The recall of information from working memory: insights from behavioural and chronometric perspectives

Towse, John N. (Department of Psychology, Lancaster University)

Cowan, Nelson (Department of Psychological Sciences, University of Missouri)

Hitch, Graham J. (Department of Psychology, University of York)

Horton, Neil J. (Department of Psychology, Lancaster University)

To appear in: *Experimental Psychology*. Please refer to the journal URL for a definitive version of the document.

Address for correspondence

Dr John N Towse

Department of Psychology

Fylde College, Lancaster University

Bailrigg, Lancaster LA1 4YF, United Kingdom

Tel: +44-(0)1524-593705 Fax: +44-(0)1524-593744

Email: j.towse@lancaster.ac.uk

Running title: Recall from working memory

Acknowledgements. We acknowledge the support of the ESRC (grant RES000220452 and RES000230859). The paper was prepared in part while the first author was a research visitor at the University of Colorado at Boulder. Scott Saults undertook blind recall timings on some trials of Experiment 1 and allowed derivation of consistency measurements. An overview of the procedures for timing spoken responses can be found at the URL:

http://www.psych.lancs.ac.uk/research/TowseWM/

Abstract

We describe and evaluate a recall reconstruction hypothesis for working memory, according to which items can be recovered from multiple memory representations. Across four experiments, participants recalled memoranda that were either integrated with or independent of the sentence content. We found consistently longer pauses accompanying the correct recall of integrated compared with independent words, supporting the argument that sentence memory could scaffold the access of target items. Integrated words were also more likely to be recalled correctly, dependent on the details of the task. Experiment 1 investigated the chronometry of spoken recall for word span and reading span, with participants completing an unfinished sentence in the latter case. Experiment 2 and 3 confirm recall time differences without using word generation requirements, while Experiment 4 used an item and order response choice paradigm with nonspoken responses. Data emphasise the value of recall timing in constraining theories of working memory functioning.

KEYWORDS: working memory; reading span; recall timing; recall method; short-term memory

The recall of information from working memory: insights from behavioural and chronometric perspectives

Introduction

Working memory reflects the ability to hold in mind transient representations while simultaneously processing and assimilating ongoing events (Baddeley & Hitch, 1974). There are a wide variety of circumstances in which we are required to carry out mental operations and remember intermediate information (for instance, retain a carry item in mental arithmetic or a referent for an anaphoric pronoun) or draw on past episodic knowledge (e.g., mapping the problem space for a current task using knowledge of related situations). They emphasise the importance of understanding active maintenance and transformation processes. Consequently, the concept of working memory has been the focus of considerable research.

The most common method for assessing working memory capacity is to draw upon at least one of a family of tasks known as working memory (WM) span. Reading span was the first such task to be reported in adults (Daneman & Carpenter, 1980). Participants read a series of unconnected sentences, the final word in each sentence yielding a memorandum to be reported afterwards in serial order. In essence, an individual's reading span score reflects how many end-of-sentence words can be remembered whilst reading. Daneman and Carpenter (1980) showed reading span to be a very good measure of reading skill (see also Daneman & Merikle, 1986). The predictive prowess of WM span tasks (including alternatives such as operation span where the processing task involves arithmetic operations, Turner & Engle, 1989) provides empirical support for the conceptual idea that the processing-plus-memory requirements tap an important skill in complex cognition.

Several theories suggest, in different ways and to different degrees, that a competitive relationship between processing and memory activities is critical to measuring WM capacity (see Miyake & Shah, 1999). In other words, the suggestion is that the maintenance of information takes place in the face of distraction or interference from concurrent processing. For example, Case (1985) proposed that limited-capacity general-purpose cognitive resources were allocated to *either* processing or memory demands. Jarrold and Bayliss (2007) discuss evidence that combining or coordinating processing with memory places demands on WM, above and beyond those imposed by each requirement *per se*. Towse, Hitch and Hutton (1998) argued that processing activity produces informational degradation because memories are not actively or continuously maintained (in this respect, see also Barrouillet, Bernadin, & Camos, 2004). Kane and Engle (2003) suggested that controlled attention is required to preserve memory representations at the same time as the concurrent processing requirements.

We can see important insights to be gained from each of these accounts, and we do not intend to arbitrate between them here. Rather, our focus is directed towards the concept that links them; the idea that processing and memory are separable and place competing demands on WM. We certainly accept that processing can interfere with retention. Nonetheless, we present data that lead us to conclude that this is not the whole story; processing may also complement or support memory (for an early and seminal version of this perspective, see Craik & Lockhart, 1972).

Our core proposal is that processing and memory need not always be thought of as completely separate. Using behavioural and chronometric evidence, we propose that psychological models can be enhanced by encompassing a broader view of the nature of WM representations present at the point of recall.

Chronometric analyses of memory recall – the timing of correct output sequences – has generally focused on short-term memory (STM) tasks such as word span where a sequence of unrelated items is presented at a regular rate and then reproduced in the original order (e.g., Dosher & Ma, 1998; for an overview, Towse & Cowan, 2005). Whilst such research has been undoubtedly productive, given the body of evidence to distinguish STM from WM (e.g., Daneman & Carpenter, 1980; Engle, Kane & Tuholski, 1999), there is a clear motivation to investigate recall timing in WM. Cowan et al. (2003) did just this. They found children's reading span recall times were dramatically longer than has been obtained in STM studies, and that for both children and adults (but especially the former) response durations for listening span exceeded those of counting span and digit span. Cowan et al. also reported a negative correlation between recall duration and children's word reading skills; moreover recall duration explained variance over and above that from span scores per se: recall evidently incorporates processes relevant to children's cognitive development and attainment.

The particularly long interword pauses in reading span and listening span led Cowan et al. (2003) to two related conclusions. First, memory representations may not always be maintained in a highly accessible state during processing. If they had been, one would expect their rapid production during recall. Second, participants sometimes draw on memory of the sentence, in terms of thematic and semantic context, to access the target items. Cowan et al. found that recall in counting span was less protracted than listening span, and attributed this to the lack of distinctive processing in the enumeration of visual displays, and thus the absence of a similar scaffolding process. Thus, reading span and listening span recall can involve the consideration of a much richer ensemble of (perhaps loosely encoded) memories of the trial episode than is the case for other tasks.

To capture these ideas we propose a ‘recall reconstruction’ hypothesis for WM performance. The central idea is that participants may bring to recall more than just representations of experimentally-assigned memoranda (i.e. target memory words). In the specific case of reading span, this can involve for example sentential information. As a consequence, we argue that the memory sequence may not be continuously and actively maintained and consequently recall involves the resuscitation of degraded information.

According to this recall reconstruction perspective, WM potentially involves the intertwined and integrated aspects of processing and memory. Processing and memory activities need not inherently be in complete opposition to each other, dependent on the specific WM task. Consider an example sentence from Daneman and Carpenter (1980): “*The lumbermen worked long hours in order to obtain the necessary amount of wood*.” According to the position just outlined, later recall of “wood” might be facilitated by gist or episodic memory about lumbermen, or indeed associations made during reading to the implicit concept “trees”. An individual need not commit a sentence to memory verbatim, but relevant linguistic information could nonetheless be accessible, either to help reconstruct the word "wood" or to discount sentence-terminal words appropriate to other serial positions. In a similar vein, Saito & Miyake (2004) have pointed to a relationship between processing activity and memory within their representation-based interference account of WM. While they concentrated on how processing events can hinder memory, via overlapping representations that interfere, their position is also that the content of processing affects memory performance.

So far as we are aware, no studies have directly investigated the link between processing meaning and memory requirements in WM. However, Osaka, Nishizaki, Komori & Osaka (2002) studied reading span, and for language-specific reasons underlined the word that was to be remembered. This was either a "focus" word – the most important word for sentence meaning- or a less-central "non-focus" word. Recall was substantially and significantly better for focus words. While all their memoranda were thematically connected with sentence material, their data support a link between processing material and what is remembered.

