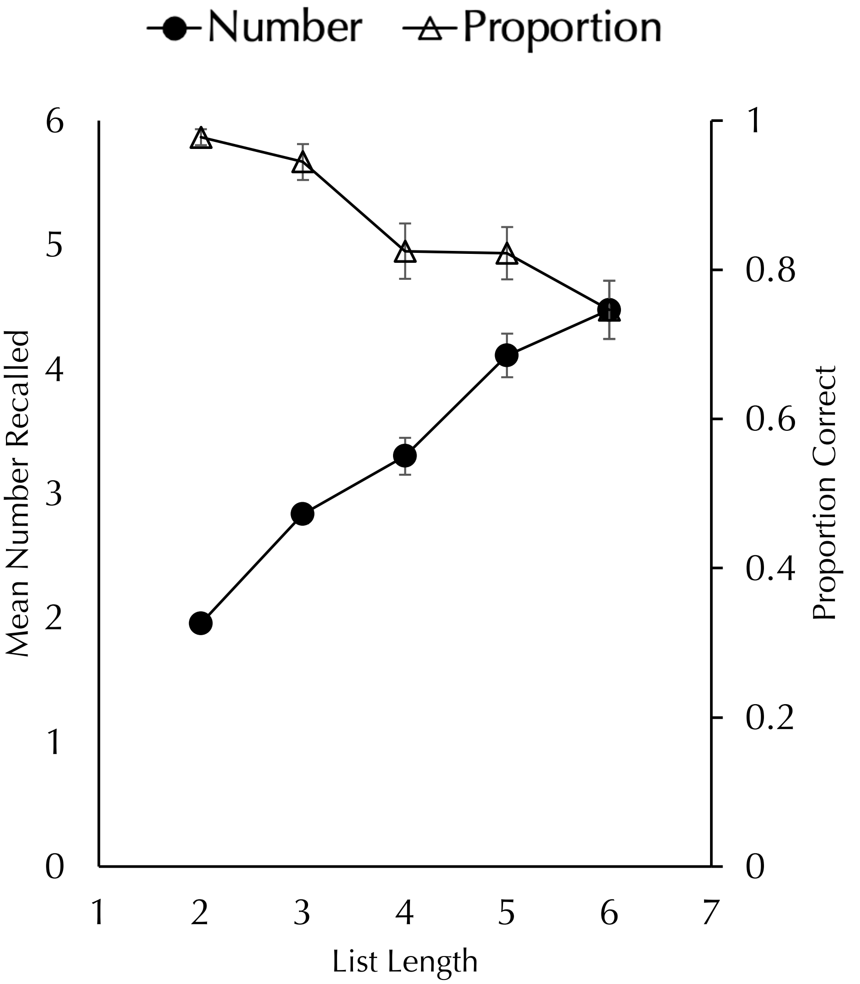

Figure 5. Mean proportion and mean number of correctly recalled words in Experiment 5 – OSPAN with unknown list lengths. Error bars represent the standard error of the mean.

The second ANOVA violated the assumption of sphericity, therefore, the Greenhouse-Geisser correct was applied. This analysis revealed a significant main effect of list length on the number of words recalled, $F(2.35, 68.28) = 68.42$, $p<.001$, $\eta_p^2=0.702$, $MSE = 0.757$. Post-hoc analysis revealed that all pairwise comparisons were significant (*all ps < .05*). The mean number of words recalled for each list length is shown in Table 5. Planned contrasts indicated a significant linear effect, $t\,(116) = 16.18$, $p < .001$ and a significant quadratic effect, $t(116) = 16.41$, $p < .001$. As list length increased, the mean number of correctly recalled words increased. Thus, both the first and second components of the LLE appear to be present.

Table 5

Mean and standard deviation scores for the number and proportion of correct responses in Experiment 5.

List length	Two	Three	Four	Five	Six
number correct	1.96	2.83	3.30	4.11	4.48
(SD)	(0.12)	(0.39)	(0.80)	(0.94)	(1.26)
proportion	0.98	0.94	0.83	0.82	0.75
correct (SD)	(0.058)	(0.13)	(0.20)	(0.19)	(0.21)

Experiment 6: OSPAN with an Extended Number of Unknown List Lengths

Despite using a different stimulus-set, the pattern of results obtained in Experiment 5 was consistent with that observed in Experiment 3. As such, in the current experiment, words from both stimulus-sets were used to examine memory performance in operation span across a wider range of list lengths. In the previous operation span experiments, it was unclear to what extent the number of words recalled would continue to increase as a function of list length. The list length in these experiments ranged from two to six because those are the typical lengths used in operation span (e.g., Conway & Engle, 1996). Therefore, Experiment 6 examined memory performance in lists ranging from three to eight words in length.

Method

Participants

Thirty different volunteers from ProlificAC participated in the second experiment. Each was paid £8 per hour, pro-rated. The following inclusion criteria were used for this and all subsequent experiments: (1) Native speaker of English; (2) approval rating of at least 90% on prior submissions at ProlificAC; and (3) age between 19 and 39. The mean age of participants was 26.6 (*SD* = 4.95, range 19-39), 22 self-identified as female and 8 self-identified as male.

Stimuli

The word stimuli were the same as those used in Experiment 1. The math questions are listed in Appendix E.

Design

The experiment followed a within-subject design with list length as a six-level within-subject factor (list lengths: three, four, five, six, seven, and eight). The dependent variables were proportion and number of correctly recalled words.

Procedure

Experiment 6 followed the same procedure outlined in Experiment 3 and 5, with the exception that participants were tested on their recall of 18 lists: three each of list three to eight words in length.

Results

The analyses of the operation span task were conducted twice: once including all participant data (N=30), and once considering only data from participants who scored 85% or above on the math question (N=27). Eliminating data from participants with lower math performance did not change the results, therefore, the analysis reported below includes all participants.[5]

Figure 6 shows the mean number and proportion of words recalled as a function of list length. Two separate analyses were conducted to verify the presence of the two components of the LLE. In both ANOVAs, the Greenhouse-Geisser correction was applied as the assumption of sphericity was violated. The first ANOVA revealed a significant main effect of list length on the proportion of words recalled, $F(3.71, 107.48) = 52.10$, $p < .001$, $\eta_p^2 = 0.642$, $MSE = 0.022$. Post-hoc analysis showed that the mean proportion of words recalled for list length three and four were significantly higher than the mean proportion recalled for list lengths five to eight. The proportion of words recalled from list length five and six were also significantly greater than the proportion of words recalled from list lengths seven and eight and a greater proportion of words were recalled from list length seven than from list length eight. No other pairwise comparisons were significant ($\alpha = .05$). Planned contrasts indicated a significant linear effect, $t\,(145) = -15.92$ $p < .001$. The mean proportion of words recalled for each list length are shown in Table 6. The first component of the LLE was present: as list length increased, the proportion of correctly recalled words decreased.

---

[5] ANOVA results when the analysis was limited to participants who scored 85% or above on the math questions showed a significant main effect of list length on the proportion of words recalled, $F(3.75, 97.40) = 42.12$, $p < .001$, $\eta_p^2 = 0.62$, $MSE = 0.02$ and a significant main effect of list length on the number of words recalled, $F(86.65, 51.08) = 15.33$, $p < .001$, $\eta_p^2 = 0.\ 37$, $MSE = 1.03$

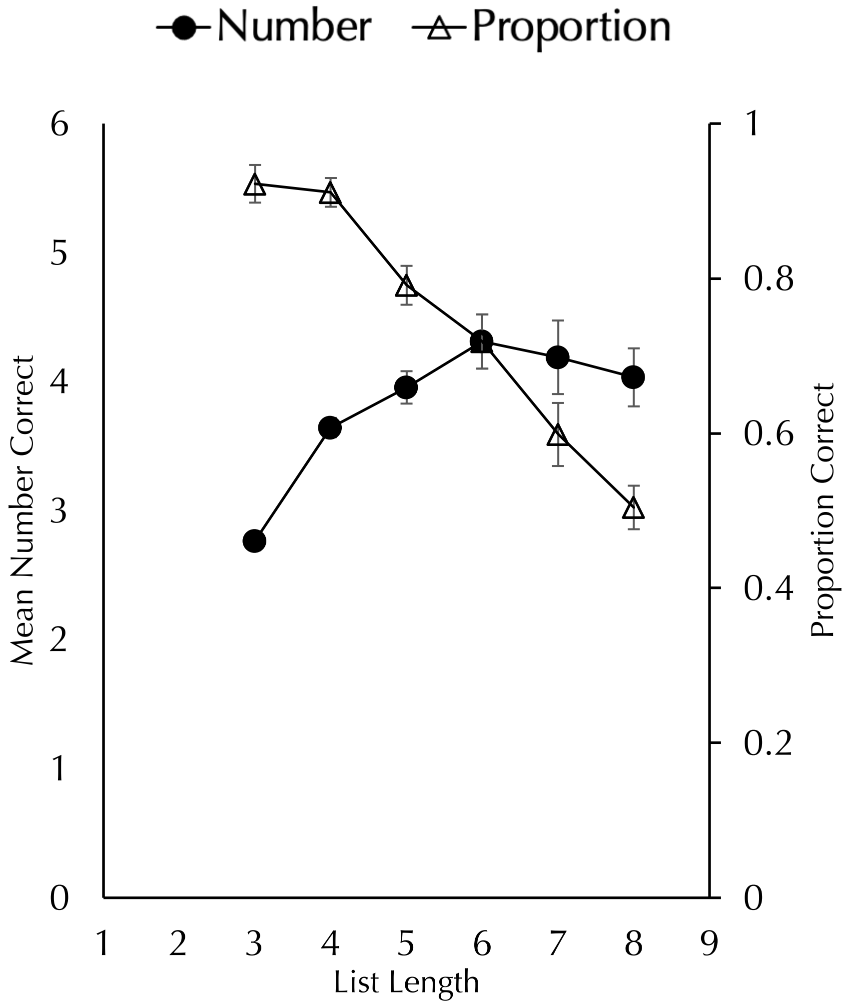

Figure 6. Mean proportion and mean number of correctly recalled words in Experiment 6 – OSPAN with unknown list lengths. Error bars represent the standard error of the mean.

The second ANOVA revealed a significant main effect list length on the number of words recalled, *F*(3.09, 89.51) = 14.645, *p*<.001, η_p^2=0.336, *MSE* = 1.050. Post-hoc analysis revealed that significantly fewer words were recalled from list length three than from all other list lengths and significantly fewer words were recalled from list length four than from list length six. No other pairwise comparisons were significant (α = .05). The mean number of words recalled for each list length is shown in Table 6. Planned contrasts indicated a significant linear effect, *t* (145) = 6.77, *p* < .001 and a significant quadratic effect, t(145) = -5.123, *p* < .001. The mean number of correctly recalled words initially increased as list length increased from three to six, before decreasing from list length six to eight, so although the first component of the list length effect was present, the second component was not.

Table 6

Mean and standard deviation scores for the number and proportion of correct responses in Experiment 6.

