



Automatic and controlled processing in sentence recall: The role of long-term and working memory

Elizabeth Jefferies^{a,*}, Matthew A. Lambon Ralph^a, Alan D. Baddeley^b

^a *Department of Psychology, University of Manchester, UK*

^b *Department of Psychology, University of York, UK*

Received 5 May 2004; revision received 8 July 2004

Available online 2 September 2004

Abstract

Immediate serial recall is better for sentences than word lists presumably because of the additional support that meaningful material receives from long-term memory. This may occur automatically, without the involvement of attention, or may require additional attentionally demanding processing. For example, the episodic buffer model (Baddeley, 2000) proposes that the executive component of working memory plays a crucial role in the formation of links between different representational formats and previously unrelated concepts. This controlled integrative encoding may be more important in sentence than word recall. Three experiments examined the effect of an attention-demanding concurrent visual choice reaction time task on the recall of auditorily presented stories, sentences, and lists of unrelated words, in order to investigate the relative importance of automatic and controlled processing for these materials. The concurrent task was found to disrupt the recall of strings of unrelated sentences more than random word lists, suggesting that controlled processing played a greater role in the sentence recall task. On subsequent learning trials, however, recall of the unrelated words was also disrupted by the concurrent task, possibly due to the development of chunking. The large dual task decrement for unrelated sentences did not generalise to the recall of more naturalistic prose, suggesting that the requirement to integrate phonological with long-term linguistic information is not attentionally demanding per se; instead, the integration of unrelated concepts is effortful once recall extends beyond the capacity of the phonological loop. Our results suggest that sentence recall reflects contributions from both automatic linguistic processes and attentionally limited working memory.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Working memory; Episodic buffer; Prose memory; Chunking; Sentence span

Historically, a useful distinction has been drawn between performance on immediate serial recall tasks such as digit span, and longer-term learning of material like prose. This distinction is reflected in the double dissociation shown by patients with impaired verbal short-term

memory (STM) and those with the classic amnesic syndrome. Amnesic patients are severely impaired on long-term prose recall tasks such as the logical memory test but have intact digit span (Baddeley & Warrington, 1970; Milner, 1966). Conversely, STM patients have greatly reduced digit span in the context of relatively intact delayed prose recall (Shallice & Warrington, 1970; Vallar & Shallice, 1990). In the original working memory (WM) model of Baddeley and Hitch (1974), span

* Corresponding author. Fax: +44 161 275 2588.

E-mail address: beth.jefferies@man.ac.uk (E. Jefferies).

for verbal materials was assumed to be limited by the time-based capacity of the phonological loop, operating independently from long-term memory (LTM). However, more recent research has emphasised the extent to which the two systems work together. Phonological STM is thought to play a crucial role in learning long-term phonological representations of new words (Baddeley, Gathercole, & Papagno, 1998; Baddeley, Papagno, & Vallar, 1988; Gathercole & Baddeley, 1989, 1990a, 1990b), and LTM can also contribute to performance on span tasks, with higher spans possible when the contribution from LTM is greater (e.g., Hulme, Maughan, & Brown, 1991). The difference between word and sentence span provides a particularly clear example of this: twice as many words can be recalled in the correct order when they form a meaningful sentence (Brener, 1940).

Different levels of long-term knowledge (sub-lexical, lexical, and super-lexical) contribute to performance on immediate serial recall tasks. At the sub-lexical level, word-like nonwords, containing sequences of phonemes that occur frequently in real words, are repeated more accurately than nonwords that are less word-like (Gathercole, Frankish, Pickering, & Peaker, 1999). At the lexical level, LTM representations of the sounds and meanings of familiar words make an important contribution to word span. Span is higher for words than for nonwords (Hulme et al., 1991), and higher for words that frequently occur in English compared with less frequent words (Hulme et al., 1997). In addition, span is greater for high compared with low imageability words (Bourassa & Besner, 1994), for concrete compared with abstract words (Walker & Hulme, 1999), and for words taken from the same semantic category (Poirier & Saint Aubin, 1995). Neuropsychological evidence also points to the role of lexical-level representations in span tasks (see Martin & Saffran, 1997; Martin & Lesch, 1996; Martin, Shelton, & Yaffee, 1994). Semantic dementia patients, who show a specific and progressive decline in semantic memory, are better at repeating words that they understand relatively well compared with words whose meanings have become degraded over the course of the disease (Jefferies, Jones, Bateman, & Lambon Ralph, *in press*; Jefferies, Jones, Bateman, & Lambon Ralph, 2004; Knott, Patterson, & Hodges, 1997; Patterson, Graham, & Hodges, 1994).

Sentence span is substantially greater than word span (Brener, 1940), reflecting the involvement of long-term representations above the lexical level in span tasks. Miller and Selfridge (1950) studied the recall of word sequences with different degrees of approximation to English prose. Higher order approximations were more meaningful and more grammatical, and resulted in better recall, suggesting that both syntactic and semantic factors may play a role in increasing sentence span above word span. Epstein (1961) found that grammatically marked se-

quences of nonsense syllables and unrelated words were easier to recall than sequences that were not tagged in this way. Similarly, Marks and Miller (1964) found that ‘anomalous sentences’ that were grammatically correct but had little overall meaning (e.g., ‘Noisy flashes emit careful floods’) were recalled more accurately than jumbled anomalous sentences with a random word order.

The semantic relationships between words can also influence immediate serial recall. Baddeley and Levy (1971) found a semantic similarity effect in the recall of meaningful noun–adjective pairs, so that sequences like ‘priest–moral, minister–religious, vicar–pious’ were recalled more poorly than sequences like ‘palace–magnificent, apple–delicious, rattlesnake–deadly.’ This semantic similarity effect disappeared when the nouns and adjectives were not semantically compatible, presumably because this caused the participants to fall back onto phonological coding. Finally, if participants are given information about the overall theme of a passage of prose in the form of a title or a picture, recall improves considerably (Bransford & Johnson, 1972; Dooling & Lachman, 1971).