Other indirect evidence is relevant to this issue. Copeland & Radvansky (2001) reported a reversed phonemic similarity effect in reading span, but a standard effect for operation span, concluding that phonemically similar lists were at an advantage because recall was facilitated by the sentence context. Hitch, Towse & Hutton (2001: Fig. 2) reported that among children, the rate of forgetting on an operation span task (as a function of arithmetic processing time) was faster than the rate of forgetting on a reading span task (as a function of sentence processing time). Distinctive sentences may help to retard forgetting because they can be used in the reconstruction of target memoranda, in a way that less distinct arithmetic problems cannot. However, Conlin, Gathercole & Adams (2005) argued against reconstructive processes in WM because they found superior recall of memoranda when categorically distinct from processing (e.g. remembering numbers after reading sentences, or remembering words after mental arithmetic). Yet this confounds categorical similarity of material with opportunity for scaffolding.

In a series of experiments, we test the recall reconstruction hypothesis directly. Its validity is important because it addresses the widespread assumption that processing and memory are necessarily competitive components of WM tasks. Yet it also broadens the conceptual focus, encouraging theoretical models of WM to incorporate recall and not just maintenance processes (Unsworth & Engle, 2006). We therefore attempt to replicate the long interword pauses found in reading span by Cowan et al. (2003) and test accounts of what underlies this phenomenon. We focus on interword pauses because it is here specifically that participants engage in mental search processes to access the next sequence item, and because other recall segments such as preparatory intervals have dissociable patterns of performance (Cowan et al., 1998; Towse & Cowan, 2005). The experiments substantially extend the analysis of Cowan et al. by providing converging paradigms to investigate recall timing, using spoken recall as well as non-spoken responses. This latter approach opens up new opportunities for studying the chronometry of recall and the nature of memory representations.

Experiment 1

To examine why reading span performance is characterised by long response durations, we manipulated the relevance of sentential processing for the memory items, based on the reading task structure used by Cowan et al. (2003), Expt. 1 (see also Towse, Hitch & Hutton, 2000) and for comparison we also included a STM task. On a reading span trial a participant might read the unfinished sentence “The rocket went into outer ___” and we would expect them to suggest “space” as the completion word. The memorandum could be (a) “space, and thus connected or ‘integrated’ with processing or (b) “bridge”, and thus unrelated to or ‘independent’ of processing. Since the condition a) word is integrated within the sentence, participants can use sentence representations (e.g., knowledge that it referred to a rocket) to inform their recall choice although drawing upon this additional information will slow down recall. This is not possible with condition b) which should therefore be recalled more quickly since participants must use alternative – albeit potentially effective – maintenance processes for the sequence of items.

Method

Participants

Twenty-four Lancaster University students (22 women and 2 men) volunteered via departmental recruitment procedures and were paid £3. They were randomly assigned to the integrated and independent condition, as described below.

Stimuli

A corpus of 88 sentences (based on medium-length stimuli described in Towse, Hamilton, Hitch & Hutton, 2000) were randomly divided into two sets, “set A” and “set B”. Allocation of sentence sets to participants was randomised with the alternative end-of-sentence items used for word span stimuli. Sentences typically contained 8-10 words and had been formulated to elicit target completion words with a high probability among children (for example, "While I was sleeping I had a strange" typically leads to the completion response "dream").

Apparatus

Computer events were driven by an Apple Macintosh ibook G4 with 14 inch screen (programmed using the “Revolution” language running under OS X) with response latencies measured in (1/60 s) ticks. Audio recordings were captured directly to minidisk (Sony MZ-N710, with a Sony ECM-DS70P microphone).

Procedure

Following task instructions, all participants undertook a STM (word span) test, and either the integrated or independent word WM (reading span) test in counterbalanced order.

Reading span trials. On each trial, a set of (between 2 and 5) incomplete sentences appeared sequentially on screen. Participants read each sentence aloud and generated a suitable completion word. Afterwards, they attempted serial recall of the memoranda. Trials commenced with 3 sets of 2 sentences. Provided at least one memory sequence was recalled correctly, an additional sentence was added to the series and 3 further trials were administered, up the maximum 5-sentence sets. Participants knew the list length prior to each trial.

In the integrated word condition, the sentence completion words formed the memoranda (if a non-expected response was produced, this was adopted as the memory target). Once the participant completed the sentence, the experimenter immediately tapped a computer key to initiate the next sentence or the recall cue, which occurred after a 1 second interval. Participants were instructed not to rehearse words whilst reading and to begin reading each sentence immediately.

In the independent word condition, participants also supplied a completion word for each sentence. This was followed by a separate, unconnected word to be read aloud, which formed the memorandum. This appeared for .5 sec surrounded by .25 sec ISIs. These timing parameters ensured the independent and integrated conditions were equivalent in overall duration. The independent words for “set A” were taken from the “set B” pool, and vice versa.

Word span trials. Participants watched the visual presentation of a sequence of unconnected words. Each word was shown centre-screen for 0.5 sec, with a 0.5 sec ISI. Initially there were 3 trials containing 2 memoranda, and sequence length increased by a single item, provided that at least one trial was successfully recalled, up to a maximum list length of 5 words.

Recall

Instructions asked participants to recall the word sequence to a trial as soon as (but not before) the computer produced an auditory-visual recall signal. Participants were asked to limit their spoken response to the recall words only (to avoid other words such as “I think”, “then it was” or “and”) and were reminded if necessary during test administration. The experimenter recorded answers onto computer after the output sequence was complete.

Results

We examined the effect of task administration order (whether word span or reading span was administered first, for each stimulus pool set and task configuration) and found no significant effects on global timing measurements. We therefore collapse across order in subsequent analyses.

Sentence reading times for the integrated word and independent word formats (M=3.39, SD=.31 and M=3.29, SD=.64 respectively) were equivalent, $t(22)=.47$, $p=.647$, $\eta^2=.009$. Following recommendations by Conway et al. (2005), and Friedman and Miyake (2004a) accuracy is measured in terms of correctly recalled words (with respect to item and position). These are shown in Table 1, expressed as a proportion of the maximum attainable. Analysis confirmed memory accuracy was substantially greater for integrated words compared to independent words, $t(22)=5.01$, $p<.001$, $\eta^2=.533$[1].

Recall times were extracted only from those trials where serially ordered items were fully correct. In some cases, data were excluded because the participant restarted their list (e.g. "yellow… dream…. no, wait, door…yellow…dream") or in some way gave an ambiguous report with respect to timing issues.

Spoken recall was segmented into three contiguous phases (see, for example, Cowan et al., 1998); the time between the recall signal and the start of recall (the preparatory interval); the time to articulate the relevant words (each word duration), and the gaps between words (interword pauses). A single trained researcher extracted timing values, for whom blind timings both correlate and correspond with those made by an independent coder (for a sample of 99 word and interval measurements, $r(97)=.993$). Specific recall time segments were screened for outliers by examining z-score distributions of each time measurement; where $z>2.58$, that interval duration was excluded. Measures of stability are reported in Table 2. To make analysis easier to present, we focus on the three recall phases,

---

[1] Where appropriate as here, probabilities have been corrected after adjusting degrees of freedom because of non-equal variances. In the case of analysis of variance, we report Greenhouse-Geisser values where warranted.

combining individual values (e.g., the first and second word in two-item sequences).

Figure 1 shows the profile of recall durations. At list length 2, the mean pause between integrated words was longer than that between independent words, although this difference was only marginally significant, $t(22)=2.00$, $p=.058$, $\eta^2=.154$. At list length 3, the effect was in the same direction, but not significant, $t(14)=.57$, $p=.580$, $\eta^2=.023$, while at list length 4 there were too few data points in the independent condition for analysis. Combining data across list length 2 and 3, pauses in the integrated word condition were twice the length of the independent word condition, $t(22)=2.13$, $p=.045$, $\eta^2=.171$.

Preparatory intervals and word durations were similar between integrated and independent words and comparisons at list length 2 and 3 were non significant (ts<1.36, ps>.187, $\eta^2<.078$). This dissociation in sensitivity is consistent with the notion that the separate recall segments can be differentiated (Cowan et al., 1998). We carried out additional analysis on STM recall, but since these are less relevant to the main experimental issue, they are reported in Appendix 1.

Discussion

Cowan et al. (2003) emphasised differences between STM and WM recall with respect to pause length in particular. Furthermore, the longer recall was most evident for WM tasks that involved linguistic-based processing. The recall reconstruction hypothesis explains this finding by proposing that the specification of a target item might include consideration of representations persisting from the sentences, and this takes time.