List length	Three	Four	Five	Six	Seven	Eight
number	2.67	3.64	3.96	4.31	4.19	4.03
correct (SD)	(0.39)	(0.40)	(0.68)	(1.13)	(1.54)	(1.21)
proportion	0.92	0.91	0.79	0.72	0.60	0.50
correct (SD)	(0.13)	(0.10)	(0.14)	(0.19)	(0.22)	(0.15)

General Discussion

The current research sought to determine whether the LLE occurs within two common short-term memory tasks: immediate serial recall and operation span. The set-size-accuracy benchmark proposed by Oberauer et al. (2018) holds that accuracy in short-term memory declines as set-size increases. Conversely, in the LLE, accuracy–when measured by the absolute number of items recalled–improves as a function of increasing set size. Therefore, evidence of a LLE in short-term memory would oppose this benchmark. Two previous studies reached opposing conclusions as to the presence of the LLE in short-term memory tasks. While Beaman (2006) found a LLE in immediate serial recall, Unsworth and Engle (2006a) found no such effect in either immediate serial recall or complex span. The current research was undertaken to reconcile these conflicting results and evaluate the strength of the set-size-accuracy benchmark.

The results of the current research indicated that the LLE does not occur in verbal immediate serial recall. Across two experiments, the proportion of words recalled decreased as list length increased. However, the absolute number of words recalled did not increase with list length, as expected in the LLE. Similarly, the LLE was not found in the operation span task. Across three experiments, the proportion of words recalled decreased with increasing list length. However, although the absolute number of words recalled initially increased as a function of list length, this trend was not maintained as list length continued to increase. These results support Unsworth and Engle (2006) and contradict Beaman (2006). Moreover, these findings are consistent with the set-size accuracy benchmark: recall accuracy decreased as set-size increased. Several plausible interpretations for these findings will be explored below.

The small sample size used in Beaman (2006) may account for why the findings of that study differ from those of Unsworth and Engle (2006) and the current research. In Beaman (2006) data was collected from only 12 participants, while in the current research thirty participants took part in each experiment. The sample size for the current studies was determined by a power analysis. By contrast, Unsworth and Engle included 235 participants in their study and found results consistent with those of the current research. A second reason may be that Beaman used a slower presentation rate of one item every 2.5 s compared to one item every 1 s in the current work and in that of Unsworth and Engle. With a longer presentation time, there is more opportunity for elaborative rehearsal.

Decay

Theories of short-term memory variably attribute the system's limited capacity to restrictions in either the amount of information that can be represented or the duration for which information can be maintained. Models of time-based forgetting propose that items held in short-term memory are susceptible to decay unless maintained by rehearsal. Under these accounts, memory accuracy decreases as set-size increases because a smaller proportion of list items can be actively rehearsed to offset decay. This explanation may account for the current research's findings. As list length increased, the proportion of list words participants could actively maintain decreased, and memory accuracy declined.

However, such an account has difficulty explaining the pattern observed in the absolute number of items recalled. For instance, in the operation span task (Experiments 3-6), the number of words recalled initially increased from list lengths two to five before subsequently decreasing. Decay-based models posit an intrinsic limit in the duration that information can be held in short-term memory without rehearsal. Thus, the initial increase in the number of words recalled is not inconsistent with decay-based explanations if the short-term memory system is understood to be capable of actively maintaining a limit of five words. However, once the number of items in a list exceeds the capacity of the short-term memory system, a participant should be capable of rehearsing a relatively consistent number of items within this limited duration, irrespective of list length. In contrast, the current research's results showed that the number of words recalled decreased as list length continued to increase. This pattern presents a similar issue for fixed-slot accounts of short-term memory, which propose that a set number of items may be maintained in memory and that newer items necessarily displace older items. Under such accounts, the number of items held in short-term memory should not decrease in supra-span lists.

Moreover, decay-based accounts of forgetting assume that rehearsal becomes less effective at compensating for decay under conditions of high cognitive load (Oberauer & Lewandowsky, 2014). In the current research, operation span should have produced greater cognitive load relative to immediate serial recall. In the former task, participants were required to retain list items in memory while simultaneously completing arithmetic operations. In the latter task, participants were not required to perform any concurrent cognitive operations during the presentation of list item. Nevertheless, recall performance was better in the operation span task than in the immediate serial recall task. Neither a decay-based nor a fixed-capacity account of short-term memory can satisfactorily explain this pattern of results.

Interference

Most decay-based explanations of forgetting in short-term memory disregards the role of interference. However, there is significant evidence that both proactive and retroactive interference leads to forgetting in short-term memory. For example, the acoustic similarity effect (Conrad, 1964) finds that recall performance suffers when lists are composed of similar-sounding items. In this effect, recall of early list items is reduced primarily due to retroactive interference, and recall of later list items is reduced primarily due to proactive interference. These and other findings suggest that an interference-based explanation of forgetting in short-term memory should be considered. Under this lens, the current results may be interpretable as a trade-off between the number of words available to be recalled and the accumulation of interference as list length increased. Thus, the initial growth observed in the number of items recalled as list length increased may reflect that longer lists provided a greater number of possible words to recall. However, at a certain point, this benefit no longer compensated for the increased interference associated with long lists, and the number of words recalled reach a plateau or began to decrease.

An alternative source of interference in the current experiment may have been derived from output interference. Output interference (Tulving & Arbuckle, 1963) occurs when the act of retrieving a list item interferes with the memory representation of other similar items. For example, Kay and Poulton (1951) demonstrated that the first four items in an eight-item sequence were better remembered when these items were recalled before, rather than after, the last four items. In output interference, the retrieval of an item strengthens its representation in memory. As a result, the same item is more likely to be retrieved during attempts to access additional items (Roediger, 1974). In the current research, the deterioration in memory accuracy as list length increased may reflect a build-up of output interference. Again, the initial increase in the number of words recalled as list length increased can be attributed to the greater number of words available to be retrieved. The subsequent plateau or decline in the number of words recalled can be attributed to a build-up of output interference. Similarly, the decrease in the proportion of words recalled as list length increased may reflect output interference.

Serial Ordered Recall

In the current research, participants were required to retain both order information and item information. Several mechanisms have been suggested to explain how order-information is stored (Brown et al., 1999; Burgess & Hitch, 1999, 2006; Neath & Crowder, 1990; Raaijmakers & Shiffrin, 1980). Many of these accounts propose that list items are represented in association with their temporal position in a list. Thus, in serial ordered recall, each item is linked to a distinct temporal context. At recall, a context can act as a retrieval cue to access a specific list item. However, in long lists, these contextual cues become less effective at differentiating between items. The reinstatement of one contextual cue may activate multiple list items associated with a similar cue. These accounts of serial recall also propose that cue distinctiveness is determined by the amount of time that elapses between the learning and retrieval episodes. Thus, item discriminability in memory is determined both by an item's temporal proximity to other items and by its temporal distance from the point of recall.

Such accounts of serial ordered recall may be applied to understand the results of the current research. In both immediate serial recall and operation span, recall performance deteriorated as list length increased. This trend may reflect a loss of cue distinctiveness as list length increased, causing words to be retrieved in an incorrect serial order. This explanation has the advantage of being able to account for the improved performance in the operation span experiments relative to the immediate serial recall experiments. In the operation span task, more time elapsed between the presentation of each word. As a result, each word would have been associated with a more distinctive temporal cue. Therefore, in the operation span task, participants would have benefited from contextual cues that were more effective at differentiating between list words than the cues available to participants in the immediate serial recall task.

Distinctiveness

The dual-store account of immediate free recall attributes retrieval of pre-recency items to long-term memory. The LLE primarily influences pre-recency items and is thus generally considered to be a long-term memory phenomenon. However, several empirical findings have cast doubt over the validity of the dual-store account of immediate free recall. As such, the fact that the LLE primarily influences pre-recency items may not justify the claim that it is a long-term memory effect. An alternative explanation of the SPC in immediate free recall relies on distinctiveness. Under this account, primacy and recency items benefit from increased positional distinctiveness due to their proximity to the start and end of the list. Thus, the positional contexts associated with these items can act as more effective retrieval cues than the less distinctive contexts associated with mid-list items. According to the perturbation model (Estes, 1972; Lee & Este, 1977), list items are stored in association with positional information. However, with the passage of time, this information diffuses along the temporal dimension so that list items are associated with a less well specified time period. As a result, the positional information associated with list item becomes less distinctive and less effective as a retrieval cue. However, items at the start and end of a list are not as affected by this diffusion process because the spread along the positional dimension can only occur in one direction. The perturbation model, therefore, accounts for the SPC in immediate free recall. The process of perturbation affects mid-list items to a greater extent than early and late list items, resulting in better memory for primacy and recency items.

Both the number of list items and the temporal distance between an item and the point of recall are believed to influence positional distinctiveness. Therefore, recency items benefit from the temporal and positional distinctiveness associated with end-of-list proximity. The additional temporal advantage of recency items is not diminished by increasing list length. In contrast, the average time interval between the presentation and recall of pre-recency items increases with list length. As a result, temporal distinctiveness decreases in these items as a positive function of list length. The greater susceptibility of earlier list items to changes in list length may explain why the LLE in immediate free recall primarily influences the pre-recency portion of the SPC. This explanation is in keeping with the current research's findings, which suggests that the LLE occurs when the benefit accrued by increasing the number of items available to be recalled outweighs the increased interference and decreased cue distinctiveness associated with longer lists.

Despite the methodological similarities between immediate free recall and immediate serial recall, the current research suggests that the LLE occurs exclusively within immediate free recall. This may be because the accurate retention of order information is not required in immediate free recall, and thus the loss of temporal cue distinctiveness has a smaller effect in this task relative to immediate serial recall. If the LLE occurs when the advantage of having more items available to be recalled compensates for the decrease in temporal cue distinctiveness, then a LLE is more likely to occur in tasks that do not require order information to be accurately recalled.

Simulations

Simulation predictions from three models, the Start-End Model, the Feature model, and SIMPLE, were generated to examine the anticipated relationship between the number and proportion of recalled items across list lengths. These models were selected because of the availability of their source code.

The Start-End Model. In SEM, parameters S_o and S represent the strength of the start marker and the degree to which that strength changes across list positions. These parameters are both assigned fixed values. The equivalent end marker parameters (E_o and E) are then defined in relation to these start maker values. As a result, the model has two free parameters, one representing a ratio of the maximum strength of the end marker to that of the start maker (F_0), and the other representing the degree of change in end marker strength relative to the degree of change in start marker strength. SEM also includes an additional free parameter representing output threshold (T_o). T_o is included in the model to account for omission errors, which occur when competition among list items results in no item being activated above a given retrieval threshold. The model also assumes that recently recalled items are temporarily suppressed to prevent their being reselected during attempts to retrieve additional items. To incorporate response suppression, a final free parameter, R_s, is included in the model. R_s represents the rate of exponential decay of response suppression.

The Start-End Model simulations in the current study were conducted using JavaScript code written by Dr. Ian Neath and available at https://memory.psych.ca/models/startend/js/ (Henson, 1998). All simulated predictions reported are based on 1000 iterations. Predictions simulated using SEM for the number and proportion of items recalled as a function of list length are shown in Figure 7. When recall performance is simulated using SEM and assuming no output threshold ($T_o = 0$) and permanent response suppression ($R_s = 0$), the model predicts that the proportion of recalled items will decrease as list length increases (Panel A). Still, the number of items recalled will continue to increase with list length. The model predicts the same pattern when output threshold is not accounted for, and response suppression is assumed to decay (T_o = 0, $R_s = 0.5$; Panel D). In contrast, when output threshold is accounted for, and response suppression is assumed to decay ($T_o = 0.35$, $R_s = 0.5$), the model predicts an inverted U-shape in the number of words recalled as a function of list length (Panel B). The same pattern is predicted when output threshold is accounted for, and response suppression is assumed to be permanent ($T_o = 0.35$, $R_s = 0$; Panel C). Based on these simulations, when output threshold is accounted for, SEM predicts an inverted U-shaped in the number of words recalled as a function of increasing list length.