A number of theories have been proposed to account for the difference between word and sentence span. They make the common assumption that sentence span is higher than word span because it derives additional support from long-term representations that are involved in sentence comprehension. Sentence recall is constrained not only by lexical-level representations but also by syntactic and semantic representations of the theme or overall meaning of the sentence. These syntactic and semantic factors increase the predictability of words in a sentence and may enhance memory in a process akin to chunking (Miller, 1956). In addition, Potter and Lombardi suggested that the surface structure of sentences could be reconstructed from long-term conceptual, lexical, and syntactic representations that were activated during the processing of the sentence (Lombardi & Potter, 1992; Potter & Lombardi, 1990, 1998).

While there is general agreement that semantic and syntactic constraints will aid sentence retrieval, there is less agreement about their role in encoding. One possibility is that sentences activate semantic and syntactic representations relatively automatically as they are processed and this activation enhances sentence retrieval consequentially. This type of support for sentences is envisaged by connectionist models that view verbal STM as arising directly from the representations that underpin language processing (MacDonald & Christiansen, 2002; Martin & Saffran, 1997). Alternatively, executive resources may play a necessary role in integrating phonological, lexical, syntactic, and semantic representations in sentence span tasks, enabling sentence span to extend far beyond the limited capacity of the phonological loop. Baddeley (2000), for example, proposed that immediate prose recall depends on the episodic buffer

component of WM: this is envisaged as an attentionally limited temporary store, capable of integrating information from multiple sources into a single episode. [Baddeley and Wilson \(2002\)](#) used the concept of an episodic buffer to explain marked individual differences in the immediate recall of prose by amnesic patients. A subset of patients who showed good immediate prose recall proved to have well preserved executive and intellectual capacity, suggesting to [Baddeley and Wilson](#) that they may have been able to utilise the attentionally limited episodic buffer system as a basis for immediate recall.

A similar distinction between automatic and attention-demanding LTM support for WM has been suggested by several other researchers. [Engle, Kane, and Tuholski \(1999\)](#) proposed that WM performance is accounted for by automatic activation of LTM and the capacity for “controlled attention.” These aspects of WM appear to interact: individuals with superior performance on complex span tasks, which make simultaneous storage and processing demands, show a greater benefit of domain-specific knowledge in their recall of baseball commentaries ([Hambrick & Engle, 2002](#)). This pattern is consistent with the notion that individuals with greater attentional capacity can more readily utilise information in LTM to support recall. Similarly, [Cowan's \(1995\)](#) “embedded processes” model of WM distinguishes between information in LTM, the activated portion of LTM, and the subset of this activation that is in the focus of attention. Attention is proposed to play a role in forming new links between disparate representations in LTM; therefore, attention-demanding binding may enable sentence span to extend beyond the recall of unrelated words by facilitating the integration of different levels of linguistic knowledge and/or the various propositions within a sentence. Connections may also be established between activated representations that are not within the focus of attention, allowing more automatic support for sentence span (see [Cowan, 1999](#)). These two theories make similar predictions to those of the episodic buffer model ([Baddeley, 2000](#)), without positing a separate integrative store.

Another relevant perspective is [Ericsson and Kintsch's \(1995\)](#) long-term working memory (LT-WM) theory. According to this viewpoint, there is a recall advantage for meaningful material that occurs because this information can be rapidly integrated with pre-existing structures in LTM. These remain readily accessible if aspects of the information are successfully maintained in STM. Consequently, LTM can act as an extension of WM (LT-WM). In some circumstances, the creation of these retrieval structures is likely to be attention demanding; for instance, with practice, subjects can dramatically improve their digit spans by associating groups of digits to pre-existing knowledge in a deliberate and apparently effortful way (e.g., encoding digits as running times for races: [Chase & Ericsson, 1981](#)). If the construction of mental models during sentence com-

prehension is analogous to this process, sentence span tasks should rely on attention. It should be noted, however, that [Ericsson and Kintsch](#) themselves did not focus on the role of attention and it is not clear if retrieval structures are always effortful to construct.

These theories might predict that normal participants should show a larger detrimental effect of divided attention on tasks that draw heavily on prior knowledge. [Britton and Tesser \(1982\)](#) observed this pattern in participants' reaction times on a secondary task. Knowledge about the primary task was manipulated by presenting additional information (for example, the previous page of text in a reading task) or by selecting novices and experts within a particular domain. Likewise, prose recall should be impaired to a greater extent than recall of an unstructured word list by an attentionally demanding secondary task during encoding. One study examined retention of a prose passage under dual-task conditions ([Baddeley & Hitch, 1974](#)). Memory for prose was found to be more impaired when participants had to remember sequences of six digits simultaneously, compared with a load of three digits, or with repeating single digits. Load, therefore, is an important factor in determining the degree to which memory for prose is disrupted by conditions of divided attention. However, the study did not compare prose with sequences of unrelated words, and was therefore not very informative about the nature of the processes disrupted by the secondary task. In addition, the secondary task required participants to say irrelevant material aloud, and would be expected to disrupt the functioning of the phonological loop as well as requiring participants to divide their attention between two tasks.

The present work examined the role of automatic and controlled processes in WM for stories, sentences, and words by means of a secondary task that was presumed to disrupt attention-demanding processes. We used the continuous choice reaction time (CRT) task employed by [Craik, Govoni, Naveh-Benjamin, and Anderson \(1996\)](#). Participants were required to press one of four keys corresponding to the position of a visual stimulus on the computer screen; a correct response caused the stimulus to move immediately to a new location. This method had a number of advantages. First, participants were required to divide their attention continuously between two tasks, placing a load on the executive component of WM. The secondary task did not involve verbal material or vocalisation, so should not have placed a load on the phonological loop component of WM. Second, [Craik et al.](#) showed that this task successfully disrupted memory performance. Third, reaction times provide a sensitive measure of the degree of disruption of the secondary task produced by the primary memory task.