The present study experimentally evaluated this hypothesis by manipulating the link between sentential processing and memory items. When processing events can scaffold recall, pauses should be extended as a richer set of representations are evaluated. Indeed, gaps between words were longer when processing and memory were linked and more sequences were correctly recalled.

All other things being equal one might expect that a difference in memory accuracy would work <u>against</u> the obtained recall time difference since accessing the correct item should be more difficult with weaker memory representations. The more rapid correct recall in the independent condition is therefore all the more telling. That this effect was not also obtained for the preparatory intervals emphasises how this recall segment has dissociable properties, and likely involves a number of processes that are not specifically tied to the production of the first item.

To forestall possible mis-interpretation, we do not suggest that the processing event in reading span provides only a supportive environment for recall. Memory for the sentence ideas, or sentence words, may well degrade access to the designated memory item, offering alternative recall candidates and adding to the problem of discriminating between memoranda and activated non target representations (Conlin et al. 2005; Saito & Miyake, 2004; see also below). Our argument is that the elicitation of recall words can be affected by memory for the processing material, and that this emergent property of the way reading span trials are often constructed is one contributory factor to the protracted recall of items reported here and elsewhere (Cowan et al., 2003).

Experiment 2

Experiment 1 asked participants to read aloud an incomplete sentence and formulate a (constrained and thus predictable) completion word. This processing requirement has been used among adults and children (e.g. Leather & Henry, 1994; Towse et al., 1998; 2000) and helps to ensure that participants attend to the sentences. At the same time, we note that the integrated condition involves memory for a self-generated item, while the independent condition involves memory for a different item that follows the self-generated word.

Slamecka and Graf (1978) demonstrated a recall accuracy advantage for self-generated items in comparison with read items although subsequent work has suggested a dissociation between (positive) item and (negative) order effects in self-generation (Nairne, Riegler & Serra, 1991). One recent account is that generation encourages semantic processing of material and context (Mulligan, 2004; see also Steffens & Erdfelder, 1998). In this sense, generation could affect reading span recall by elaborating and enriching the memory representation, similar to the time-consuming recall reconstruction processes envisaged here. Notwithstanding this overlap, to understand better the locus of the accuracy and interword pause effects, we removed the generation requirements from the reading span task in the next study.

Method

Participants

Twenty-eight Lancaster University students (22 women and 6 men) were paid £3 to complete reading span trials and ancillary tasks (not described here). They were randomly assigned to the integrated and independent condition.

Stimuli and Apparatus

We used the same apparatus and sentences as in Experiment 1 except that the latter were presented in completed form; there were no missing words.

Procedure

Participants completed either the (a) integrated or (b) independent word task. (a) Each sentence appeared in black type with the final word in purple - to demarcate that it was a to-be-remembered item. There was a 1-sec ISI. (b) Each (completed) sentence appeared entirely in black type. The purple memory word appeared for .5 sec (surrounded by .25 sec ISIs). Timing parameters were again designed to equate durations across task format. Instructions were the same except that rather than furnishing a sentence completion, participants were instructed to "read the sentences aloud and think about the sentence meaning as you do so."

Results

Sentence reading times for integrated and independent conditions ($\underline{M}$=2.95, $\underline{SD}$=.50 and $\underline{M}$=2.84, $\underline{SD}$=.33 respectively) were equivalent, $\underline{t}(26)=.71$, $\underline{p}=.488$, $\eta^2=.019$, and just slightly quicker than in Experiment 1 (sampling differences and the absence of sentence completion requirements could explain the discrepancy). In terms of the proportion of correctly recalled words, shown in Table 1, the integrated format enjoyed an advantage, but in this dataset it was not significant, $\underline{t}(26)=1.46$, $\underline{p}=.156$, $\eta^2=.076$. Notably, the major difference from Experiment 1 is that without the generation requirements, performance in the independent condition has improved.

As in Experiment 1, recall time outliers were screened prior to compilation of trial data[2]. Recall times are illustrated in Figure 2 (Table 2 reports stability measures). At list length 2, the pauses between integrated words were significantly longer than between independent words, $t(26)=2.24$, $p=.034$, $\eta^2=.162$. The average list length 3 pause was also significantly longer in the integrated condition, $t(24)=2.54$, $p=.020$, $\eta^2=.212$, and the pause difference averaged across list lengths 2 and 3 was likewise significant, $t(26)=3.08$, $p=.005$, $\eta^2=.267$. Word duration differences at list length 2 were in the same direction as above but only marginally significant, $t(26)=1.96$, $p=.061$, $\eta^2=.129$, while there was no reliable difference in for list length 3, $t(24)=.22$, $p=.828$, $\eta^2=.002$. There were no differences in preparatory intervals (e.g. averaged across both list lengths, $t(26)=.32$, $p=.751$, $\eta^2=.004$). At list length 4 there were few data points for meaningful analysis.

Discussion

The data offer further support for the recall reconstruction hypothesis. We again found evidence that interword pauses in recall are longer when there is a connection between the memoranda and the processing context associated with them. There was also a trend for superior levels of recall in the integrated condition but this was not significant in the current dataset.

Thus, word generation *per se* cannot be wholly responsible for slow but accurate recall in the integrated condition (see Conlin et al., 2005, for additional evidence of phenomena robust across generation effects). The clear point of change across experiments lies in recall from the independent condition. Generating a

---

[2] Two independent judges extracted timing measurements, using the same procedures as Expt. 1. Independent t-tests on all list-length 2 and list-length 3 segments indicated comparable judgements between coders (all ps>.10)

word to complete a sentence (Experiment 1), as opposed to just reading a sentence (Experiment 2), makes it harder to then recall a separate word that follows. This could be because in the independent condition the generated - irrelevant word - affects the encoding of the subsequent item or maintenance of items already encoded (this is investigated further in Experiment 4 below). Regardless, data demonstrate that longer interword pauses in recall are due to the processing-memory connection, not the processing task per se.

Experiment 3

Both previous experiments show that participants take reliably longer to recall words semantically linked to the sentences they accompany. However, the accuracy advantage for integrated words was significant only in Experiment 1. Therefore, we collected data from an additional experiment comparing reading span for integrated and independent words. This allows us to replicate the recall timing effect and assess further the issue of recall accuracy. The number of trials at each sequence length was also larger and used a different corpus of sentences.

Method

Participants

Thirty-three Lancaster University students (27 women and 6 men) formed a subset of a larger experiment. They were paid £4 to complete the reading span trials and additional tasks (not described here), and were randomly assigned to the integrated and independent condition.

Apparatus and Procedure

The same apparatus as in Experiment 2 was employed. The core stimulus pool comprised 90 sentences from the Friedman & Miyake (2004b) corpus. This was divided into two subsets (A & B) as before. Participants completed either the (a) integrated or (b) independent reading span task. (a) Each sentence appeared in black type with the final to-be-remembered word (integrated with sentence meaning) displayed in purple. There was a 0.5 sec ISI. (b) The entire sentence appeared in black type with the independent memory word that followed it shown in purple type for 0.5 sec. Participants were administered a total of 15 trials; 5 for each list length 2-4, in ascending sequence length order. Verbal instructions were the same as Experiment 2.

Results

Sentence reading times for integrated and independent conditions ($\underline{M}$=5.22, $\underline{SD}$=.69 and $\underline{M}$=4.86, $\underline{SD}$=.56 respectively) did not differ significantly, $\underline{t}(29)=1.58$, $\underline{p}=.125$, $\eta^2=.079$. The sentences were longer than those used in previous experiments, leading to extended reading times, but the pattern of performance is the same. In terms of the number of correctly recalled words, shown in Table 1, the integrated format again enjoyed an advantage, but in this dataset it was not significant, $\underline{t}(31)=1.82$, $\underline{p}=.078$, $\eta^2=.097$.