The inverted U-shaped pattern predicted by these latter SEM simulations is consistent with the pattern of results observed in the current research. If successful retrieval is understood to occur when a target item is reactivated to a greater extent than competing items, then it is logical to assume that omission errors occur when this reactivation is weak. In SEM, the output threshold parameter functions to simulate this outcome. The model suggests that more items fail to reach the activation threshold necessary for retrieval in longer lists. This pattern is predicted because of the decreased positional distinctiveness associated with items in longer lists. As list length is initially increased, simulated memory performance benefits from the increased number of items available to be retrieved. However, as list length continues to increase, this benefit no longer compensates for the cost associated with the decrease in positional distinctiveness. In the context of the current research, words from longer lists would have been associated with less distinctive positional contexts as list length increased. As a result, the reinstatement of contextual cues would have been less effective at reactivating the associated list words. Consequently, a greater number of words would have been insufficiently reactivated to reach the output threshold.

The characterization of working memory as a limited capacity system underlies the Oberauer et al. (2018) set-size-accuracy benchmark. The tenability of a limited capacity memory system depends on the existence of forgetting mechanisms such as decay or displacement. However, the Start-End Model offers an explanation of the observed decrease in the number of items recalled as a function of set-size that does not rely on the notion of a capacity limitation. Instead the model suggests that memory performance decreases as list length increases due to interference among list items and the inadequate availability of retrieval cues at test. Therefore, although the empirical data collected in the current study are congruent with the set-size-accuracy benchmark and with the results of Unsworth and Engle (2006) as opposed to Beaman

(2006), concluding that these results reflect distinct characteristics of short-term or working memory would be premature.

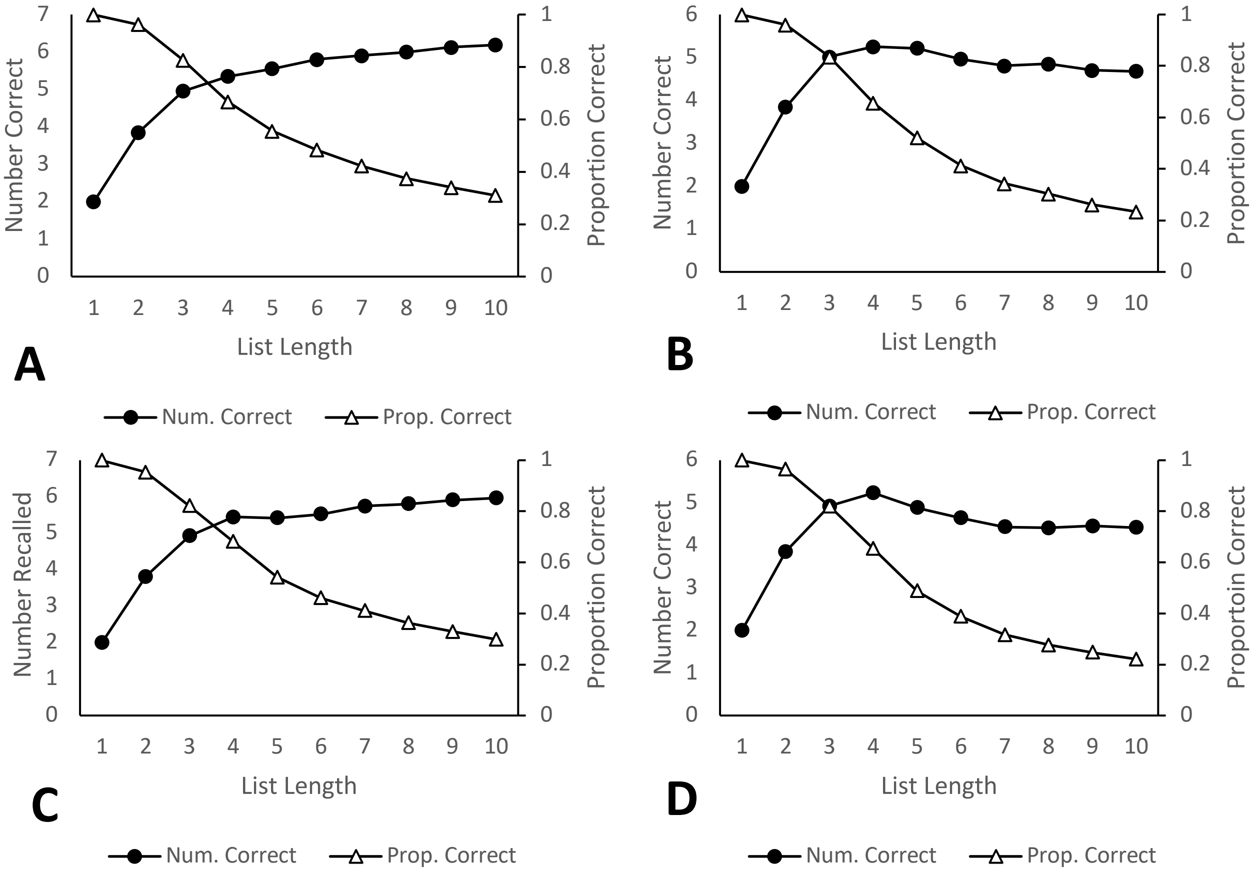

Figure 7. Predictions of the Start-End Model (SEM) for proportion and number of correctly recalled items as a function of list length. Simulated predictions were based on 1000 iterations. Panel A represents simulation predictions assuming no output threshold ($T_o = 0$) and permanent response suppression ($R_s = 0$). Panel B represents simulation predictions accounting for output threshold ($T_o = 0.35$) and response threshold decay ($R_s = 0.5$). Panel C represents simulations predictions accounting for output threshold ($T_o = 0.35$), but not response threshold decay ($R_s = 0$). Panel D represents simulations predictions accounting for response threshold decay ($R_s = 0.5$) but not output threshold ($T_o = 0$).

The Feature Model. The Feature model was designed to account for effects observed in immediate memory (Nairne, 1988; 1990). The model assumes that items are encoded as traces composed of modality-dependent and modality-independent features. Information associated with the presentation context of an item is considered modality-dependent, while information generated through the internal categorization and processing of an item is considered modality-independent. These multi-attributional memory traces are assumed to be simultaneously stored in primary and secondary memory. Primary memory traces are then subject to degradation through interference, while secondary memory traces are assumed to be permanent in nature. Recall is based on a matching process in which a degraded primary memory trace is compared to traces sampled from a secondary memory search set. The probability of successful retrieval is determined by the amount of feature overlap between the residual primary memory trace and the intact secondary memory trace. Trace degradation may reduce the similarity advantage that a primary memory trace has to its corresponding secondary memory trace over other primary memory representations. When this is the case, overall recall performance is reduced.

In the Feature model, memory traces are represented as vectors of features. Each feature is randomly assigned a value of either +1 or -1. With visually presented items, the number of modality-independent features is generally set to 20, and the number of modality-dependent features is set to two. An overwriting process represents interference. Individual features within a vector are overwritten, with probability F, if the corresponding feature matches a feature in the following item. While subsequently presented list items can only overwrite modality-dependent features, modality-independent features can additionally be overwritten by internally generated traces, produced as by-products of cognitive activity. As such, modality-dependent features are important when distinguishing between externally presented and internally generated representations.

During retrieval, items are assumed to be sampled from secondary memory based on their similarity to a degraded primary memory trace. Similarity is derived from a distance value, d, calculated from the number of mismatching features, M, between the primary and secondary memory traces. Equation 1 is used to calculate distance, where N represents the number of features compared, a is a scaling parameter and, M_k represents the number of times that feature position x_{ik} does not match feature position x_{jk}.

$$d_{ij} = \frac{a \sum M_k}{N} \qquad (1)$$

Similarity between a primary and secondary memory trace, $s(i,j)$, is then calculated using Equation 2.

$$s(i,j) = e^{-d_{ij}} \qquad (2)$$

The probability of sampling a secondary memory trace, SM_j, given a particular primary memory trace, PM_i, is calculated using Equation 3.

$$P_s(SM_j | PM_i) = \frac{s(i,j)}{\sum_{k=1}^{n} s(i,k)} \qquad (3)$$

Items sampled from secondary memory must then be recovered before being output. The probability of successfully recovering an item from secondary memory, P_r, is calculated using

Equation 4, where *r* represents the number of times the item has been sampled, and *c* is a scale constant.

$$P_r = e^{-cr} \tag{4}$$

In serial ordered recall, to output the correct response, both item information and order information are required. According to the Feature model, order information is stored along with primary memory traces. The model assumes that order information is encoded as a point along a serial position dimension for each item. In serial ordered recall, primary memory traces are first sampled to determine which trace was in position one. Once identified, this primary memory trace is used to sample from among a secondary memory search set. However, for each primary memory trace, the point representing order information may, over time, drift along the serial position dimension. This drifting occurs according to the perturbation process. The probability of perturbation during a given time interval is given by *θ*, which is held constant at 0.5. It is assumed that perturbation is equally likely in either direction but cannot extend beyond the positional limits imposed by the task. The one free parameter associated with the perturbation process is π, representing the number of opportunities to perturb. This value is assumed to be the same for all items. The probability that item, *I*, will occupy a given position, *p*, during the next time interval, *t+1,* is given by Equations 5, 6 and 7. Equation 5 is used for items in non-terminal positions, while Equations 6 is used for the item in the first position, and Equation 7 is used for the item in the last position, *n.*

$$\mathrm{I}_{p,t+1} = (1-\theta)\mathrm{I}_{p,t} + \left(\frac{\theta}{2}\right)\mathrm{I}_{p-1,t} + \left(\frac{\theta}{2}\right)\mathrm{I}_{p-1,t} \tag{5}$$

$$\mathrm{I}_{1,t+1} = \left(1 - \frac{\theta}{2}\right)\mathrm{I}_{1,t} + \left(\frac{\theta}{2}\right)\mathrm{I}_{2,t} \tag{6}$$

$$\mathrm{I}_{n,t+1} = \left(1 - \frac{\theta}{2}\right)\mathrm{I}_{n,t} + \left(\frac{\theta}{2}\right)\mathrm{I}_{n-1,t} \tag{7}$$

Feature model simulations in the current study were conducted using JavaScript code written by Dr. Ian Neath and available at https://memory.psych.ca/models/feature/js/ (Nairne, 1990; Neath & Nairne, 1995; Neath, 2000). All simulated predictions reported are based on 20000 iterations. The predictions generated for the number and proportion of items recalled as a function of list length are shown in Figure 8. A replication of Beaman's (2006) simulation confirmed his results. Using the standard set of parameters outlined in Neath (2000), the Feature model predicts that the proportion of items recalled will decrease, and the number of items recalled will increase as a function of increasing list length (Panel A). This same qualitative pattern is predicted with and without accounting for perturbation (Panel B). Similarly, changing the scaling parameter, a, does not change the general trend predicted by the model (Panels C and D). In contrast to these predictions, the empirical findings of the current research showed an inverted U-shape in the number of items recalled as a function of increasing list length. The inability of the Feature model to predict these results suggests a problem with the model.