The stories, sentences, and word lists were presented to each participant three times in successive trials allowing the role of learning to be examined. With multiple presentations, encoding tends to switch from phonolog-

ical to semantic, as reflected by both similarity effects in serial recall (Baddeley, 1966a) and the development of semantically based clustering in free recall (Tulving, 1962). It seems likely, therefore, that with repeated presentations, attention-demanding binding mechanisms will become more important in the recall of unrelated words (especially given that we used a fast presentation rate of two items per second, which would have discouraged the formation of associations between unrelated items on the first presentation). Consequently, any effect of divided attention on word recall was expected to increase across learning trials.

It was anticipated that recall would be substantially higher for sentences than word lists. If sequences of a single fixed length were studied, it might be difficult to avoid either ceiling effects in sentence recall or floor effects in word recall. Therefore, word and sentence span was established for each participant at the start of the testing session, and the length of material presented subsequently was set at 50% greater than span. This method helped to ensure that the word and sentence span tasks were of equivalent initial difficulty. The use of supra-span material made it possible to examine learning of the sentences and words over several trials.

Experiment 1 compared recall of strings of unrelated sentences and words under baseline control conditions and while participants performed an attention-demanding concurrent task. This methodology allowed a comparison of the relative importance of attention-demanding processes in word and sentence recall. In Experiment 2, this methodology was extended to look at different lengths of material, ranging from just above span to far beyond span. Beyond span, the amount of material to be recalled is assumed to exceed the capacity of the phonological loop. Performance is likely to be supplemented by the involvement of LTM. If the contribution from LTM is greater for sentences than words, participants may recall sentences better than word lists above span, despite the fact that sequence length was adjusted so as to match performance at span. Experiment 3 examined participants' recall of three types of verbal material under single and dual-task conditions: (1) unrelated words, (2) sequences of short unrelated sentences, and (3) related sentences that formed a story. The comparison of these two types of sentence material helped to establish exactly which aspects of the encoding of sentences required controlled processing.

Experiment 1

Method

Design

A repeated measures design was used. Recall was tested: (1) with or without a concurrent choice reaction

time (CRT) task, (2) with sentences or unrelated words as the material to be recalled, and (3) across three repeated presentations (allowing the effect of learning to be explored). The order of conditions was fully counter-balanced across subjects.

Participants

The participants were 24 psychology undergraduates, who took part for course credit. The participants in all three experiments were aged between 18 and 45, spoke English as a first language, and had no reported history of hearing or speech difficulties. Three subjects with exceptionally slow reaction times (more than 3.5 *SD* away from the mean) were excluded from the analysis and replaced, because of the risk that they were not performing both tasks simultaneously.

Materials

The primary task involved serial recall of unrelated words, and words that formed sentences. The number of words that participants were asked to recall depended on their span level, established at the start of the testing session (see Procedure below). Participants were presented with sentences and word lists that were 50% greater than their spans in order to reduce ceiling effects.

A total of 86 sentences were generated for this study. They were factual in content, contained between four and seven words and were designed to read naturally. The sentences used simple syntactic structures and avoided embedded clauses and conjunctions. Two, three or four unrelated sentences were strung together to produce longer sentence material (examples are provided in the Appendix). This method was preferred to using increasingly long single sentences, as the potential confound between grammatical complexity and sentence length was avoided. However, difficulty may not have increased smoothly with the number of words tested because lists containing a larger number of sentences may have been harder to recall. As the sentences comprising each list were unrelated in meaning, sentence spans may be lower than those reported elsewhere in the literature. Sentences were not repeated during the course of the experiment.

The lists of unrelated words were generated from the sentences (examples are provided in the Appendix). As word span is generally smaller than sentence span, only a subset of the sentence words were included in the word lists. The words used in the sentences were divided into four categories: (1) nouns, (2) verbs, (3) adjectives and adverbs, and (4) function words, according to the role each word played in the sentence. Estimates of the frequency of each word in written English were obtained from Kucera and Francis (1967) using the MRC psycholinguistic database (Coltheart, 1981). The four word types were divided into five frequency categories, and the unrelated word lists were formed by drawing words

pseudo-randomly from each of these categories so that the overall proportion of words in each category was the same for the sentences and word lists. Word lists that contained related words or multiple copies of a single word were discarded. As this method matched the unrelated word lists to the entire set of sentences, there may have been some unsystematic differences in word characteristics between the conditions for individual participants, given that the materials presented depended on span.

Two parallel versions of the test material were created for each span level. One version was used in conjunction with the secondary CRT task and the other was used under control conditions of undivided attention. Each version was presented concurrently with the CRT task for half of the participants. The sentences and word lists were recorded using a microphone attached to a computer, and were copied onto a compact disc. They were spoken in a monotone female voice, at the rate of 2 words per second. Each word was recorded separately and neighbouring words were placed together using computer software (Wave Studio, Creative), to ensure that there were no differences in prosody between the sentences and word lists.

The secondary choice reaction time (CRT) task was based on a task used by Craik et al. (1996). The computerised task was presented using SuperLab software (Cedrus), and involved a visual display of four boxes in a line. A star appeared at random in one of the boxes, and the subject was required to press a corresponding key on the keyboard as fast as possible (z, x, ., or /). The task was self-paced, with the next trial being presented immediately after a correct response. The advantage of this method was that errors could be expressed as a time cost, allowing a single measure (time taken to produce correct response) to indicate performance. The computer recorded reaction times (RT) for correct and incorrect responses.

Procedure

Span was assessed for unrelated words and sentences at the start of the testing session. The word lists and sentences were produced using the method described above, and were played back from a tape. Testing was started at four words for word span and at eight words for sentence span. No participants failed to obtain the minimum span tested. Participants were asked to repeat back the words they had heard in the correct order, taking care to use exactly the same words. Four strings of words were played at each span level. Span was defined as the longest length at which two of the four strings were repeated correctly. Testing was stopped when participants failed to recall all four of the strings.