To increase statistical power we combined accuracy scores with data from Experiment 2. The trial structure changed across experiment with more shorter sequence length trials in Experiment 3, and indeed the proportion of words correctly recalled was higher in Experiment 3, $\underline{F}(1,57)=13.4$, $\underline{p}=.001$, $\eta_p^2=.191$. Nonetheless it is the difference between sentence formats that is relevant here, and indeed more words were recalled from the integrated format, $\underline{F}(1,57)=5.13$, $\underline{p}=.027$, $\eta_p^2=.083$. There was no interaction between experiment and task format, $\underline{F}<1$, $\underline{p}=.747$, $\eta_p^2=.002$. Thus, the integrated format does lead to greater levels of recall, even without a word generation requirement in processing, though the effect is significant only with data aggregated across the two studies.

One judge (who contributed to measurements in Experiment 2) extracted timing measurements from auditory recordings of correct reading span sequence, and outliers were screened as before. Recall times are illustrated in Figure 3. At list length 2, the pauses between integrated words were significantly longer than between independent words, $t(31)=2.88$, $p=.010$, $\eta^2=.211$. Moreover, the word duration was significantly longer in the integrated condition, $t(31)=2.65$, $p=.012$, $\eta^2=.185$. At list length 3, the pause was also significantly longer in the integrated condition, $t(24)=2.37$, $p=.029$, $\eta^2=.190$ while the word durations did not differ, $t(24)=.079$, $p=.938$, $\eta^2<.001$. The difference in the interword pause averaged across list lengths 2 and 3 was significant, $t(24)=2.68$, $p=.017$, $\eta^2=.230$.

Preparatory intervals did not differ between the independent and integrated conditions at list length 2, $t(31)=.81$, $p=.425$, $\eta^2=.021$, but were longer in the independent condition for list length 3, $t(24)=2.19$, $p=.038$, $\eta^2=.167$. Although participants were asked to read aloud the independent word immediately, some nonetheless articulated this item whilst the recall cue occurred, and consequently their ‘preparatory interval’ included additional reading. When we adjusted for the ‘over-running’ of the read word, the silent preparatory interval for the independent condition became shorter than for the integrated condition at list length 2 (*M*s=.46 vs. .82), $t(31)=4.21$, $p<.001$, $\eta^2=.364$, and there was no longer any difference at list length 3 (*M*s=.88 vs. .96), $t(24)=.44$, $p=.664$, $\eta^2=.008$.

Discussion

This study confirms and extends the results from Experiments 1 and 2. We again found that participants produced consistently longer pauses between recall words when these words were semantically related to the sentences that had been read, compared to when the words were unrelated to the sentences preceding them.

The difference in recall accuracy between the integrated and independent conditions was clearly largest in Experiment 1. Recall accuracy was roughly comparable in the integrated condition, regardless of whether participants either generated a sentence completion and final word. However, accuracy was relatively poor in Experiment 1 when participants generated a word to complete the sentence and then remembered a separate item, rather than when they read an entire sentence and remembered a separate item in Experiments 2 & 3. We conclude that the memory for the independent words must be fragile, and thus retention can be disrupted by competing representations such as a generated item. Yet this reinforces a central contention of the present paper: in reading span, participants arrive at recall with more than just the experimentally-defined memoranda in mind (in this case, a generated word not relevant to the memory set). 'Processing' and 'memory' may represent separate experimentally-defined phases of the working memory span trial, but memory is not a neatly segregated modular activity.

Experiment 4

We next undertook a conceptual replication of the preceding studies but rather than using spoken output, participants compiled a response sequence from a visually-presented set of choices with a touch-screen device. Such an automated response method in which the computer measures inter-response delays offers a potentially complementary source of evidence about recall timing, avoiding the requirement that participants assemble words into articulatory programs (see Chuah & Maybery, 1999; Maybery, Parmentier & Jones, 2002, for timing data involving spatial stimuli).

Participants are shown a number of candidate words and they attempt to select the designated memoranda in the correct serial order, whilst avoiding incorrect words, yielding a set of inter-response intervals rather than separate word and pause times. There were two types of incorrect lures; (a) words from the processing sentence, since it is known that participants sometimes verbally recall nontarget sentence material (Chiappe, Hasher & Segal, 2000; Friedman & Miyake, 2004b); (b) Target words from the preceding trial. Thus, the participant must overcome the impact of proactive interference from earlier trials (Lustig, Hasher & May, 2001), or at least make source-information judgements about current and past memories (Hedden & Park, 2003).

The experimental procedure returns to the 'read-and-complete' sentence processing procedure used in Experiment 1. However, the response format differs in a crucial respect. The recall choice display comprises correct answers, sentence words and previous trial words. Thus, for the independent condition the self-generated sentence terminal word is not presented. The contrast between Experiment 1 and 2/3 shows that this nontarget word impairs the accuracy of recall in the independent word condition. Consequently, this experiment offers an illuminating complement to and extension of the preceding studies. It addresses the question of whether the generated item disrupts recall even when it is not available as a response choice.

Participants were administered two assessments of reading span. Responses either disappeared from the recall screen when they were selected or remained on screen. This manipulation addresses the contribution of visual screen and search complexity to recall performance. Recall word repetition is unusual in spoken recall sequences, and hence removing selected answers allows participants to focus on remaining answers. However, this affects screen complexity – and potentially response selection time - as recall proceeds. Therefore a condition in which answers remain available provides an important comparison.

Method

Participants

Twenty-seven University of Missouri students (16 women and 11 men) participated for partial fulfilment of course credit requirements. One participant was subsequently excluded since they consciously ignored serial recall instructions.

Stimuli and Apparatus

Computer events were driven by an Apple Macintosh ibook (using “Revolution” software under OS X) with response latencies measured in (1/60 s) ticks. A Liyama touchscreen monitor (model INTH380-BS plus Keyspan RS232-USB Adaptor) displayed the experimental screen and recorded participants’ responses. Some sentences were adjusted for idiomatic phrases since we sampled from North American participants (whereas English students are familiar with eating “fish and chips”, American students are more familiar with a reference to “hamburger and fries”). We added to the set of memory word stimuli such that memoranda could be selected without replacement.

Procedure

Except in the following respects, the procedure for the delivery of processing and memory followed Experiment 1. Trials at all list lengths (i.e. 2 – 5) were administered Independent memoranda were presented for .75 sec with .125 sec ISIs (maintaining duration equivalence with integrated words). If a participant produced a non-expected completion (suggesting "Food and water makes plants" ..."live" instead of "grow") the experimenter would identify the target word (say "*or grow*") before proceeding to the next experimental event[3]. Participants completed both a 'remain' and 'disappear' response selection condition, which were presented in counterbalanced order, with a minimal break between each.

Recall

Upon completion of the sentence processing phase, the computer presented a response choice screen that comprised the target memoranda as well as incorrect words. For each correct choice, there was also a 'processing-phase' lure – a word sharing semantic content that had appeared in the sentence (or very occasionally, when no suitable candidate was available, an associated prime word). There were also two 'protrusion' lures – correct answers from the previous trial (the first trial employed two randomly selected words). Each of these ((list length * 2) + 2) candidate answers was assigned at random to one of 16 pre-specified screen locations (arranged in an underlying pattern of four rows and columns). Participants selected their responses by tapping the relevant locations on-screen in the appropriate order. The computer recorded recall selections and latencies.

[3] This was necessary because recall items here were fixed prior to presentation, although it was used very rarely.

In the 'disappear' condition, the chosen response was removed from the screen upon selection. In the 'remain' condition, responses continued to be visible after they had been chosen. In either case there was an auditory signal to confirm the computer's registration of the response selection. Participants were informed about the recall configuration at the start of the condition.

Results

Inspection of Table 1 indicates that participants again recalled more integrated words correctly (in both the disappear and remain condition) than independent words. Analysis of variance confirmed a significant memory advantage for integrated words, $\underline{F}(1,24)=16.2$, $\underline{p}=.001$, $\eta_p^2=.402$, but no difference between the screen formats, $\underline{F}<1$, $\eta_p^2<.001$, and no interaction, $\underline{F}<1$, $\eta_p^2=.004$.

We did not anticipate reading time differences between the disappear and remain condition, since they differ only in screen dynamics, and they were comparable ($\underline{M}=3.50$ vs. $\underline{M}=3.60$ respectively), $\underline{t}(25)=.75$, $\eta^2=.022$. There were also no reliable reading time differences for integrated and independent words, for either the disappear or remain conditions ($\underline{t}s(24)<.97$, $\eta^2<.038$).