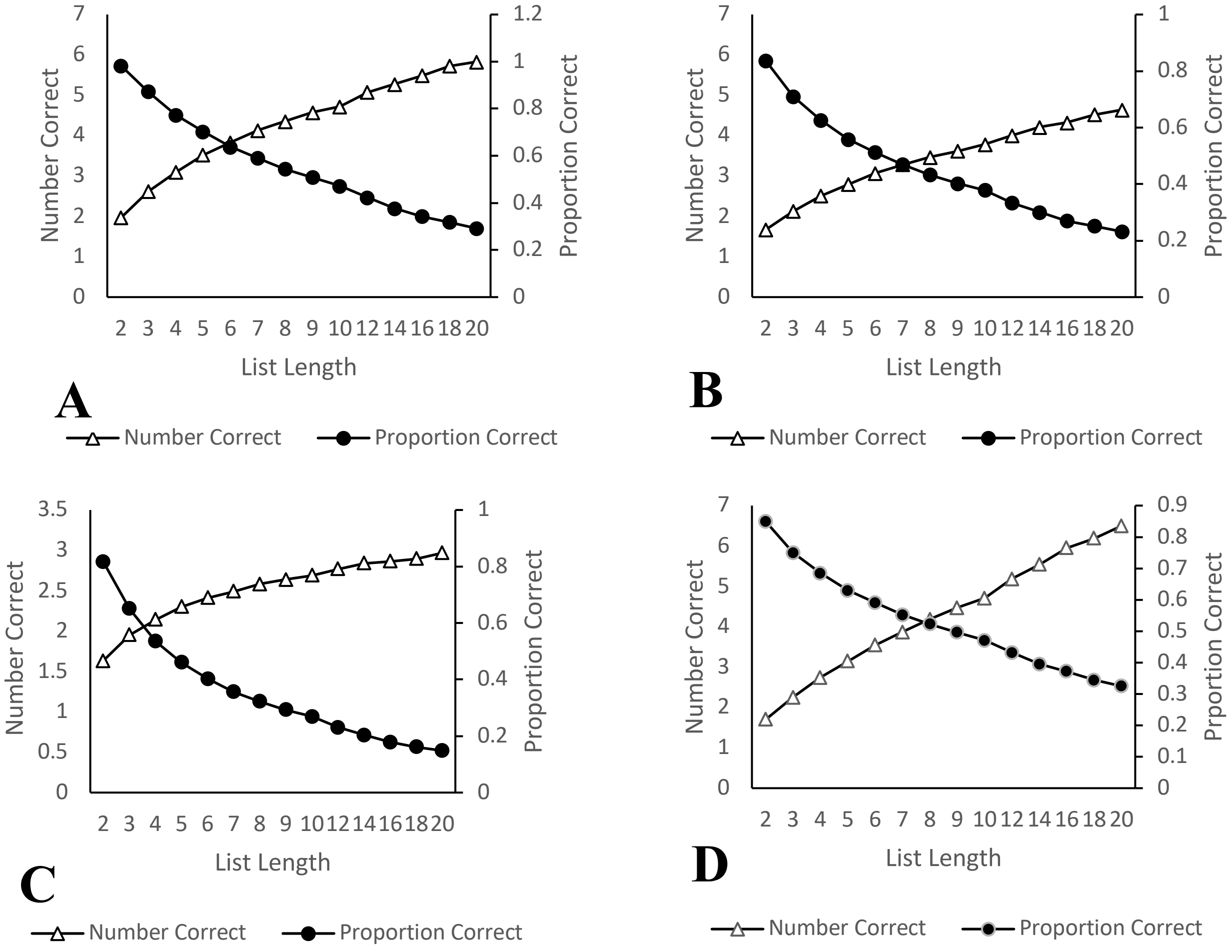

Figure 8. Predictions of the Feature model for proportion and number of correctly recalled items as a function of list length. All simulated predictions reported are based on 20000 iterations. Panel A represents simulation predictions using the standard parameters from Neath (2000). Panel B represents simulations predictions using the standard parameters from Neath (2000) and accounting for perturbation (θ= 0.05, *π* = 5). Panel C represents simulations predictions with a reduced distance scaling value (*a* = 7) and accounting for perturbation (θ = 0.05, *π* = 5). Panel D represents simulations predictions with an increased scaling value (*a* = 13) and accounting for perturbation (θ = 0.05, *π* = 5).

SIMPLE. SIMPLE (Brown, Neath & Chater, 2002) is a local relative distinctiveness model premised on the assumption that items are represented in memory along one or more dimensions. According to the model, items more isolated in psychological space are more distinctive and memorable than items with many close neighbours. The number and type of dimensions on which items are represented vary with the demands of a given memory task. In the context of a standard serial recall task, the model assumes that items are represented in memory along a temporal or ordinal dimension. At test, location along this dimension is used as a retrieval cue to access the associated item representation in memory. Under SIMPLE, the probability of successful retrieval is determined by the extent to which a retrieval cue is more effective at cueing the target memory than other competing memories. Thus, the probability of retrieving a list item is considered to be the inverse of the summed confusability of an item with all other item representations in memory. Confusability is assumed to be determined by the ratio of the items' respective positional distances from the point of retrieval.

Under the SIMPLE model, similarity, $\eta_{i,j}$, between two log-transformed memory representations, M_i and M_j, is given by Equation 8. In this equation, c is a constant and α is set to 1.0 for an exponential function and 2.0 for a Gaussian function relating similarity to distance. The parameter c is the main free parameter of the model and represents the rate at which similarity decreases with psychological distance. As c increases, the confusability between two items decreases more rapidly with psychological distance. The probability of retrieving item i, R_i, given a particular positional cue for item j, C_j, can be understood as a ratio of the similarity between the value at position i in psychological space and position j, divided by the sum of ratios between the value at position i and the value at all other positions. This relationship is given by Equation 9, in which n represents the number of items in the memory set. Equation 10 is used to account for omissions by transforming the estimated recall probability, *P*, resulting in the output probability, P_o. This transformation adds two additional free parameters, *t* and *s,* to the model. Parameter *t* is the output threshold, and parameter *s* is the slope of the transforming function. As *s* decreases, the change in the probability of retrieval as a function of psychological distance becomes more gradual.

$$\eta_{i,j} = e^{-c|M_i - M_j|^{\alpha}} \tag{8}$$

$$P(R_i|C_j) = \frac{\eta_{i,j}}{\sum_{k=1}^{n} \eta_{j,k}} \tag{9}$$

$$P_o = \frac{1}{1 + e^{-s(P-t)}} \tag{10}$$

Two different versions of SIMPLE have been fit to immediate serial recall data. One assumes a single underlying dimension of ordinal position and the other assumes a single underlying dimension of time. Past work has shown each version can fit such data very well (e.g., Surprenant, Neath, & Brown, 2006). Both versions were assessed using JavaScript packages written by Dr. Ian Neath.

To simulate the results of a serial recall experiment with the positional version of SIMPLE, a single underlying dimension, ordinal position, was assumed. For each list length, the first item was assigned a value of 1, the second item a value of 2, and so on. Using this approach, SIMPLE was fit to the data collected in Experiment 2 of the current research. Figure 9 shows the predictions of the model when the free parameters were selected to minimize the difference between the data and the model. SIMPLE gave a very good fit to the data, with R^2 values ranging from 0.999 for list length three, to 0.972 for list length eight. In list lengths three to five, the model slightly underpredicts both the number and proportion of items recalled, while in list lengths six to eight, these values are slightly overpredicted. Nevertheless, the overall pattern predicted by the model is consistent with the results of the current research. SIMPLE, therefore, offers an explanation of the results observed in the current research.

Similarly, R^2 values were very high when the model parameters were selected to fit the data for each list length individually (R^2 values ranged from 0.999 for list length three, to 0.988 for list length eight). Again, the ability of the model to closely predict the empirical data collected in Experiment 2 suggests that SIMPLE provides a candidate explanation of the results of the current research.

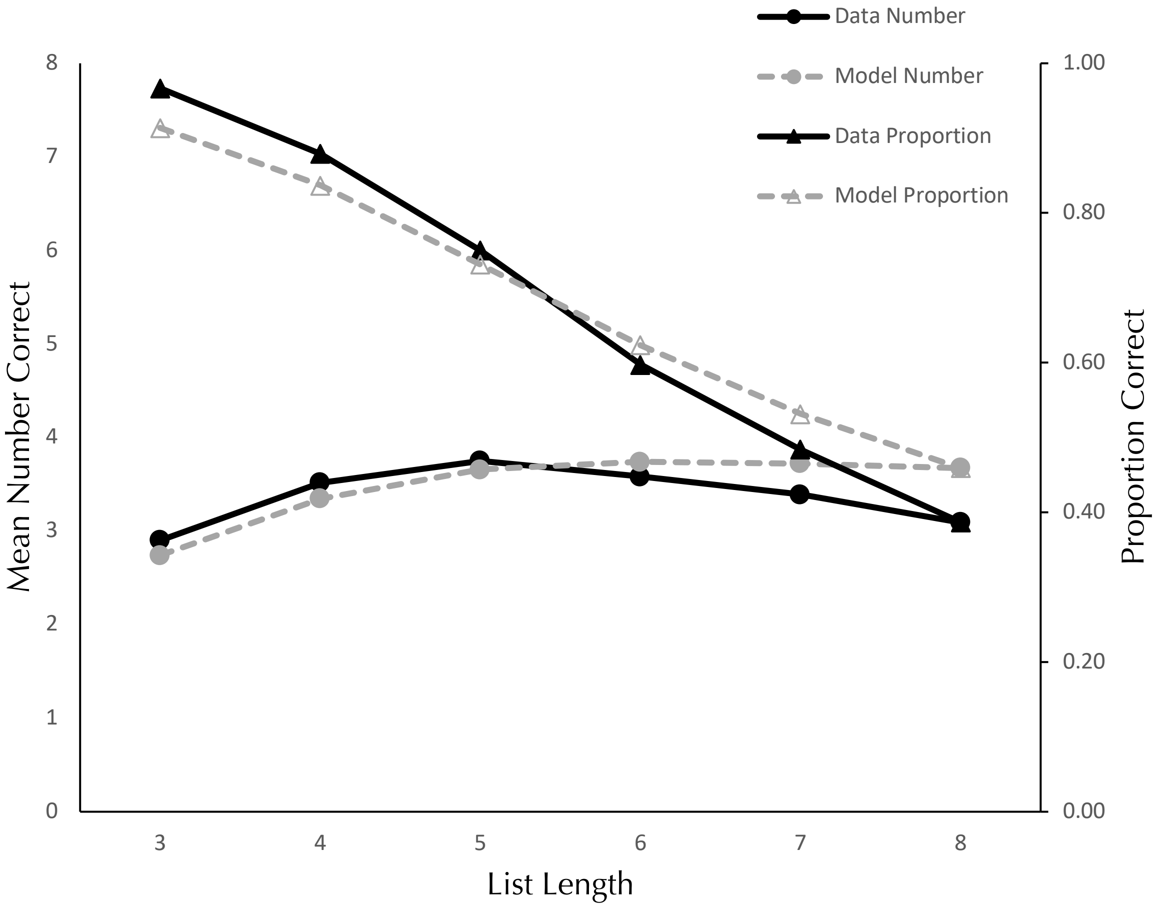

Figure 9. Data from Experiment 2 compared with the predictions of the positional version of SIMPLE with parameters c= 5.10, *s*=8.606, *t*=0.625.