Each participant received practice at the CRT task, followed by a total of six blocks of practice on the sentence and word list recall tasks while performing the

CRT task concurrently. The length of the material on the practice trials was 50% greater than the mean sentence and word spans obtained during pilot testing. In each block of practice, the same words or sentences were played twice, on two successive trials. There was a period of 10s before the words were played, in which the CRT task was performed alone. Subjects were required to continue with the CRT task while they listened to the words, and during a 10-s retention interval. If they stopped or slowed noticeably on the CRT task, they were prompted to continue. When a tone was sounded, participants stopped performing the CRT task, and attempted to recall what they had heard in serial order. After 15s, the tone was repeated to indicate that the next trial was starting, and the CRT task should be continued. The test trials were identical to the practice trials, except that the retention interval was 20s and each sentence or word list was presented three times in successive learning trials. Trials without the concurrent CRT task had the same structure, but participants were only required to remember the words they heard. Written records were made of participants' recall attempts and, where necessary, these were checked against an audio recording of the experimental session.

Results

Span measures

Mean word span was 6.50 words (standard deviation ($SD = 1.06$), and mean sentence span was 13.71 words ($SD = 1.90$). There was a significant correlation between the two span measurements ($r = .61$, $p < .001$) and sentence span was significantly larger than word span ($t(23) = 23.50$, $p < .001$).

Recall accuracy

Figs. 1A and B show the mean percentage of words recalled in each condition. As participants were asked to recall the items in serial order, it was possible to examine both the percentage of words recalled in total, and the percentage of words recalled in the correct order. The results for these two measures were very similar and for the sake of brevity only total recall is presented for all three experiments.

A repeated-measures analysis of variance (ANOVA) revealed significant main effects of material, secondary task, and trial. The sentences were recalled better than the word lists ($F(1, 23) = 20.25$, $p < .001$), the secondary task significantly impaired retention ($F(1, 23) = 8.77$, $p < .01$), and recall improved with repeated presentations of the same material ($F(2, 46) = 240.17$, $p < .001$). There was a significant interaction between material and trial ($F(2, 46) = 15.26$, $p < .001$) caused by faster learning of the sentences than the word lists. Bonferroni tests indicated that the recall of words and sentences did not differ on Trial 1 ($t(23) < 1$), whereas the sentences

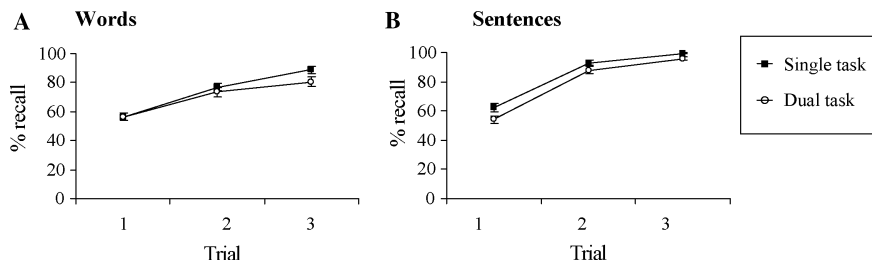


Fig. 1. Recall of words and sentences in Experiment 1 (50% above span). Figures show percentage of items recalled in any order. Error bars show standard error of mean.

were recalled more accurately than the words on Trials 2 and 3 ($t(23) = 5.50, p < .001$ and $t(23) = 5.57, p < .001$, respectively).

The three-way interaction between material, secondary task, and trial was also significant ($F(2,46) = 5.13, p < .05$). The size of the dual task decrement for sentences did not change significantly across the three trials in which the same material was repeated ($F(2,46) = 1.20, ns$). In contrast, the secondary task effect for word lists increased over trials ($F(2,46) = 4.15, p < .05$). Bonferroni tests revealed that the secondary task did not significantly affect retention of the word lists on either Trials 1 or 2 ($t(23) < 1$ and $t(23) = 1.23, ns$, respectively) but did impair retention of the word lists on Trial 3 ($t(23) = 3.44, p < .01$). This finding suggests that presenting the word lists repeatedly facilitated the use of attention-demanding encoding processes.

The secondary task impaired retention of the sentences to a greater extent than the word lists on Trial 1 ($F(1,23) = 4.75, p < .05$) but not on Trials 2 or 3 ($F(1,23) < 1$ and $F(1,23) = 2.52, ns$). Bonferroni tests indicated that on Trial 1, the secondary task impaired recall of the sentences ($t(23) = 2.63, p < .05$) but did not impair recall of the word lists ($t(23) < 1$). Therefore, attentional processes may have played a larger role in sentence than word span on Trial 1 when the material was novel. On later trials this interaction did not occur, presumably because attention-demanding processes played a greater role in word recall following learning.

Recall errors

Different types of errors during word and sentence recall may reflect differences in the coding used in these tasks. Incorrect responses were categorised as omission, order, and item errors. Omission errors were calculated by subtracting the correct responses and other error types from the number of presented words. Order errors were target words produced in the wrong place in the sequence. Item errors were further classified as phonological, semantic, morphological, repetition, and unrelated errors. Phonological errors shared at least half of their phonemes with a target word. Semantic errors were related in meaning to one of the target words. Morphological errors retained the correct word stem, but were incorrectly marked (e.g., 'walks' became 'walked'). These errors were also semantically and phonologically related to the target word, but were placed in this separate category. Repetition errors were target words produced more than once. Errors that did not fall into one of these categories were classified as 'unrelated.' A small number of 'unclassifiable' errors fell into more than one category and were not analysed further.