Selection times for correct sequences were screened for outliers as previously. Figure 4 describes pauses for each list length and response format. Graphs indicate list position effects – a speeding up in selection through the list – in *both* the disappear and remain conditions. Analysis revealed significantly quicker selections in the disappear condition at list length 3 only and no interactions. We therefore simplify results by collapsing across this variable.

The average inter-item response durations for each list length were broadly consistent with previous experiments. Although list length 2 and 3 differences were not significant [t(24)=.26, p=.796, η^2=.003 and t(22)=.31, p=.756, η^2=.004 respectively], at list length 4 the integrated word response pauses were significantly longer than independent response pauses, t(20)=2.56, p=.029, η^2=.247. There was also a difference at list length 5 but this was marginal, t(14)=1.82, p=.090, η^2=.191. We then combined the correct recall times for all available list lengths, and this confirmed the longer pauses in the integrated condition, t(24)=2.74, p=.021, η^2=.238. In the round, the independent words are selected more quickly, but with a response choice paradigm this is most salient at longer list lengths.

Analysis of selection errors

The task design places clear constraints on the nature of selection errors. Serial order errors could occur but participants could also choose an incorrect word. However, opportunities for the different types of error are not constant across trials; at list length 2 for example, there is only one order error possible (the string A-B recalled as B-A) while at list length 3, there are 5 order error permutations. Furthermore, the number of protrusion error lures was a constant two items across list length (necessary to minimise visual screen complexity) and thus protrusions are less likely to occur through random selection at larger list lengths. For clarity and brevity, we present analysis of errors after combining data for the disappear and remain conditions.

In what follows we consider data based on performance up to and including the span-terminal level[4]. Table 3 reports the distribution of response choices. Errors are not randomly distributed: there are more order errors than processing-

[4] In the 'remain' condition, it is possible to produce a 'repeat' error; many of these reflected registration issues for touch screen responses (e.g. immediate repetitions with an interval <0.5 sec). For simplicity, they were coded here as order errors.

phase errors in both the independent condition, $t(12)=3.60$, $p=.004$, $\eta^2=.519$, and integrated condition, $t(12)=5.50$, $p<.001$, $\eta^2=.716$. Furthermore, although there were more processing-phase lures than protrusion lures, error proportions for these two categories were not significantly different, either for the independent or integrated conditions, $ts(12)=.48$ & $-.03$ respectively, $\eta^2<.019$.

Error frequency varied across experimental condition. The proportion of protrusion errors was higher with independent words than integrated words, $t(24)=2.63$, $p=.015$, $\eta^2=.224$. The proportion of all selections that were order errors was marginally higher with independent words, $t(24)=1.93$, $p=.069$, $\eta^2=.134$, but the proportion of order errors expressed as a function of correct item information was significantly higher for independent words (M=.251, SD=.136) than integrated words (M=.146, SD=.083), $t(24)=2.35$, $p=.027$, $\eta^2=.187$[5].

In the independent condition there is no link between the processing material and the target memory word and so there is nothing to bind the processing episode to word activation levels. Since answers to previous trials are likely to retain activation, they become more susceptible to being chosen. In contrast, for the integrated condition the processing context may help rule out these protrusion lures, making source-monitoring decisions more accurate. In other words, an important function of the integration between processing and memory is to tie memoranda more distinctly specific trial episodes.

Finally, we note that participants were tempted by the presence of processing-phase lures. All 13 participants in the independent condition selected this lure type at least once, as did 12 / 13 participants in the integrated condition.

[5] We are grateful to an anonymous reviewer for suggesting this comparison.

Discussion

This experiment addresses several issues. First, it provides output timing data from a complex memory task using touchscreen responses rather than spoken recall. The latter has been highly important in increasing our understanding of memory phenomena (e.g., Cowan et al., 1992; Haberlandt, Lawrence, Krohn, Bowe & Thomas, 2005; Tehan & Lalor, 2000) but of course it is possible that some phenomena are properties of the specific methods of responding. Spoken recall generally requires item and order information, yet potentially these dimensions can be systematically manipulated in the current paradigm by varying the selection choices available. From a pragmatic standpoint, measuring spoken recall is a highly labour intensive process. Consequently, it is valuable to additional evidence available from different, more easily registered methods of output.

In these terms, the experiment has been a success. Response processes are consistent and systematic, with individuals prone to confusing the sequential order of items and substituting no-longer-relevant words and words never explicitly designated as memory items. These results are consistent with reports from verbal sequence recall (Caretti, Cornoldi, De Beni & Romano 2005; Friedman & Miyake, 2004b) confirming both item and order constraints in WM.

General Discussion

Four experiments involving either spoken recall or manual item selection provide support for the recall reconstruction hypothesis. This posits that the sequence of reading span items may not be fully specified at the point when they are cued; instead, access to memoranda embedded within sentence processing may be subsequently scaffolded by those sentences, especially when task requirements stress the memory word identity (via a word generation requirement). The experiments present convergent and complementary evidence in the form of recall timing together with information about recall accuracy and error types. Data consistently suggest that the correct recall of words connected with processing events is slower than the correct recall of words unrelated to processing, even though recall accuracy for the integrated format is higher (which, other things being equal, should lead to faster recall).

Across studies reading times were comparable for integrated and independent word conditions and sentence completions were invariably suitable, strongly implying that sentences were processed appropriately in both conditions. We therefore suggest that a sentence context can support access to relevant words that might otherwise not be recalled, albeit at the cost of larger pauses. At the same time, recall time differences are not significant at every list length and some accuracy differences emerge only with a larger sample size, emphasising that (a) recall timing can be variable and recall reconstruction may not be required on every trial; and (b) participants may be able to draw on other strategies to retain the memoranda that can in some situations be effective. One possibility is that participants rely on an imagery-focused strategy (Caretti, Borella & De Beni, 2007). Alternatively, there may be an increased reliance on active sequence maintenance or primary memory, an interpretation that is consistent with the more rapid recall that we have observed.

The response choice method supports the basic finding that pauses are shorter in the independent condition although significant effects were not found at all sequence length. We suspect that there is a less stringent requirement in this paradigm to maintain the fidelity of item (as opposed to order) information, because the correct items are always present at recall. Item representations must be retained since various types of lures are used. Nonetheless, perhaps imperfect representations are sufficient for successful reconstruction.

Conway et al. (2005) have noted in passing that reading span studies have involved both the independent and integrated word format. We are not making claims about which is the theoretically preferred task form, especially in the context of individual differences since one might wish to minimise or emphasise the reconstructive element in recall. Nonetheless, is it clear each task configuration can have consequences for what reading span measures.

Several theoretical views consider complex spans' processing-memory relationship in competitive terms, even where inherently they may not need to do so. According to Daneman and Carpenter (1980), the processing activity within a working memory span trial uses general cognitive resources that are consequently denied to retention activities. Span therefore reflects the competitive balance of resources between processing and memory. The present data show instead that recall is partly a function of the compatibility between processing and memory, and that processing activity produces representations that affect recall, providing a source of both recall facilitation and interference. Processing and retention in reading span are therefore not as functionally distinct as considered hitherto.

The recall reconstruction hypothesis – proposing that longer pauses in the recall of information from working memory derive from memory search through the processing episodes – has resonances with conclusions from conceptual span

(Haarmann, Davelaar & Usher, 2003), which involves partial recall of sequences via category cues. Haarmann et al. (2003) report overlap between reading span and conceptual span, even though the latter does not require the conventional 'processing plus memory' combination. The present data encourage the view that both these paradigms overlap in terms of recall processes and both reflect a link between encoding context and recall.