A second SIMPLE model simulation was performed in which the underlying dimension was assumed to be temporal position. The temporal version of SIMPLE was fit to the data collected in Experiment 2 of the current research. Figure 10 shows the predictions of the model when the free parameters were selected to minimize the difference between the data and the model. The temporal version of SIMPLE also gave a very good fit to the data, with R^2 values ranging from 0.999 for list length three, four and five, to 0.973 for list length six. At list length three, the model slightly underpredicts both the number and proportion of items recalled, while in list lengths five to eight, the model slightly overpredicts these values. Nevertheless, the overall pattern predicted by the model is consistent with the results of the current research. The temporal version of SIMPLE, therefore, also offers a potential explanation of the results observed in the current research. This version of the model also showed a very high fit with the data when parameters were selected to fit the data for each list length individually (R^2 values ranged from 0.999 for list length three, to 0.974 for list length eight). The outcomes of these simulations therefore suggest that the temporal version of SIMPLE may account for the results of the current research.

As with the Start-End Model, both the positional and temporal versions of SIMPLE offer an explanation of the data collected in the current study. By supporting the findings of Unsworth and Engle (2006) and contradicting those of Beaman (2006), the current results align with Oberauer et al. (2018)'s set-size-accuracy benchmark. However, the capacity of SIMPLE to accurately predict the pattern observed in the empirical data suggests that assuming the results reflect the operations of a limited capacity short-term or working memory system is not necessarily justified. Instead, the data are well accounted for by SIMPLE – a timescale invariant model.

The Set-Size-Accuracy Benchmark

Oberauer et al. (2018) proposed that the decline in short-term memory accuracy as a positive function of list length provides evidence for the capacity limitations that are central to the conceptualization of short-term memory. For this reason, the set-size-accuracy benchmark was given a high priority rating. Although the current research does not explicitly contradict this benchmark, some of the findings are problematic for limit capacity accounts of short-term memory. For instance, such accounts have difficulty explaining why participants showed better

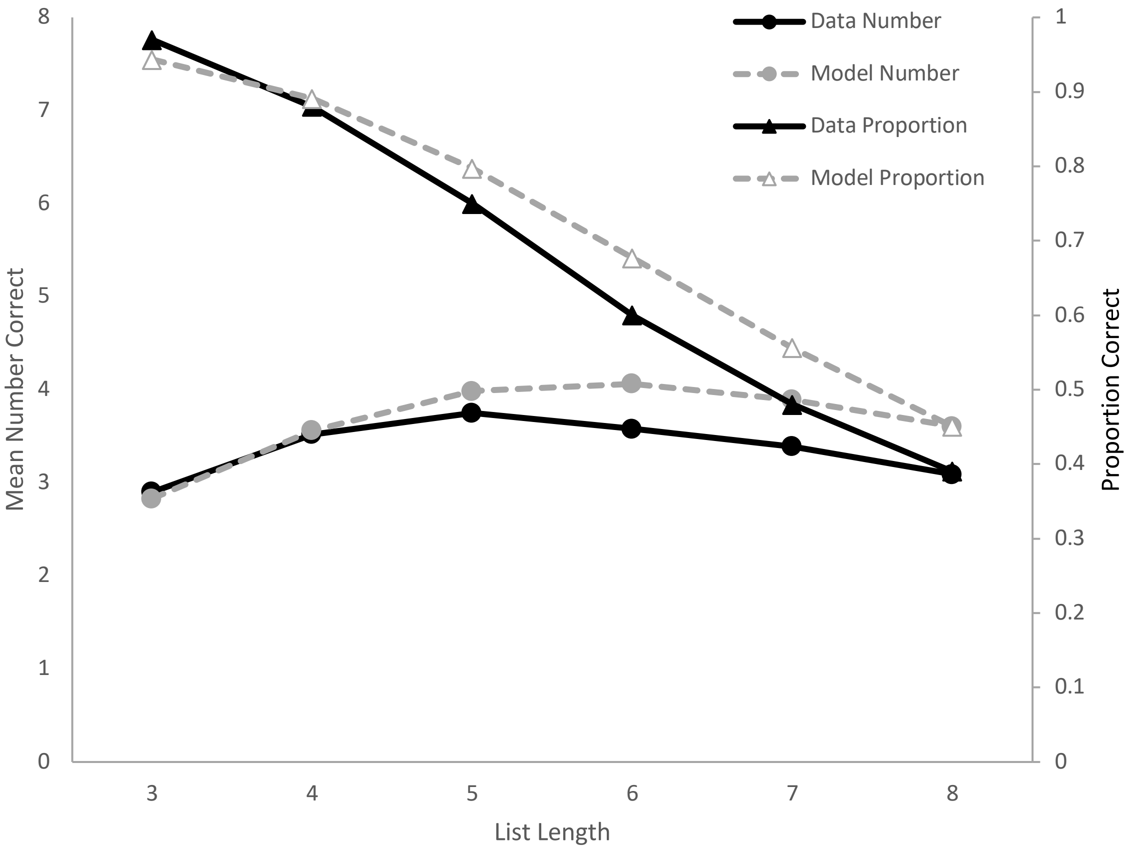

Figure 10. Data from Experiment 2 compared with the predictions of the positional version of SIMPLE with parameters c= 11.647, *s*=8.122, *t*=0.584.

recall in the operation span relative to the immediate serial recall task. The operation span task should have occupied more cognitive resources and resulted in lower recall performance. In contrast, a distinctiveness-based explanation can account for these results. This explanation opposes the notion that the set-size-accuracy relationship is evidence of a limited capacity short-term memory. Instead, it offers an account of the relationship between set size and accuracy without presupposing distinct short- and long-term memory systems. Thus, although any model or theory of memory should be capable of explaining the tendency of performance to decrease as set-size increases, such explanations need not depend on a limited capacity short-term memory system.

Known and Unknown List Lengths

One additional aim of the current research was to determine how the knowledge of a list's length, before its presentation, affects recall performance. In Beaman (2006), participants were told the number of items in each list before its presentation, while participants in Unsworth and Engle (2006) were not given this information. Therefore, this methodological difference might have been the cause of the opposing results found in these two studies. The current research's result found that prior knowledge of list length improved overall performance but did not change the general pattern of results. The relatively better performance in the known list length condition is consistent with Pollack, Johnson and Knaff (1959) and likely reflects the ability to employ more effective mnemonic strategies during the encoding of lists with known relative to unknown lengths (Bunting et al., 2006; Grenfell-Essam & Ward, 2012).

Limitations and Future Directions

The current research was designed to assess whether the LLE occurs in tasks targeting short-term memory. Although I found no evidence of the LLE in immediate serial recall or operation span, it would be valuable to extend this examination to include other short-term memory tasks. Similarly, the current research was limited to verbal stimuli, which were presented visually. Thus, there is no evidence that the result would generalize to other stimuli or presentation modalities. Again, it would be valuable to extend this investigation to determine whether the LLE may occur in short-term memory tasks under different conditions.

A distinctiveness-based account of serial ordered recall was suggested as a candidate explanation for the current results. However, this study was not explicitly designed to assess the impact of distinctiveness on recall performance. Future research might vary the duration of the inter-presentation intervals to determine how increasing or decreasing temporal distinctiveness impacts recall performance. If increasing the distinctiveness of the temporal contexts associated with list items results in a LLE, the proposed distinctiveness explanation would be supported. Alternatively, varying the structure of lists by presenting items interpolated with a pause at differing intervals would allow the impact of positional distinctiveness to be investigated. If the addition of a positional marker within the list increases recall accuracy or results in a LLE, this would also provide evidence favouring the distinctiveness explanation.

Lastly, it would be beneficial to contrast recall accuracy in immediate serial recall and immediate free recall directly. Different list lengths are typically used to investigate performance in these two tasks, making direct comparison difficult. However, Ward et al. (2010) noted strong similarities between performance in immediate serial recall and immediate free recall when participants were tested using lists ranging from one to 15 words. A similar comparison in which both the number and proportion of recalled words are considered might shed more light on the importance of requiring serial ordered recall. Bhatarah, Ward and Tan (2008) used a pre-cue/post-cue technique to demonstrate that differences between the two tasks emerge during retrieval and were not due to differential encoding processes. Thus, the impact of positional and temporal cue distinctiveness may be inferred from a direct comparison of recall accuracy across multiple list lengths in immediate serial recall and immediate free recall.

Final Conclusions

As hypothesized, the current research's findings indicate that the LLE does not occur in immediate serial recall or operation span. Interestingly, performance was comparatively better in the operation span task than in the immediate serial recall task. One aim of the current research was to reconcile the differing results of two previous studies. As anticipated, the present findings were consistent with the results of Unsworth and Engle (2006) and opposed the results of Beaman (2006). A second aim of the current research was to assess the validity of the set-size-accuracy benchmark proposed by Oberauer et al. (2018). The results aligned with this benchmark in showing that recall accuracy declines as set size increases. However, an explanation of these findings was proposed, which conflicts with the idea that the set-size-accuracy relationship is evidence of a limited capacity short-term memory system. Instead, it was suggested that recall accuracy declines as a function of increasing list length due to increased interference and a decrease in the positional and temporal distinctiveness in longer lists. Both the Start-End Model and SIMPLE were able to predict the patterns of results observed in the current research. By contrast, the Feature model could not predict the observed pattern, suggesting a problem with the model. Lastly, a final aim of the current research was to determine what effect prior knowledge of list length has on recall performance. The results showed that informing participants of the list length of an upcoming trial improved overall recall performance but did not affect the general pattern of results.

References

Adam, K. C. S., Mance, I., Fukuda, K., & Vogel, E. K. (2015). The contribution of attentional lapses to individual differences in visual working memory capacity. *Journal of Cognitive Neuroscience*, *27*(8), 1601–1616. https://doi.org/10.1162/jocn_a_00811

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *The psychology of learning and motivation: II* (pp. xi, 249–xi, 249). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, *18*(4), 362–365. https://doi.org/10.1080/14640746608400055

Baddeley, A. D. (1992). Working memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Baddeley, Alan D., & Hitch, G. (1974). Working Memory. In *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Elsevier. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology*, *8*(4), 485–493. https://doi.org/10.1037/0894-4105.8.4.485

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2019). From short-term store to multicomponent working memory: The role of the modal model. *Memory & Cognition*, *47*(4), 575–588. https://doi.org/10.3758/s13421-018-0878-5

Baddeley, A. D., & Logie, R. H. (1999). Working memory: the multiple-component model. In *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.005

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 575–589. https://doi.org/10.1016/S0022-5371(75)80045-4

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83–100. https://doi.org/10.1037/0096-3445.133.1.83

Beaman, C. P. (2006). The relationship between absolute and proportion scores of serial order memory: Simulation predictions and empirical data. *Psychonomic Bulletin & Review*, *13*(1), 92–98. https://doi.org/10.3758/BF03193818

Bhatarah, P., Ward, G., Smith, J., & Hayes, L. (2009). Examining the relationship between free recall and immediate serial recall: similar patterns of rehearsal and similar effects of word length, presentation rate, and articulatory suppression. *Memory & Cognition*, *37*(5), 689–713. https://doi.org/10.3758/MC.37.5.689

Bhatarah, P., Ward, G., & Tan, L. (2006). Examining the relationship between free recall and immediate serial recall: The effect of concurrent task performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 215–229. https://doi.org/10.1037/0278-7393.32.2.215

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. (2013). *Perception and Communication*. Elsevier.

Brown, G. D. A., Neath, I., & Chater, N. (2002). A ratio model of scale invariant memory and identification. *Memory Lab Technical Report, Purdue University*.