Table 1 shows the mean number of omission, morphological, phonological, semantic, repetition, and unrelated errors for words and sentences as a percentage of the number of items presented, and order errors as a percentage of total item recall. A series of t tests, corrected for multiple comparisons, revealed that omission and order errors occurred more frequently in word recall

Table 1
Recall errors in the three experiments

		Order ^a	Omission ^b	Total item ^b	Semantic ^b	Phonological ^b	Morphological ^b	Repetition ^b	Unrelated ^b
Exp. 1	Words	14.3 (9.7)	26.5 (10.2)	1.7 (1.6)	0.1 (0.5)	0.1 (0.4)	0.9 (1.2)	0.6 (1.1)	0.0 (0.0)
	Sentences	3.1 (3.5)	14.7 (6.1)	4.2 (2.8)	1.9 (1.7)	0.1 (0.2)	0.9 (1.5)	0.3 (0.5)	0.8 (1.0)
Exp. 2 (All lengths)	Words	19.3 (9.4)	36.3 (6.1)	10.7 (7.0)	0.3 (0.7)	2.2 (1.5)	3.4 (1.9)	2.6 (2.1)	1.9 (1.8)
	Sentences	16.2 (15.9)	26.2 (8.1)	6.6 (3.7)	2.1 (1.2)	0.4 (0.6)	1.2 (1.1)	1.1 (1.0)	1.7 (1.4)
Exp. 3	Words	10.0 (5.8)	21.6 (7.2)	13.0 (5.7)	0.3 (0.6)	4.9 (4.1)	2.9 (2.4)	2.3 (2.2)	2.1 (2.3)
	Sentences	3.8 (3.5)	14.0 (8.4)	8.5 (4.7)	2.8 (1.7)	0.7 (0.9)	0.9 (0.9)	1.9 (2.1)	2.1 (1.9)
	Stories	2.9 (2.9)	5.1 (3.2)	8.8 (5.9)	1.9 (2.2)	0.6 (1.2)	1.1 (1.4)	2.7 (2.1)	2.3 (1.9)

^a Mean % of total items recalled (SD).

^b Mean % of items presented (SD).

($t(23) = 4.69, p < .001$ and $t(23) = 5.33, p < .001$, respectively). Semantic and unrelated errors were more common in sentence recall ($t(23) = 5.37, p < .001$ and $t(23) = 4.11, p < .05$, respectively), possibly reflecting the greater role played by semantic coding in this task. All of these effects reached significance in nonparametric tests (Wilcoxon's signed ranks).

Reaction times

RT was recorded during three periods in each trial: before the words began to play (baseline), while the words were playing (encoding), and after they had finished playing but before recall could commence (retention). Mean reaction times are shown in Table 2.

In the data set as a whole, reaction times were faster during sentence than word recall ($F(1, 23) = 22.84, p < .001$). In addition, reaction times were slower when the CRT task was performed concurrently with the memory task compared with when it was performed alone ($F(2, 46) = 53.05, p < .001$). RT was slower in the retention period relative to baseline (Bonferroni: $t(23) = 6.06, p < .001$) and was slower during encoding than retention ($t(23) = 4.71, p < .001$).

There was a significant interaction between material and trial ($F(2, 46) = 3.29, p < .05$) apparently because RT increased for words but not sentences over the three trials in which the same material was repeated. Bonferroni: tests indicated that there was no significant RT difference between the words and sentences on Trial 1 ($t(23) = 1.60, ns$) whereas RT was slower for the words on Trials 2 and 3 ($t(23) = 3.59, p < .05$, and $t(23) = 4.13, p < .001$, respectively). This finding is consistent with the larger dual task decrement in retention observed for the word lists on later trials, and suggests that attention-demanding processes played a greater role in word span after learning had occurred.

There was a significant interaction between material and period ($F(2, 46) = 11.74, p < .001$) because the overall RT cost of performing the memory task was greater for the words than for the sentences. There was no significant difference between words and sentences in the baseline period ($t(23) < 1$), but reaction times were slower for the words during the encoding and retention periods ($t(23) = 4.73, p < .001$, and $t(23) = 3.42, p < .05$, respectively).

Discussion

In this experiment, participants were required to perform a choice reaction time (CRT) task whilst encoding and retaining sentences and word lists for serial recall. The CRT task was found to disrupt memory for novel sentences to a greater extent than word lists. Memory for the two materials was equivalent under single task conditions, suggesting that the differential effect of the secondary task could not be attributed to overall difficulty. Although the chunking processes underlying the span advantage for sentences could have been relatively automatic (relying, for example, on pre-existing associations between representations in LTM), our results are consistent with the view that attention played an important role in sentence recall, in line with the proposals of Baddeley (2000) and Cowan (1995). These accounts suggest that attention is required to forge novel associations in WM: the formation of such links may be crucial for sentence comprehension/recall, given the need to assimilate the words comprising a sentence (and in this experiment specifically, the need to bind together several sentences with unrelated meanings; see Experiment 3). In contrast, the recall of relatively meaningless lists of unrelated words may initially depend more on automatic phonological coding that is comparatively resistant to the effects of divided attention.

A larger number of semantic errors occurred during the sentence span task, suggesting that semantic coding played a greater role in the retention of sentences than word lists. Fewer order errors occurred for the sentences, presumably because the sentence structure placed constraints on the likely order of the words.

Over three trials in which the same material was repeated, the sentences were learned more quickly than the word lists. In addition, the secondary task produced a greater disruption to the recall of unrelated words following learning, so that on later trials, an equivalent secondary task decrement was observed for the two materials. This finding suggests that although attention-demanding processes played a rather limited role in the memory task for novel lists of unrelated words, they made a more substantial contribution following learning.

Table 2
Mean (SD) reaction times in milliseconds in Experiment 1

	Words			Sentences		
	Baseline	Encoding	Retention	Baseline	Encoding	Retention
Trial 1	477 (87)	790 (230)	661 (167)	474 (53)	713 (139)	612 (200)
Trial 2	482 (84)	918 (490)	782 (369)	487 (91)	651 (184)	588 (129)
Trial 3	482 (84)	901 (328)	679 (230)	466 (75)	665 (138)	578 (156)
Average	480	870	707	476	676	593

How can we account for this finding? On the first trial, participants may have been largely unable to forge links between the unrelated words, especially given the rapid presentation rate of 2 items per second, and consequently they may have relied on relatively automatic phonological coding. As the word set became more familiar however, participants may have been able to enhance their recall by forming novel associations between the list items: according to the theories of [Baddeley \(2000\)](#) and [Cowan \(1995\)](#), such integrative encoding is attention-demanding, potentially explaining the increase in the impact of divided attention on word recall.