The recall reconstruction hypothesis for reading span also has some parallels with the 'regeneration' account of sentence recall (Potter & Lombardi, 1990; 1998; Schweppe & Rummer, 2007). When individuals recall sentences verbatim, they are prone to substitute semantically compatible words presented prior to recall, and make subtle grammatical changes to the sentence structure when primed by other sentences. The regeneration account proposes that these errors occur because participants remember general semantic features and relations, with sentences regenerated from partial traces. The current proposal is that with an integrated reading span trial, individuals may likewise use sentence representations to augment recall. In a related vein, Schweickert (1993) has argued for 'redintegrative' processes in immediate serial recall, whereby gaps in incomplete representations of words can be filled in by lexical knowledge. In sum, whilst the present recall reconstruction hypothesis is a novel account of complex span performance, it echoes and indeed elaborates other approaches in the literature that emphasise the potential breadth of sources for making memory decisions and the constructive nature of temporary memory (e.g. Coane, McBride, Raulerson & Jordan, 2007; Craik & Lockhart, 1972).

We conclude that reconstructive processes can operate on what are often degraded representations during reading span recall. These processes may include memory search and decision-making drawing upon 'contextual' information from a variety of domains or content (e.g., Unsworth & Engle, 2006). Moreover, the present data show that WM theories typically adopt too simplistic an approach to the relationship between mental operations and information retention. Whereas most theories propose some form of competitive relationship, our data show that processing operations leave their own memory traces that can influence span performance. In producing evidence for this view, the paper emphasises the value of recall timing as a way of understanding immediate memory processes.

References

Baddeley, A. D., & Hitch, G. J. (1974). Working Memory. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.

Barrouillet, P., Bernadin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory span. *Journal of Experimental Psychology: General, 133*(1), 83-100.

Caretti, B., Borella, E., & Den Beni, R. (2007). Does Strategic Memory Training Improve the Working Memory Performance of Younger and Older Adults? *Experimental Psychology, 54*(4), 311-320.

Caretti, B., Cornoldi, C., De Beni, R., & Romano, M. (2005). Updating in working memory: A comparison of good and poor comprehenders. *Journal of Experimental Child Psychology, 91*(1), 45-66.

Case, R. (1985). *Intellectual development: birth to adulthood*. New York, Academic Press.

Chiappe, P., Hasher, L., & Siegel, L. A. (2000). Working memory, inhibitory control and reading ability. *Memory & Cognition, 28*(1), 8-17.

Chuah, Y.M.L., & Maybery, M.T. (1999). Verbal and spatial short-term memory: Two sources of developmental evidence consistent with common underlying processes.

International Journal of Psychology, *34*, 374-377.

Coane, J. H, McBride, D. M., Raulerson, B. A., & Jordan, J. S. (2007). False memory in a short-term memory task. *Experimental Psychology*, *54*, 62-70.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.

Copeland, D. E., & Radvansky, G. A. (2001). Phonological similarity in working memory. *Memory & Cognition, 29*(5), 774-776.

Cowan, N., Day, L., Saults, J. S., Keller, T. A., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language, 31*, 1-17.

Cowan, N., Towse, J. N., Hamilton, Z., Saults, J. S., Elliott, E. M., Lacey, J. F., et al. (2003). Children's working memory processes: A response-timing analysis. *Journal of Experimental Psychology: General, 132*(1), 113-132.

Cowan, N., Wood, N. L., Wood, P. K., Keller, T. A., Nugent, L. D., & Keller, C. V. (1998). Two separate verbal processing rates contributing to short-term memory span. *Journal of Experimental Psychology: General, 127*(2), 141-160.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal Verbal Learning and Verbal Behavior, 11*, 671-684.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*, 422-433.

Dosher, B. A., & Ma, J. (1998). Output loss or rehearsal loop? Output-time vs. pronunciation time limits in immediate recall for forgetting matched materials. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*(2), 316-335.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory* (pp. 102-134). New York: Cambridge University Press.

Friedman, N. P., & Miyake, A. (2004a). The reading span test and its predictive power for reading comprehension. *Journal of Memory and Language, 31*, 136-158.

Friedman, N. P., & Miyake, A. (2004b). The relations among inhibition and interference control functions: A latent variable analysis. *Journal of Experimental Psychology: General, 133*, 101-135.

Haberlandt, K., Lawrence, H., Krohn, T., Bowe, K., & Thomas, J. G. (2005). Pauses and durations exhibit a serial position effect. *Psychonomic Bulletin & Review, 12*(1), 152-158.

Haarmann, H. J., Davelaar, E. J., & Usher, M. (2003). Individual differences in semantic short-term memory capacity and reading comprehension. *Journal of Memory and Language, 48*, 320-345.

Hedden, T., & Park, D. (2003). Contributions of source and inhibitory mechanisms to age-related retroactive interference in verbal working memory. *Journal of Experimental Psychology: General, 132*(1), 93-112.

Hitch, G. J., Towse, J. N., & Hutton, U. M. Z. (2001). What limits working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General, 130*(2), 184-198.

Jarrold, C., & Bayliss, D. (2007). Variation in working memory due to typical and atypical development. In A. R. A. Conway, C. Jarrold, M. Kane, A. Miyake & J. N. Towse (Eds.), *Variation in working memory* (pp. 134-161). New York: Oxford University Press.

Kane, M. J., & Engle, R. W. (2003). Working memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop Interference. *Journal of Experimental Psychology: General, 132*(1), 47-70.

Leather, C. V., & Henry, L. A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. *Journal of Experimental Child Psychology, 58*, 88-111.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General, 130*(2), 199-207.

Maybery, M., Parmentier, F. B. R., & Jones, D. M. (2002). Grouping of list items reflected in the timing of recall: implications for models of serial verbal memory. *Journal of Memory and Language, 47*, 360-385.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory*. New York: Cambridge University Press.

Mulligan, N. (2004). Generation and memory for contextual detail. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*(4), 838-855.

Nairne, J. S., Riegler, G. L., & Serra, M. (1991). Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17*(4), 702-709.

Osaka, M., Nishizaki, Y., Komori, M., & Osaka, N. (2002). Effect of focus on verbal working memory: Critical role of the focus word in reading. *Memory & Cognition, 30*(4), 562-571.

Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language, 29*, 633-654.

Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language, 38*, 265-282.

Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing-storage relationship in reading span performance. *Journal of Memory and Language, 50*, 425-443.

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, *21*, 168-175.

Schweppe, J., & Rummer, R. (2007). Shared representations in language processing and verbal short-term memory: The case of grammatical gender. *Journal of Memory and Language, 56*, 336-356.

Slamecka, N. J., & Graf, P. (1978). The generation effect; Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory, 4*, 592-604.

Steffens, M. C., & Erdfelder, E. (1998). Determinants of positive and negative generation effects in free recall. *Quarterly Journal of Experimental Psychology, 51A*(4), 705-733.

Tehan, G., & Lalor, D. M. (2000). Individual differences in memory span: The contribution of rehearsal, access to lexical memory, and output speed. *Quarterly Journal of Experimental Psychology, 53A*(4), 1012-1038.

Towse, J. N., & Cowan, N. (2005). Working memory and its relevance for cognitive development. In W. Schneider, R. Schumann-Hengsteler & B. Sodian (Eds.), *Young Children's Cognitive Development: Interrelationships among Executive Functioning, Working Memory, Verbal Ability, and Theory of Mind* (pp. 9-37). Mahwah, NJ:

Lawrence Erlbaum Associates.

Towse, J. N., Hamilton, Z., Hitch, G. J., & Hutton, U. (2000). *Sentence completion norms among adults: A corpus of sentences differing in length* (Technical Report No. CDRG 7): Royal Holloway, University of London.

Towse, J. N., Hitch, G. J., & Hutton, U. (1998). A reevaluation of working memory capacity in children. *Journal of Memory and Language, 39*(2), 195-217.

Towse, J. N., Hitch, G. J., & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition, 28*(3), 341-348.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127-154.

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*, 68-80.

Appendix 1: Additional analysis of data from Experiment 1

As is evident from Table 1, participants achieved good levels of accuracy on the STM task; 7 participants remembered all 42 words while everyone remembered at least half of the words. This high level of performance is advantageous in maximising the data density for analysis of *correct* recall.

Several recall phenomena are evident from Figure 5. First, mean word articulation at approximately .4 sec, is similar to reading span data. Second, pauses between words are much shorter, but non-negligible, at less than .2 sec. Third, the preparatory intervals are approximately three times longer than interword pauses, consistent with the interpretation that they reflect a different mental activity. Data contrast dramatically with reading span recall that is much longer (note a different scale is used across Figures), and there are large, almost qualitative changes in the pattern of responding across list lengths.