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

Brown, G. D. A., Vousden, J. I., McCormack, T., & Hulme, C. (1999). The development of memory for serial order: a temporal-contextual distinctiveness model. *International Journal of Psychology*, *34*(5/6), 389–402. https://doi.org/10.1080/002075999399747

Brown, J. (1958). Some Tests of the Decay Theory of Immediate Memory. *Quarterly Journal of Experimental Psychology*, *10*(1), 12–21. https://doi.org/10.1080/17470215808416249

Bunting, M., Cowan, N., & Scott Saults, J. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology*, *59*(10), 1691–1700. https://doi.org/10.1080/17470210600848402

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*(3), 551–581. https://doi.org/10.1037/0033-295X.106.3.551

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, *55*(4), 627–652. https://doi.org/10.1016/j.jml.2006.08.005

Conrad, R. (1957). Decay theory of immediate memory. *Nature*, *179*(4564), 831–832. https://doi.org/10.1038/179831b0

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, *55*(1), 75–84. https://doi.org/10.1111/j.2044-8295.1964.tb00899.x

Conrad, R., & Hille, B. A. (1958). The decay theory of immediate memory and paced recall. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *12*(1), 1–6. https://doi.org/10.1037/h0083723

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, *55*(4), 429–432. https://doi.org/10.1111/j.2044-8295.1964.tb00928.x

Conway, A. R. A., & Engle, R. W. (1996). Individual differences in working memory capacity: more evidence for a general capacity theory. *Memory, 4*(6), 577–590. https://doi.org/10.1080/741940997

Cortis, C. (2015). *Comparing the immediate free recall of verbal and visuo-spatial stimuli:List length, capacity and output order effects in single- and dual-modality tasks* [Phd, University of Essex]. http://repository.essex.ac.uk/15490/

Cortis, C., Dent, K., Kennett, S., & Ward, G. (2015). First things first: Similar list length and output order effects for verbal and nonverbal stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(4), 1179–1214. https://doi.org/10.1037/xlm0000086

Cortis Mack, C., Cinel, C., Davies, N., Harding, M., & Ward, G. (2017). Serial position, output order, and list length effects for words presented on smartphones over very long intervals. *Journal of Memory and Language*, *97*, 61–80. https://doi.org/10.1016/j.jml.2017.07.009

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory* (1st ed., pp. 62–101). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.006

Conway, A. R. A., & Engle, R. W. (1996) Individual differences in working memory capacity: More evidence for a general capacity theory. Memory, 4(6), 577-590, DOI: 10.1080/741940997

Crannell, C. W., & Parrish, J. M. (1957). A comparison of immediate memory span for digits, letters, and words. *Journal of Psychology; Provincetown, Mass., Etc.*, *44*, 319–327. https://search.proquest.com/docview/1290654804/citation/725BEBF280CA43DCPQ/1

Crowder, R. G. (1969). Behavioral strategies in immediate memory. *Journal of Verbal Learning & Verbal Behavior*, *8*(4), 524–528. https://doi.org/10.1016/S0022-5371(69)80098-8

Crowder, R. G. (1969). On free recall and free-recall scoring in immediate memory. *Psychonomic Science*, *14*(6), 255–256. https://doi.org/10.3758/BF03329109

Crowder, R. G. (1993). Short-term memory: Where do we stand? *Memory & Cognition*, *21*(2), 142–145. https://doi.org/10.3758/BF03202725

Drewnowski, A. (1980). Attributes and priorities in short-term recall: a new model of memory span. *Journal of Experimental Psychology: General*, *109*(2), 208–250. https://doi.org/10.1037/0096-3445.109.2.208

Engle, R. W., & Oransky, N. (1999). Multi-store to dynamic models of temporary storage. In *The Nature of Cognition* (pp. 515–555). MIT Press.

Estes, W. K. (1972). Interactions of signal and background variables in visual processing. *Perception & Psychophysics*, *12*(3), 278–286. https://doi.org/10.3758/BF03207204

Glanzer, M. (1972). Storage mechanisms in recall. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 5, pp. 129–193). Academic Press. https://doi.org/10.1016/S0079-7421(08)60441-7

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., Bradley, M. M., Stevenson, J. A., Kraus, T. A., Tkachuk, M. J., Gretz, A. L., Fish, J. H., & Turpin, B. M. (1980). A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(4), 355–369. https://doi.org/10.1037/0278-7393.6.4.355

Grenfell-Essam, R., & Ward, G. (2012). Examining the relationship between free recall and immediate serial recall: The role of list length, strategy use, and test expectancy. *Journal of Memory and Language*, *67*(1), 106–148. https://doi.org/10.1016/j.jml.2012.04.004

Grenfell-Essam, R., Ward, G., & Tan, L. (2013). The role of rehearsal on the output order of immediate free recall of short and long lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 317–347. https://doi.org/10.1037/a0028974

Hulme, C., Neath, I., Stuart, G., Shostak, L., Surprenant, A. M., & Brown, G. D. A. (2006). The distinctiveness of the word-length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(3), 586–594. https://doi.org/10.1037/0278-7393.32.3.586

Hulme, C., Suprenant, A. M., Bireta, T. J., Stuart, G., & Neath, I. (2004). Abolishing the Word-Length Effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(1), 98–106. https://doi.org/10.1037/0278-7393.30.1.98

Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., & Koeppe, R. A. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience*, *9*(4), 462–475. https://doi.org/10.1162/jocn.1997.9.4.462

Katz, L. (1968). The limited-capacity hypothesis: Effects of sequence length and instructions in free recall. *Journal of Verbal Learning and Verbal Behavior*, *7*(5), 942–944. https://doi.org/10.1016/S0022-5371(68)80101-X

Kay, H., & Poulton, E. C. (1951). Anticipation in memorizing. *British Journal of Psychology. General Section*, *42*(1), 34–41. http://www.proquest.com/docview/1293573006/citation/4666258708744B8FPQ/1

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, *1*(3), 153–161. https://doi.org/10.1016/S0022-5371(62)80023-1

Kintsch, W., & Buschke, H. (1969). Homophones and synonyms in short-term memory. *Journal of Experimental Psychology*, *80*(3, Pt.1), 403–407. https://doi.org/10.1037/h0027477

Kleiman, G. M. (1975). Speech recoding in reading. *Journal of Verbal Learning and Verbal Behavior*, *14*(4), 323–339. https://doi.org/10.1016/S0022-5371(75)80013-2

Lakens, D., & Caldwell, A. R. (2019). Simulation-based power analysis for facotrial ANOVA designs. *PsyArXiv*. https://doi.org/10.31234/osf.io/baxsf

Lee, C. L., & Estes, W. K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Behavior*, *16*(4), 395–418. https://doi.org/10.1016/S0022-5371(77)80036-4

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. https://doi.org/10.1038/36846

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General*, *118*(4), 346–373. https://doi.org/10.1037/0096-3445.118.4.346

McKeown, D., & Mercer, T. (2012). Short-term forgetting without interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1057–1068. https://doi.org/10.1037/a0027749

Melton, A. W., & Irwin, J. M. (1940). The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *The American Journal of Psychology*, *53*(2), 173–203. JSTOR. https://doi.org/10.2307/1417415

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Miller, R. R. (2021). Failures of memory and the fate of forgotten memories. *Neurobiology of Learning and Memory*, *181*, 107426. https://doi.org/10.1016/j.nlm.2021.107426

Miyake, A., & Shah, P. (1999). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge University Press.

Murdock, B. B. (1974). *Human memory: Theory and data* (pp. x, 362). Lawrence Erlbaum.

Murdock, B. B. (1968). Serial order effects in short-term memory. *Journal of Experimental Psychology*, *76*(4, Pt.2), 1–15. https://doi.org/10.1037/h0025694

Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *21*(3), 263–276. https://doi.org/10.1037/h0082978

Nairne, J. S. (1988). A framework for interpreting recency effects in immediate serial recall. *Memory & Cognition*, *16*(4), 343–352. https://doi.org/10.3758/BF03197045

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, *18*(3), 251–269. https://doi.org/10.3758/BF03213879

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, *7*(3), 403–423. https://doi.org/10.3758/BF03214356

Neath, I., & Brown, G. D. A. (2006). SIMPLE: further applications of a local distinctiveness model of memory. In *Psychology of Learning and Motivation* (Vol. 46, pp. 201–243). Academic Press. https://doi.org/10.1016/S0079-7421(06)46006-0

Neath, I., & Crowder, R. G. (1990). Schedules of presentation and temporal distinctiveness in human memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 316–327. https://doi.org/10.1037/0278-7393.16.2.316

Neath, I., & Surprenant, A. M. (2003). *Human Memory: An Introduction to Research, Data, and Theory*. Thomson/Wadsworth.

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

Norris, D., Baddeley, A. D., & Page, M. P. A. (2004). Retroactive effects of irrelevant speech on serial recall from short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1093–1105. https://doi.org/10.1037/0278-7393.30.5.1093

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

Oberauer, K., & Kliegl, R. (2001). Beyond resources: formal models of complexity effects and age differences in working memory. *European Journal of Cognitive Psychology*, *13*(1–2), 187–215. https://doi.org/10.1080/09541440042000278

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, *55*(4), 601–626. https://doi.org/10.1016/j.jml.2006.08.009

Oberauer, K., & Lewandowsky, S. (2014). Further evidence against decay in working memory. *Journal of Memory and Language*, *73*, 15–30. https://doi.org/10.1016/j.jml.2014.02.003

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–958. https://doi.org/10.1037/bul0000153

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*(3), 193–198. https://doi.org/10.1037/h0049234

Plancher, G., & Barrouillet, P. (2020). On some of the main criticisms of the modal model: Reappraisal from a TBRS perspective. *Memory & Cognition*, *48*(3), 455–468. https://doi.org/10.3758/s13421-019-00982-w

Pollack, I., Johnson, L. B., & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology*, *57*(3), 137–146. https://doi.org/10.1037/h0046137

Postman, L., Stark, K., & Fraser, J. (1968). Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior*, *7*(3), 672–694. https://doi.org/10.1016/S0022-5371(68)80124-0

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134. https://doi.org/10.1037/0033-295X.88.2.93

Raaijmakers, J. G. W. (1979). *Retrieval from long-term store: A general theory and mathematical models*. [S.l. : s.n.].

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 14, pp. 207–262). Academic Press. https://doi.org/10.1016/S0079-7421(08)60162-0

Reitman, J. S. (1974). Without surreptitious rehearsal, information in short-term memory decay. *Journal of Verbal Learning and Verbal Behavior*, *13*(4), 365–377. https://doi.org/10.1016/S0022-5371(74)80015-0

Ricker, T. J., Vergauwe, E., & Cowan, N. (2016). Decay theory of immediate memory: from Brown (1958) to today (2014). *Quarterly Journal of Experimental Psychology*, *69*(10), 1969–1995. https://doi.org/10.1080/17470218.2014.914546

Roberts, W. A. (1972). Free recall of word lists varying in length and rate of presentation: a test of total-time hypotheses. *Journal of Experimental Psychology*, *92*(3), 365–372. https://doi.org/10.1037/h0032278

Roediger, H. L. (1974). Inhibiting effects of recall. *Memory & Cognition*, *2*(2), 261–269. https://doi.org/10.3758/BF03208993

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

Schneider, W., & Detweiler, M. (1988). Connectionist/control architecture for working memory. *Psychology of Learning and Motivation* (pp. 54–114). Academic Press.