Reaction times on the secondary CRT task were consistent with this interpretation. On the initial trial, there was no RT difference for words and sentences. It seems unlikely, therefore, that the larger dual task decrement in retention observed for sentences could be accounted for in terms of differential trade-offs between the memory and CRT tasks. During subsequent presentations of the same material, RT increased for the unrelated words, consistent with a greater role of attention-demanding encoding following learning. On these later trials, the sentences had benefited from rapid learning and were recalled more accurately than the word lists so the faster RT associated with the sentence recall task may also have arisen because the sentences were less taxing.

In sum, these results suggest that, on the first trial, attention played a more substantial role in the memory task for sentences than word lists. On subsequent trials, the unrelated word lists also became sensitive to disruption by the secondary task. We have argued that this pattern reflects the importance of attention-demanding integrative encoding in sentence span, and a shift from relatively automatic phonological to more controlled integrative encoding of unrelated word lists following learning. However, one caveat should be noted: the effect of divided attention purportedly arose during encoding; however, as the CRT task was performed both while the sentences and words were presented and throughout a retention interval in this experiment, it is not clear whether the dual task decrement reflected the disruption of attentional control processes during encoding or retention or both. In the following experiment, the retention period was eliminated from the experimental design, allowing a more specific attribution of the dual task decrement to the encoding period.

Experiment 2

This experiment examined the recall of word lists and sentences with and without a concurrent secondary task, to provide a replication and extension of Experiment 1. Different lengths of material were examined, ranging from 50% above span (as in Experiment 1) to 200%

above span. On trials where length was increased far above span, percentage recall might be expected to decline more gradually for sentences than word lists because of the greater involvement of LTM in the immediate recall of sentences. Word span might be expected to be less resistant to supra span increases in length because of its greater reliance on the limited capacity phonological loop.

Method

Design

A repeated-measures design was used to examine the impact of material, secondary task, trial, and list length on immediate recall. Participants were presented with material at three lengths: 50% above span, 100% above span, and 200% above span. The order of the conditions was pseudo-randomly assigned across the subjects in a way that controlled for practice effects.

Participants

The participants were 24 psychology undergraduates who took part for course credit. Three participants with exceptionally slow reaction times (more than 3.5 *SD* away from the mean) were excluded from the analysis and replaced.

Materials

The materials were constructed from the pool of sentences used in the previous experiment. As in Experiment 1, the sentence material for each trial was formed from several unrelated sentences (three to five sentences for the 50% above span condition; four to seven for the 100% above span condition; and 6–10 for the 200% above span condition). Sentences were not repeated during the course of the experiment. Two parallel versions of the test material were created for each span level. One version was used in conjunction with the secondary CRT task, and the other was used under single task conditions. Both versions occurred in the single task condition for half the participants. The secondary task was the choice reaction time task used in Experiment 1. Other details are as described for Experiment 1.

Procedure

Span was assessed for unrelated words and sentences at the start of the testing session, using the method described for Experiment 1. Participants then received practice at the CRT task alone, followed by six blocks of practice at the sentence and word recall tasks while performing the CRT task concurrently. The length of the practice material was 50, 100, and 200% greater than the average span for words and sentences obtained in Experiment 1. The practice and test trials in Experiment 2 had a similar structure to those in Experiment 1,

although the retention period was removed to allow a more specific attribution of the influence of the secondary task to encoding processes. In trials involving the CRT task, there was a 10s period of CRT before the words began to play, which was used to indicate baseline reaction times, and participants were required to carry out the CRT task while the sentences or words were playing. When a tone sounded, participants stopped the CRT task, and attempted to recall what they had heard in correct serial order. Trials without the concurrent CRT task had the same structure but participants were only required to try to remember the words they heard. As in Experiment 1, each sentence or word list was repeated three times to allow the effect of learning to be examined.

Results

Span measures

Mean word span was 5.88 words ($SD = 0.80$) and mean sentence span was 12.88 words ($SD = 1.75$). There was a significant correlation between the two span measurements ($r = .42$, $p < .05$) and sentence span was

significantly larger than word span ($t(23) = 21.60$, $p < .001$).

Recall accuracy

Figs. 2A and B indicate the percentage of words recalled in any order for words and sentences when length was 50% above span, Figs. 2C and D indicate recall at 100% above span and Figs. 2E and F recall at 200% above span. The main effects of material, dual task, length, and trial were all significant. Sentences were recalled more accurately than the unrelated words ($F(1,23) = 68.65$, $p < .001$), recall was poorer under dual-task conditions ($F(1,23) = 11.24$, $p < .01$), percentage recall was lower on longer lists ($F(2,46) = 210.68$, $p < .001$) and recall improved over repeated presentations of the same material ($F(2,46) = 445.31$, $p < .001$)

Shortest list length (50% above span). To allow direct comparison with the results of Experiment 1, the data from the three list lengths will be described separately, starting with the 50% above span length that was used in the previous experiment. At this list length, there was a significant interaction between material and trial