Table 1. Memory performance as the proportion of words correctly recalled in the appropriate serial position. Standard deviations in parentheses. Short-term memory (STM) trials in Experiment 1 differ only with respect to the working memory (WM) task that was also completed.

	Expt.1(STM)	Expt.1(WM)	Expt.2	Expt.3	Expt.4(dis.)	Expt.4(rmn..)
Integrated WM	.913 (.117)	.647 (.214)	.616 (.141)	.695 (.101)	.789 (.152)	.802 (.122)
Independent WM	.901 (.079)	.300 (.109)	.520 (.200)	.624 (.120)	.582 (.176)	.571 (.208)

Table 2. Stability of recall timing: correlations between recall durations of list-length 2 & 3 in Experiment 1, 2, and 3 (asterisks represent significant correlations, at least p<.05).

	Expt.1 STM	Expt.1 WM	Expt.2 WM	Expt.3 WM	Expt.4 WM
Preparatory Intervals	$r(22)=.661$*	$r(14)=.062$,	$r(24)=.535$*	$r(24)=.137$	
Interword Pauses	$r(22)=.571$*	$r(14)=.168$,	$r(24)=.484$*	$r(24)=.609$*	
Word durations	$r(22)=.766$*	$r(14)=.694$*	$r(24)=.626$*	$r(24)=.051$	
Recall intervals					$r(22)=.465$*

Table 3. Proportion of recall choices falling into different response categories in Experiment 4, for all trials up to and including span length. Standard deviations in parentheses.

	Response choice			
	Correct	Order error	Proc.-phase error	Protrusion error
Integrated	.802 (.112)	.134 (.070)	.032 (.021)	.032 (.045)
Independent	.626 (.118)	.212 (.129)	.076 (.034)	.086 (.058)

Figure legends

Figure 1. Duration and standard error bars of correct sequences for reading span trials in Experiment 1, as a function of the phase of recall. PI = preparatory interval. Words = average duration of recalled items. Pause= interword pause duration (averaged at list length 3).

Figure 2. Duration and standard error bars of correct sequences for reading span in Experiment 2, as a function of the phase of recall. PI = preparatory interval. Words = average duration of recalled items. Pause= interword pause duration (averaged at list length 3).

Figure 3. Duration and standard error bars of correct sequences for reading span in Experiment 3, as a function of the phase of recall. PI = preparatory interval. Words = average duration of recalled items. Pause= interword pause duration (averaged at list length 3).

Figure 4. Duration of recall delays between the correct selection of responses in Experiment 4, as a function of output position. Graph includes standard error bars. Top panel=data from integrated word condition (when response choices disappear after selection or remain after selection on the left and right respectively). Bottom panel=data from independent word condition (when response choices disappear after selection or remain after selection on the left and right respectively).

Figure 5. Duration and standard error bars of correct sequences for word span trials in Experiment 1, as a function of the phase of recall and the sequence length. PI = preparatory interval. Words = average duration of recalled items. Pause= interword pause duration (averaged at list length 3 and beyond).

Figure 1.

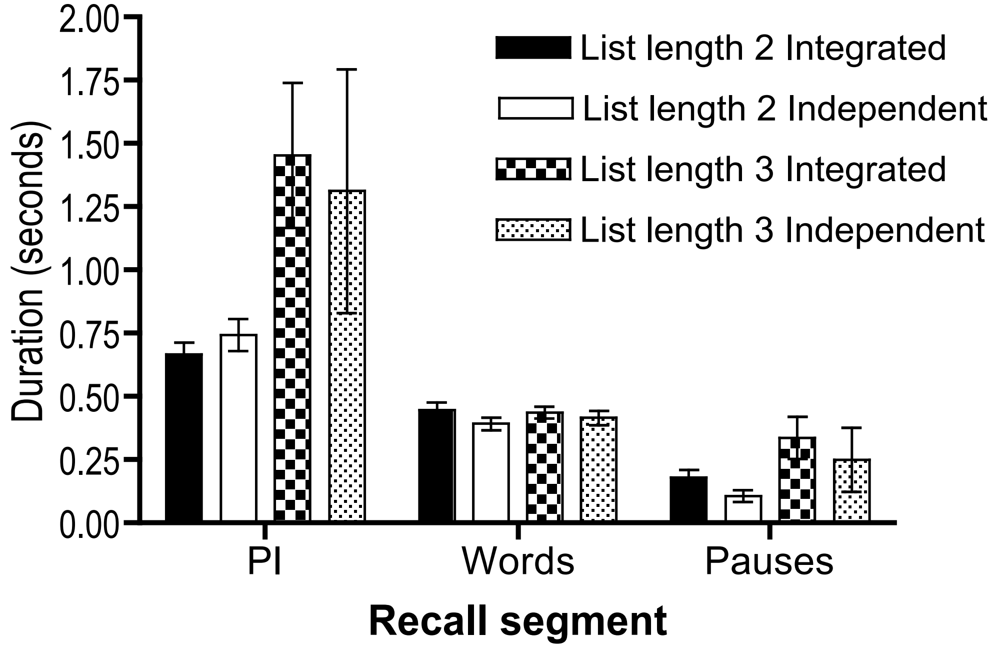

2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Duration (seconds)
List length 2 Integrated
List length 2 Independent
List length 3 Integrated
List length 3 Independent
PI
Words
Pauses
Recall segment

Figure 2.

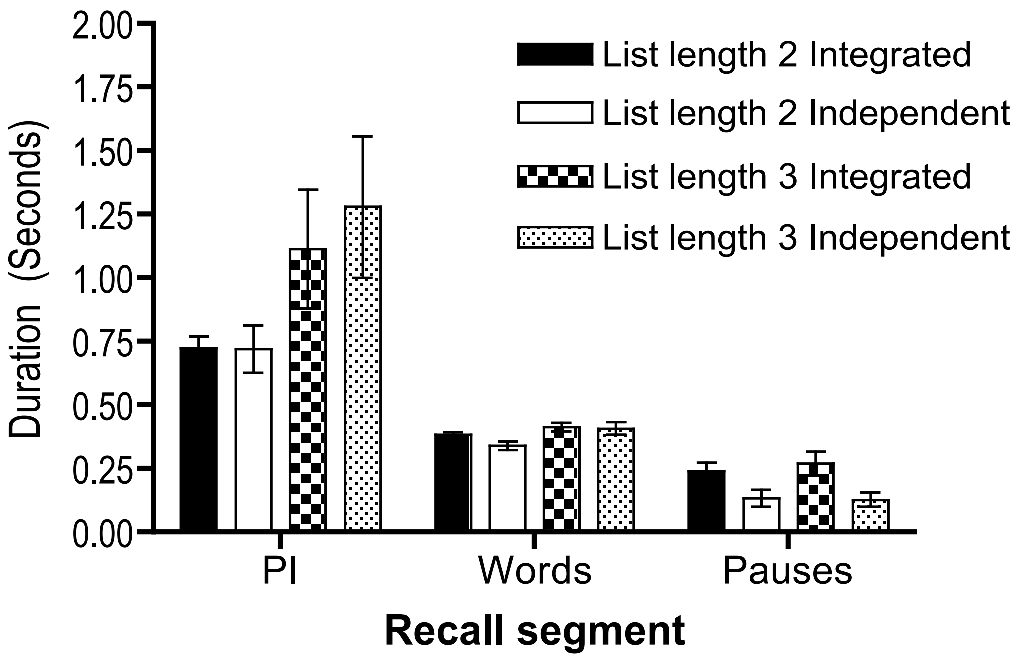

2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Duration (Seconds)
List length 2 Integrated
List length 2 Independent
List length 3 Integrated
List length 3 Independent
PI
Words
Pauses
Recall segment

Figure 3.