Seamon, J. G., & Murray, P. (1976). Depth of processing in recall and recognition memory: Differential effects of stimulus meaningfulness and serial position. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(6), 680–687. https://doi.org/10.1037/0278-7393.2.6.680

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of*

Experimental Psychology: General, *125*(1), 4–27. https://doi.org/10.1037/0096-3445.125.1.4

Spurgeon, J., Ward, G., & Matthews, W. J. (2014). Examining the relationship between immediate serial recall and immediate free recall: Common effects of phonological loop variables but only limited evidence for the phonological loop. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1110–1141. https://doi.org/10.1037/a0035784

Spurgeon, J., Ward, G., Matthews, W. J., & Farrell, S. (2015). Can the effects of temporal grouping explain the similarities and differences between free recall and serial recall? *Memory & Cognition*, *43*(3), 469–488. https://doi.org/10.3758/s13421-014-0471-5

Surprenant, A. M., Neath, I., & Brown, G. D. A. (2006). Modeling age-related differences in immediate memory using SIMPLE. *Journal of Memory and Language, 55*(4), 572-586. https://doi.org/10.1016/j.jml.2006.08.001

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1589–1625. https://doi.org/10.1037/0278-7393.26.6.1589

Tulving, E., & Arbuckle, T. Y. (1963). Sources of intratrial interference in immediate recall of paired associates. *Journal of Verbal Learning and Verbal Behavior*, *1*(5), 321–334. https://doi.org/10.1016/S0022-5371(63)80012-2

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

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

Verhaeghen, P., & Basak, C. (2005). Ageing and switching of the focus of attention in working memory: results from a modified n-back task. *The Quarterly Journal of Experimental Psychology Section A*, *58*(1), 134–154. https://doi.org/10.1080/02724980443000241

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

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review*, *72*(2), 89–104. https://doi.org/10.1037/h0021797

Welch, G. B., & Burnett, C. T. (1924). Is primacy a factor in association-formation. *The American Journal of Psychology*, *35*(3), 396–401. https://doi.org/10.2307/1414018

Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *2*(5), 440–445. https://doi.org/10.1016/S0022-5371(63)80045-6

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00433

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11–11. https://doi.org/10.1167/4.12.11

Woods, D. L., Wyma, J. M., Herron, T. J., & Yund, E. W. (2016). An improved spatial span test of visuospatial memory. *Memory*, *24*(8), 1142–1155. https://doi.org/10.1080/09658211.2015.1076849

Appendix A

MEMORIAL
UNIVERSITY

Interdisciplinary Committee on
Ethics in Human Research (ICEHR)

St. John's, NL Canada A1C 5S7
Tel: 709 864-2561 icehr@mun.ca
www.mun.ca/research/ethics/humans/icehr

ICEHR Number:	**20201563-SC**
Approval Period:	February 27, 2020 – February 28, 2021
Funding Source:	Supervisor's NSERC [RGCS # 20131064]
Responsible Faculty:	Dr. Ian Neath, Department of Psychology
Title of Project:	*The list length effect in short term memory*

February 27, 2020

Ms. Molly MacMillan
Department of Psychology, Faculty of Science
Memorial University of Newfoundland

Dear Ms. MacMillan:

Thank you for your submission to the Interdisciplinary Committee on Ethics in Human Research (ICEHR), seeking ethical clearance for your research project. The Committee appreciates the care and diligence with which you prepared your application.

The project is consistent with the guidelines of the *Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans* (TCPS2). *Full ethics clearance* is granted for one year from the date of this letter. ICEHR approval applies to the ethical acceptability of the research, as per Article 6.3 of the *TCPS2* (2014). Researchers are responsible for adherence to any other relevant University policies and/or funded or non-funded agreements that may be associated with the project.

The *TCPS2* **requires** that you submit an Annual Update to ICEHR before February 28, 2021. If you plan to continue the project, you need to request renewal of your ethics clearance and include a brief summary on the progress of your research. When the project no longer involves contact with human participants, is completed and/or terminated, you are required to provide an annual update with a brief final summary and your file will be closed. If you need to make changes during the project which may raise ethical concerns, you must submit an Amendment Request with a description of these changes for the Committee's consideration. If funding is obtained subsequent to ethics approval, you must submit a Funding and/or Partner Change Request to ICEHR so that this ethics clearance can be linked to your award.

All post-approval event forms noted above must be submitted by selecting the ***Applications: Post-Review*** link on your Researcher Portal homepage. We wish you success with your research.

Yours sincerely,

Russell J. Adams, Ph.D.
Chair, Interdisciplinary Committee on
Ethics in Human Research
Professor of Psychology and Pediatrics
Faculties of Science and Medicine

RA/th

copy: Supervisor – Dr. Ian Neath, Department of Psychology
Director, Research Grant and Contract Services

Appendix B

Word Stimuli Used in Experiment 1, 2, 5 and 6

act, add, age, aid, aim, air, ape, arm, art, back, badge, bag, bait, ball, band, bank, bar, bark, barn, base, bat, batch, bath, bay, beach, beak, beam, bear, beard, beast, beat, bed, beef, bell, belt, bench, bend, bet, bid, bike, bin, bind, bird, birth, bit, black, blaze, block, bloom, blue, bluff, blur, board, boast, boat, boil, bold, bone, book, boom, boost, boot, booth, born, boss, bound, bow, bowl, box, boy, brain, brake, brass, brave, bread, break, breed, brick, brief, broad, brook, broom, brow, brown, brush, buck, build, bulb, bump, bunch, bunk, burst, bus, bush, bust, buy, buyer, cab, cake, calf, call, calm, camp, can, cane, cap, car, card, care, cart, case, cast, cat, catch, cause, cave, cell, cent, chain, chair, chalk, chap, charm, chart, chat, check, cheek, cheer, chest, chew, chief, child, chin, chip, choir, chop, claim, class, clay, clear, clerk, click, climb, cloak, clock, close, cloth, cloud, club, coach, coal, coast, coat, code, cold, comb, come, cook, cool, cop, cord, core, corn, cost, cot, couch, count, cow, crab, cream, crew, crop, cross, crowd, crown, crush, cube, cue, cup, cure, curl, curve, dad, dam, damp, dance, dare, dark, dash, date, dawn, day, deal, deck, deed, deep, deer, den, depth, desk, dew, dig, dip, dirt, dish, dive, dock, dog, doll, dome, door, dorm, dot, draw, dream, dress, drill, drop, drum, dry, duck, due, duke, dusk, dye, ear, earl, earth, ease, east, eat, edge, eel, egg, eight, elm, end, eye, face, fact, fade, faint, fair, faith, fame, fan, fare, farm, fate, feast, feat, feed, feel, fence, field, fig, file, fill, film, find, fine, firm, fish, fit, fix, flag, flap, flash, flat, fleet, flick, fling, float, flock, floor, flow, flush, fly, foam, fog, foil, fold, folk, food, foot, force, forge, fork, form, fox, frame, free, fruit, fry, full, fund, fur, fuse, gain, game, gap, gas, gasp, gate, gay, gaze, gear, get, ghost, gift, gin, girl, give, glad, glass, gleam, globe, glove, glow, goal, goat, golf, good, goose, gown, grab, grade, grain, grass, graze, great, green, grill, grin, grip, group, grove, guard, guess, guest, guide, gum, guy, gym, hail, hair, hall, halt, ham, hand, haste, hat, hatch, haul, haunt, have, hawk, hay, haze, head, heap, heart, heat, heave, hedge, heel, help, hen, herb, herd, hide, high, hill, hind, hint, hip, hire, hold, hole, home, hood, hook, hop, hope, horn, horse, host, hour, house, hug, huge, hut, ice, inch, ink, inn, jam, jar, jaw, jaws, jeans, jeep, jet, job, join, joke, joy, jug, juice, jump, keep, key, kick, kid, kids, kind, king, kit, knee, knife, knock, know, lab, lace, lad, lake, lamb, lamp, land, lane, lap, large, law, lawn, lay, lead, leaf, leak, lean, leap, leave, ledge, leg, lens, lid, life, lift, light, like, limb, lime, line, link, lip, list, load, loaf, loan, lock, lodge, log, look, loop, loot, lord, love, low, luck, lunch, lung, lure, mail, main, make, male, malt, man, map, mask, mass, mat, match, mate, meal, meat, meet, melt, mend, mild, mile, milk, mill, mind, mint, miss, mist, mix, month, mood, moon, mop, mound, mount, mouse, mouth, move, mud, mug, mule, must, myth, nail, name, nap, neck, need, nerve, nest, net, new, news, nice, niece, night, nod, noon, north, nose, note, nurse, nut, oak, oath, ounce, pace, pack, pad, page, paint, pair, pal, pale, palm, pan, park, part, pass, past, paste, patch, path, pause, pay, peace, peak, pearl, peel, peer, pen, pet, phase, phone, pick, pie, piece, pier, pig, pike, pile, pill, pin, pine, pink, pint, pipe, pitch, place, plain, plan, plate, play, plea, plot, plug, poet, point, poke, pole, poll, pond, pool, pop, porch, pork, port, pose, post, pot, pound, press, price, pride, prime, prize, proof, prop, pull, pulp, pulse, pump, purse, push, put, quick, quote, race, rack, rag, rail, rain, raise, ram, ramp, ranch, range, rank, rate, raw, reach, read, real, realm, rear, red, reef, reel, rent, rest, rice, ride, ridge, rig, right, rim, ring, ripe, rise, road, roast, robe, rock, rod, role, roll, roof, room, root, rope, rose, round, route, row, rub, rug, rule, rum, run, rust, sack, safe, sail, saint, sake, salt, sand, sauce, save, saw, say, scale, scene, scent, scoop, screw, sea, seal, seat, seed, seek, self, sell, send, sense, serve, set, shade, shake, shape, share, shark, shave, shawl, shed, sheep, sheet, shelf, shell, shift, shine, ship, shirt, shoe, shop, short, show, shrug, shy, side, sigh, sight, sign, silk, sink, sip, site, size, skid, skill, skin, skirt, sky, slab, slate, sleep, slice, slick, slide, slip, slit, slope, slot, small, smash, smell, smile, snap, snow, soak, soap, soil, sole, son, song, soul, sound, soup, south, space, span, spare, spear, spell, spin, spine, spoon, spot, spray, spur, spy, stack, staff, stage, stake, stall, star, stare, state, stay, steak, steam, steel, steep, steer, stem, step, stew, stick, stir, stock, stone, stop, store, stout, stove, straw, stray, stuff, style, suit, suite, sum, sun, sure, surf, surge, sway, sweep, sweet, swell, swing, sword, tag, tail, take, tale, talk, tall, tan, tank, tap, tape, task, taste, tea, teach, team, tease, teeth, tell, tent, term, test, theme, thick, thing, think, throw, thumb, thump, tide, tie, tilt, time, times, tin, tip, toast, toe, tone, tool, tooth, top, torch, toss, touch, tough, tour, towel, tower, town, toy, trace, track, trade, trail, train, tray, tread, treat, tree, trick, trim, trip, troop, trot, trout, truck, truth, try, tub, tube, tug, tune, turf, turn, tweed, twin, type, urge, van, vase, vat, vein, verse, vest, vet, view, voice, vote, vow, wage, waist, wait, wake, walk, wall, want, ward, wash, watch, wave, wax, way, wear, web, wedge, week, west, wet, whale, wheat, wheel, white, whole, width, wife, wig, wild, will, win, wind, wine, wing, wipe, wire, wise, wish, wit, wolf, wood, wool, word, work, world, worm, wrap, wrist, yard, year, yeast, yield, young, youth, zinc, zone