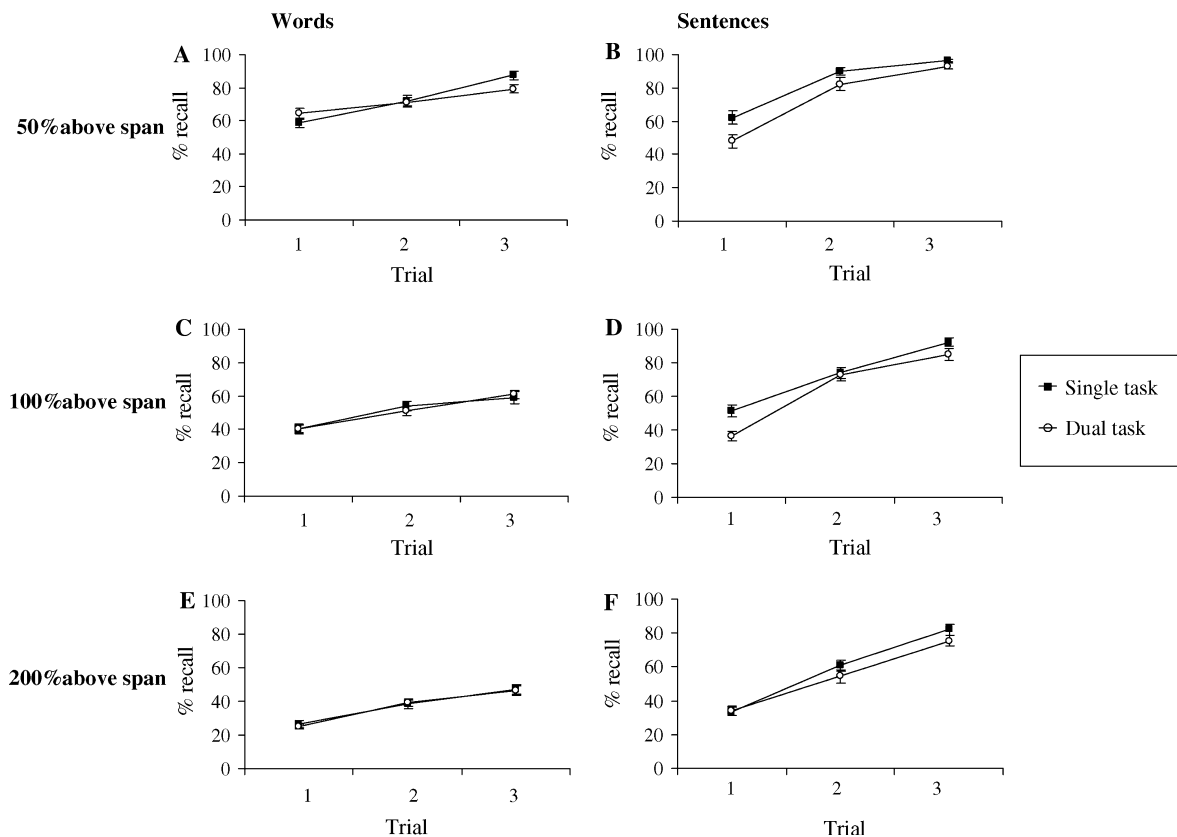


Fig. 2. Recall of words and sentences in Experiment 2. Figures show percentage of items recalled in any order. Error bars show standard error of mean.

($F(2,46) = 24.35, p < .001$) suggesting that, as in Experiment 1, the sentences were learned more rapidly than the lists of unrelated words. Bonferroni: t tests indicated that the recall of the words and sentences did not differ on Trial 1 ($t(23) = 1.79, ns$), whereas the sentences were recalled more accurately than the words on Trials 2 and 3 ($t(23) = 5.03, p < .001$ and $t(23) = 6.12, p < .001$, respectively).

As in Experiment 1, the three-way interaction between material, secondary task, and trial was significant ($F(2,46) = 7.97, p < .01$). Divided attention impaired word recall to a greater extent following learning ($F(2,46) = 4.05, p < .05$). Bonferroni: t tests revealed that the secondary task did not affect retention of the word lists on either Trials 1 and 2 ($t(23) < 1.40, ns$), but did impair retention of the word lists on Trial 3 ($t(23) = 2.79, p < .05$). Therefore, attentional processes may have played a greater role in the word span task following learning. In contrast, the size of the dual task decrement for sentences actually decreased across the three trials due to ceiling effects ($F(2,46) = 4.46, p < .05$). The secondary task impaired sentence recall on Trial 1 ($t(23) = 3.53, p < .01$), but not on Trials 2 and 3 ($t(23) < 1.89, ns$).

This experiment also replicated the finding of a larger dual task decrement for sentences than for unrelated words on Trial 1 ($F(1,23) = 11.74, p < .01$), suggesting that attentional processes played a larger role in sentence than word recall when the material was novel. This interaction did not occur on Trials 2 and 3 ($F(1,23) < 2.40, p < .14$), consistent with the notion that learning increased the role of attentional processes in the word span task.

Therefore, at 50% above span in both experiments: (1) divided attention impaired the recall of novel sentences more than word lists and (2) the effect of the secondary task increased for word lists when they were presented repeatedly.

Longer list lengths (100 and 200 % above span). The material by trial interaction remained significant when list length was increased to 100% above span ($F(2,46) = 43.57, p < .001$) and 200% above span ($F(2,46) = 51.24, p < .001$), reflecting the fact that the sentences were learned faster than the word lists with repeated presentations. In fact, the strength of the material by trial interaction grew with list length ($F(4,92) = 4.92, p < .01$), apparently because on the shorter list lengths, sentence recall on later trials was limited by ceiling effects.

The three-way interaction between material, dual task, and trial was marginally significant for the 100% above span condition ($F(2,46) = 4.23, p < .05$) but did not reach significance for the 200% above span condition ($F(2,46) = 1.88, ns$). The strength of this three-way interaction appeared to weaken as list length increased, consistent with the four-way interaction between mate-

rial, dual task, trial, and list length that was obtained in the experiment as a whole ($F(4,92) = 6.33, p < .001$). For the two longer list lengths, divided attention did not impair the recall of unrelated words on any trial, even following learning (all $t(23) < 1.08, ns$). It is possible that the number of items in the lists, combined with the rapid presentation speed, discouraged participants from adopting attention-demanding integrative encoding to enhance word recall.

The larger dual task decrement for sentences compared with words, observed on Trial 1 in the 50% above span condition, also reached significance for lists 100% above span ($F(1,23) = 6.54, p < .05$) but did not occur for lists 200% above span ($F(1,23) < 1$). On Trials 2 and 3, there was no material by dual task interaction for any list length (all comparisons were ns). In summary, we found that conditions of divided attention impaired the recall of novel sentences to a greater extent than unrelated words when list length was relatively short but not when it was longer.