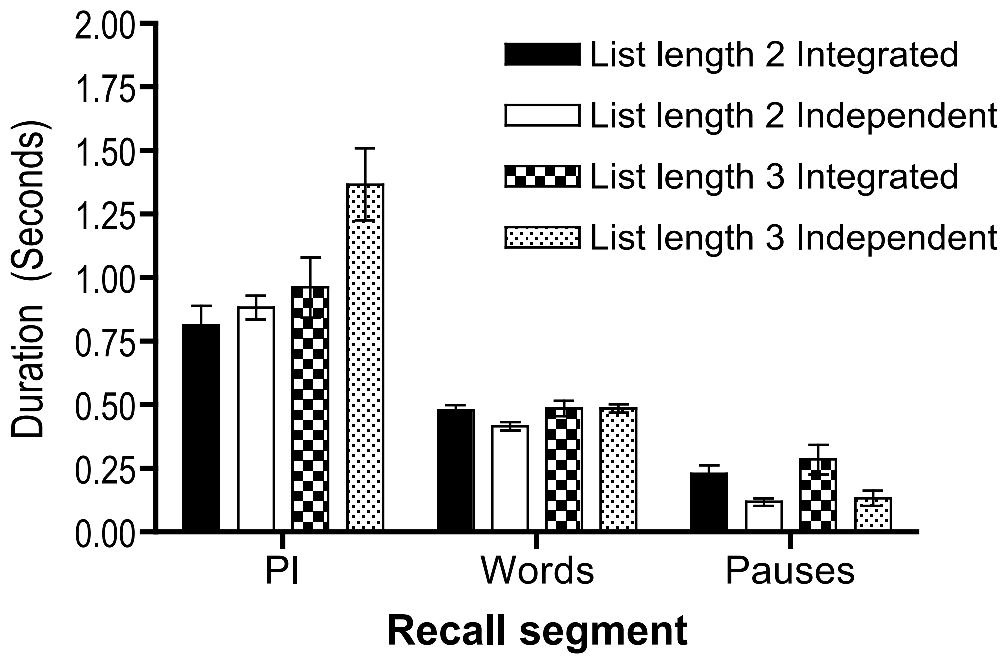

List length 2 Integrated
List length 2 Independent
List length 3 Integrated
List length 3 Independent
Duration (Seconds)
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
PI
Words
Pauses
Recall segment

Figure 4.

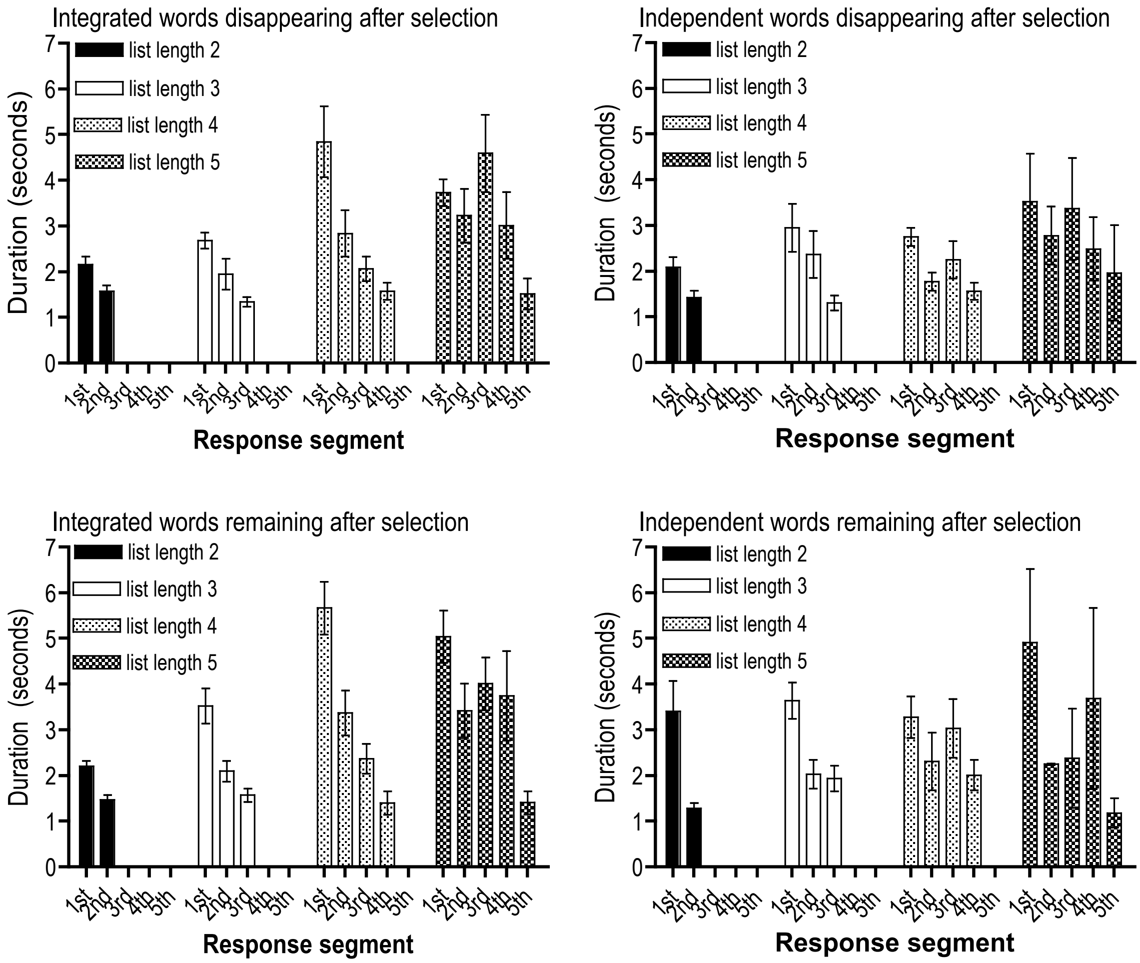
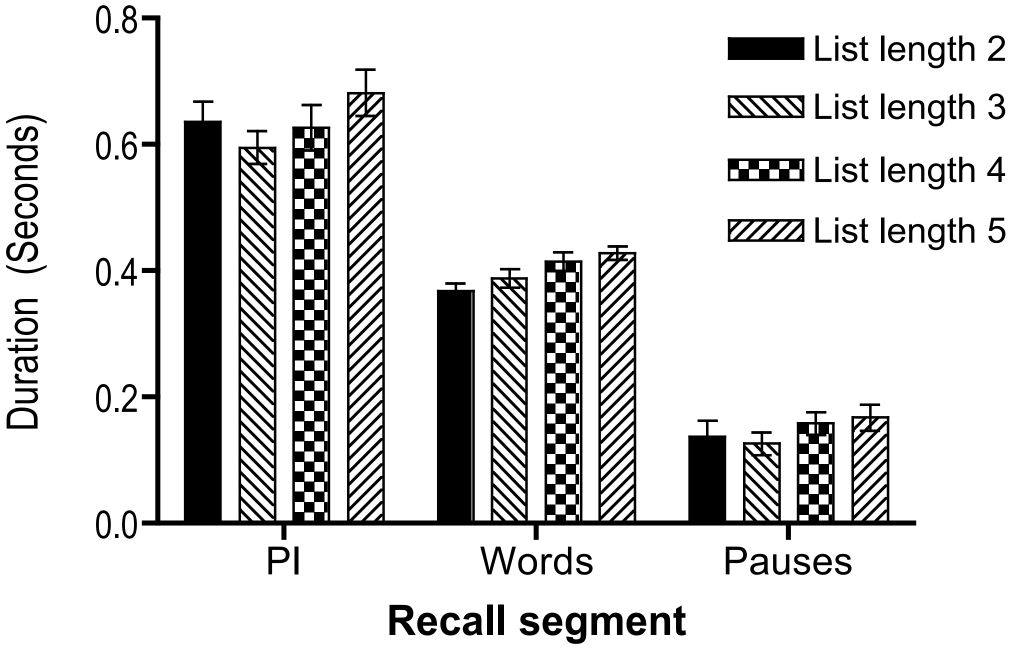

Integrated words disappearing after selection
Independent words disappearing after selection
Integrated words remaining after selection
Independent words remaining after selection
list length 2
list length 3
list length 4
list length 5
Duration (seconds)
Response segment
1st
2nd
3rd
4th
5th
0
1
2
3
4
5
6
7

Figure 5.


0.8
0.6
0.4
0.2
0.0
Duration (Seconds)
List length 2
List length 3
List length 4
List length 5
PI
Words
Pauses
Recall segment