Appendix C

Word Stimuli Used in Experiment 3

ants, arm, aunt, back, bar, beach, beans, bear, bench, bike, branch, brass, bread, bus, bush, cake, calf, card, cave, chalk, chart, cheek, church, class, clouds, cone, corn, cot, dad, deck, dock, drill, ears, east, face, fern, fish, flame, flute, fork, game, germs, ground, hall, hat, heart, hill, hole, jail, jam, jar, job, kid, lamp, man, mold, nerve, net, paint, pipe, rat, sea, street, tin, wax, world

Appendix D

Math Questions Used for the Operation Span Task in Experiment 3, 4, and 5

is (2 x 1) - 1 = 1?
is (2 x 2) + 1 = 4?
is (2 x 3) + 1 = 4?
is (3 / 1) - 2 = 3?
is (3 / 1) + 3 = 6?
is (3 x 1) + 2 = 2?
is (3 x 2) + 1 = 6?
is (3 x 3) - 1 = 8?
is (4 / 1) - 4 = 2?
is (4 / 1) + 1 = 4?
is (4 / 2) + 1 = 6?
is (4 x 2) + 1 = 9?
is (4 x 4) + 1 = 17?
is (5 / 1) + 4 = 9?
is (5 / 5) + 1 = 2?
is (5 x 1) - 1 = 4?
is (5 x 1) - 1 = 5?
is (5 x 2) + 1 = 6?
is (6 / 2) - 2 = 2?
is (6 / 2) - 2 = 2?
is (6 / 2) + 1 = 4?
is (6 / 3) + 2 = 4?
is (6 x 1) - 4 = 1?
is (6 x 2) - 2 = 10?
is (6 x 2) - 3 = 10?
is (6 x 3) - 2 = 11?
is (6 x 3) + 2 = 17?
is (6 x 4) + 1 = 25?
is (7 / 1) + 2 = 7?
is (7 / 1) + 6 = 12?
is (7 x 1) + 6 = 13?
is (7 x 2) - 1 = 14?
is (7 x 2) - 3 = 11?
is (7 x 7) + 1 = 49?
is (8 / 1) - 6 = 4?
is (8 / 2) - 1 = 3?
is (8 / 2) + 4 = 2?
is (8 / 4) - 1 = 1?
is (8 / 8) + 1 = 2?
is (8 x 1) + 5 = 13?
is (8 x 1) + 8 = 16?
is (8 x 4) - 2 = 32?
is (8 x 4) + 2 = 34?
is (9 / 1) - 5 = 4?
is (9 / 1) - 7 = 4?
is (9 / 1) + 8 = 18?
is (9 / 3) - 1 = 2?
is (9 / 3) - 1 = 2?
is (9 / 3) - 2 = 1?
is (9 x 1) - 1 = 8?
is (9 x 1) + 9 = 1?
is (9 x 2) - 3 = 16?
is (9 x 2) + 1 = 18?
is (9 x 3) - 2 = 25?
is (9 x 3) - 3 = 24?
is (10 / 1) - 1 = 9?
is (10 / 1) + 1 = 10?
is (10 / 1) + 3 = 13?
is (10 / 2) - 3 = 2?
is (10 / 2) - 4 = 3?
is (10 / 2) + 4 = 3?
is (10 / 2) + 4 = 9?
is (10 / 2) + 4 = 9?
is (10 x 2) - 6 = 12
is (10 x 2) + 3 = 23?
is (10 / 10) - 1 = 2?

Appendix E

Math Questions Used for the Operation Span Task in Experiment 6

is (2 x 1) - 1 = 1?

is (2 x 1) - 1 = 2?

is (2 x 1) + 1 = 3?

is (2 x 1) + 1 = 4?

is (2 x 2) - 1 = 3?

is (2 x 2) - 1 = 2?

is (2 x 2) + 1 = 4?

is (2 x 2) + 1 = 5?

is (2 x 2) - 2 = 2?

is (2 x 2) - 3 = 2?

is (2 x 3) - 1 = 4?

is (2 x 3) + 1 = 4?

is (2 x 3) + 2 = 6?

is (2 x 3) - 2 = 4?

is (2 x 4) - 1 = 9?

is (2 x 4) + 1 = 9?

is (2 x 4) - 2 = 4?

is (2 x 4) + 2 = 10?

is (3 / 1) - 2 = 3?

is (3 / 1) + 3 = 6?

is (3 / 3) + 2 = 3?

is (3 / 3) + 3 = 3?

is (3 x 1) + 2 = 2?

is (3 x 1) - 2 = 1?

is (3 x 1) + 1 = 2?

is (3 x 1) - 1 = 2?

is (3 x 1) + 3 = 7?

is (3 x 1) + 3 = 6?

is (3 x 2) + 1 = 6?

is (3 x 2) - 1 = 5?

is (3 x 2) + 3 = 8?

is (3 x 2) - 3 = 3?

is (3 x 3) - 1 = 8?

is (3 x 3) + 1 = 8?

is (3 x 3) - 2 = 7?

is (3 x 3) + 2 = 8?

is (4 / 1) - 4 = 2?

is (4 / 1) + 1 = 4?

is (4 / 1) - 2 = 2?

is (4 / 1) + 3 = 4?

is (4 / 2) + 1 = 6?

is (4 / 2) - 1 = 1?

is (4 / 2) + 2 = 6?

is (4 / 2) - 2 = 2?

is (4 x 1) + 1 = 5?

is (4 x 1) - 1 = 5?

is (4 x 2) + 1 = 9?

is (4 x 2) - 1 = 9?

is (4 x 2) + 2 = 10?

is (4 x 2) - 2 = 4?

is (4 x 2) + 3 = 11?

is (4 x 2) - 3 = 4?

is (4 x 3) + 1 = 12?

is (4 x 3) - 1 = 11?

is (4 x 3) + 2 = 16?

is (4 x 3) - 2 = 10?

is (4 x 3) + 3 = 16?

is (4 x 3) - 3 = 9?

is (4 x 4) + 1 = 17?

is (4 x 4) - 1 = 15?

is (4 x 4) - 2 = 14?

is (4 x 4) + 2 = 10?

is (4 x 4) - 3 = 12?

is (4 x 4) + 3 = 19?

is (5 / 1) - 1 = 6?

is (5 / 1) + 1 = 6?

is (5 / 1) + 1 = 7?
is (5 / 1) - 2 = 3?
is (5 / 1) + 2 = 6?
is (5 / 1) + 3 = 9?
is (5 / 1) - 3 = 2?
is (5 / 1) + 4 = 9?
is (5 / 1) - 4 = 1?
is (5 / 5) + 1 = 2?
is (5 / 5) - 1 = 2?
is (5 / 5) + 2 = 7?
is (5 / 5) + 2 = 3?
is (5 / 5) + 3 = 8?
is (5 / 5) + 3 = 4?
is (5 x 1) - 1 = 4?
is (5 x 1) - 1 = 5?
is (5 x 1) + 1 = 6?
is (5 x 1) + 1 = 7?
is (5 x 1) - 2 = 3?
is (5 x 1) - 2 = 2?
is (5 x 1) + 2 = 7?
is (5 x 1) + 2 = 8?
is (5 x 1) + 3 = 9?
is (5 x 1) + 3 = 8?
is (5 x 1) - 3 = 2?
is (5 x 1) - 3 = 3?
is (5 x 1) - 4 = 1?
is (5 x 1) + 4 = 9?
is (5 x 1) + 4 = 10?
is (5 x 2) + 1 = 6?
is (5 x 2) - 1 = 8?
is (5 x 2) + 1 = 10?
is (5 x 2) - 1 = 9?
is (5 x 2) + 2 = 10?
is (5 x 2) - 2 = 12?
is (5 x 2) + 2 = 12?
is (5 x 2) - 2 = 10?
is (5 x 2) + 3 = 10?
is (5 x 2) - 3 = 8?
is (5 x 2) + 3 = 13?
is (5 x 2) - 3 = 9?
is (5 x 2) + 4 = 6?
is (5 x 2) - 4 = 8?
is (5 x 2) + 4 = 14?
is (5 x 2) - 4 = 6?
is (6 / 2) - 2 = 2?
is (6 / 2) - 2 = 2?
is (6 / 2) + 1 = 4?
is (6 / 2) - 1 = 2?
is (6 / 3) + 2 = 4?
is (6 / 3) + 2 = 6?
is (6 x 1) - 2 = 3?
is (6 x 1) - 2 = 4?
is (6 x 1) + 2 = 4?
is (6 x 1) + 2 = 8?
is (6 x 1) - 3 = 3?
is (6 x 1) + 3 = 6?
is (6 x 1) - 4 = 1?
is (6 x 1) + 4 = 10?
is (6 x 2) - 1 = 13?
is (6 x 2) + 1 = 13?
is (6 x 2) - 1 = 11?
is (6 x 2) + 1 = 11?
is (6 x 2) - 2 = 10?
is (6 x 2) + 2 = 12?
is (6 x 2) - 3 = 10?
is (6 x 2) + 3 = 9?
is (6 x 3) - 2 = 11?
is (6 x 3) + 2 = 17?
is (6 x 3) - 3 = 15?
is (6 x 3) + 2 = 20?
is (6 x 4) + 1 = 25?
is (6 x 4) - 1 = 23?
is (6 x 4) + 1 = 22?
is (6 x 4) - 6 = 18?
is (7 / 1) + 2 = 7?

is (7 / 1) + 6 = 12?

is (7 x 1) + 6 = 13?

is (7 x 2) - 1 = 14?

is (7 x 2) - 3 = 11?

is (7 x 7) + 1 = 49?

is (8 / 1) - 6 = 4?

is (8 / 2) - 1 = 3?

is (8 / 2) + 4 = 2?

is (8 / 4) - 1 = 1?

is (8 / 8) + 1 = 2?

is (8 x 1) + 5 = 13?

is (8 x 1) + 8 = 16?

is (8 x 4) - 2 = 32?

is (8 x 4) + 2 = 34?

is (9 / 1) - 5 = 4?

is (9 / 1) - 7 = 4?

is (9 / 1) + 8 = 18?

is (9 / 3) - 1 = 2?

is (9 / 3) - 1 = 2?

is (9 / 3) - 2 = 1?

is (9 x 1) - 1 = 8?

is (9 x 1) + 9 = 1?

is (9 x 2) - 3 = 16?

is (9 x 2) + 1 = 18?

is (9 x 3) - 2 = 25?

is (9 x 3) - 3 = 24?

is (10 / 1) - 1 = 9?

is (10 / 1) + 1 = 10?

is (10 / 1) + 3 = 13?

is (10 / 1) - 3 = 8?

is (10 / 2) - 3 = 2?

is (10 / 2) - 4 = 3?

is (10 / 2) + 4 = 3?

is (10 / 2) + 4 = 9?

is (10 / 2) - 4 = 9?

is (10 x 2) - 6 = 12?

is (10 x 2) + 3 = 23?

is (10 x 2) - 4 = 14?

is (10 x 2) + 4 = 24?

is (10 / 5) + 1 = 3?

is (10 / 5) - 1 = 1?

is (10 / 5) + 2 = 3?

is (10 / 5) - 2 = 1?

is (10 / 5) + 3 = 5?

is (10 / 5) + 3 = 8?

is (10 / 10) - 1 = 2?

is (10 / 10) - 1 = 2?