The material by length interaction in the experiment as a whole reached significance ($F(2,46) = 13.33, p < .001$). As the length of material to be remembered was increased, percentage recall declined less rapidly for sentences compared with words. Bonferroni: tests indicated that, in the experiment overall, there was a small but significant recall difference between words and sentences on the shortest length, collapsing across the three trials ($t(23) = 2.97, p < .05$). Sentence recall was substantially better than word recall on the longer lengths, 100 and 200% above span ($t(23) = 6.77, p < .001$ and $t(23) = 8.71, p < .001$, respectively). The greater involvement of LTM in sentence span may have made this task more resistant to supra span increases in length. Word span may have been less resistant because of its heavy reliance on the limited capacity phonological loop.

Recall errors

Errors were coded as for Experiment 1. Table 1 shows the mean number of omission, morphological, phonological, semantic, repetition, and unrelated errors for words and sentences as a percentage of the number of items presented. Order errors are also shown, expressed as a percentage of total item recall. A series of t tests, corrected for multiple comparisons, revealed that there were more omission ($t(23) = 6.60, p < .001$), morphological ($t(23) = 5.27, p < .001$), phonological ($t(23) = 5.95, p < .001$), and repetition errors ($t(23) = 2.98, p < .05$) in word recall. In contrast, semantic errors were more common in sentence recall ($t(23) = 8.13, p < .001$), possibly reflecting the greater role played by semantic coding in this task. All of these effects reached significance in nonparametric tests (Wilcoxon's signed ranks).

Although there was no overall difference in order errors for words and sentences, order errors were more

numerous for the words at the shortest list length, consistent with the results of Experiment 1 (words mean = 16.8%, $SD = 7.8$; sentences mean = 8.5%, $SD = 11.6$; Bonferroni: $t(22) = 4.35$, $p < .001$). In addition, there was a material by dual task interaction for order errors in this experiment ($F(1,22) = 6.41$, $p < .05$). Post hoc analyses indicated that the secondary task significantly impaired order memory for word lists (single task condition = 15.7%, $SD = 9.9$; dual-task condition = 22.9%, $SD = 10.9$; Bonferroni: $t(22) = 3.82$, $p < .01$). In contrast, the secondary task did not impair order memory for sentences (single task condition = 15.7%, $SD = 15.6$; dual-task condition = 16.7%, $SD = 17.0$; Bonferroni: $t(22) < 1$). This finding is consistent with the notion that attention plays a role in integrative encoding in verbal STM. For sentences, there are long-term linguistic constraints on word order and these might operate relatively automatically. In contrast, the order of unrelated words is arbitrary and attention might be required to form links between these previously unrelated items. It should be noted, however, that this interaction did not reach significance (and is not discussed) for Experiments 1 and 3. It seems likely that this was because the longer list lengths in the current experiment made an important contribution to the material by trial interaction, although the higher order interaction with length was not significant.

Reaction times

RT was recorded during two periods in each trial: before the words began to play (baseline), and while they were playing (encoding). Mean RT is shown in Table 3.

Overall, there were significant main effects of material ($F(1,23) = 46.66$, $p < .001$) and time period ($F(1,23) = 68.13$, $p < .001$). RT was slower for words than for sentences and slower during the encoding period relative to baseline.

There was a significant three-way interaction between material, trial, and time period ($F(2,46) = 3.85$, $p < .05$). The RT cost of encoding increased for words ($F(2,46) = 4.58$, $p < .05$) but not sentences ($F(2,46) < 1$) over the three successive trials, in line with the findings of Experiment 1. For unrelated words, there was a marginally significant increase in RT during encoding between Trials 1 and 2 (uncorrected $t(23) = 2.36$, $p < .05$) but not between Trials 2 and 3 ($t(23) < 1$). There was no change in baseline RT across the three trials ($t(23) < 1.17$, *ns*). This pattern is suggestive of a greater role for attention-demanding encoding processes in word recall following learning and is consistent with the larger dual task decrement in retention observed for words on later trials.

As in Experiment 1, there was no RT difference between sentences and words on Trial 1 for the 50% above span length ($t(23) < 1$), suggesting that the larger dual task decrement in retention observed for sentences observed under these conditions could not be accounted for in terms of differential trade-offs between the memory and CRT tasks.

The overall RT cost of encoding was greater for words than for sentences ($F(1,23) = 32.42$, $p < .001$). In addition, the interaction between time period and trial in the data set as a whole was marginally significant, reflecting the pattern observed for unrelated words ($F(2,46) = 3.18$, $p = .05$).

Table 3
Mean (SD) reaction times in milliseconds in Experiment 2

		Words		Sentences	
		Baseline	Encoding	Baseline	Encoding
50% above span	Trial 1	452 (88)	685 (214)	427 (76)	661 (212)
	Trial 2	453 (99)	781 (272)	445 (78)	629 (212)
	Trial 3	449 (101)	764 (262)	448 (130)	634 (214)
	Average	452	744	440	641
100% above span	Trial 1	440 (71)	720 (230)	440 (92)	616 (218)
	Trial 2	437 (71)	734 (281)	440 (93)	611 (175)
	Trial 3	441 (86)	806 (398)	446 (96)	641 (214)
	Average	439	753	442	623
200% above span	Trial 1	460 (105)	731 (267)	438 (86)	610 (176)
	Trial 2	437 (71)	840 (298)	451 (113)	628 (204)
	Trial 3	461 (99)	767 (366)	429 (82)	650 (244)
	Average	453	780	439	630
Average	Trial 1	451	712	435	629
	Trial 2	442	785	446	623
	Trial 3	450	779	441	642
	Average	448	759	440	631

Discussion

The experiment replicated the key findings of Experiment 1. For sequences 50% above span, a concurrent choice reaction time (CRT) task during encoding was found to disrupt the retention of novel sentences to a greater extent than word lists, even though the two materials were recalled equally well in the absence of the CRT task. This finding suggests that attention may play a role in the binding that underpins sentence recall, as envisaged by [Baddeley \(2000\)](#).

The secondary task had relatively little effect on the retention of the unrelated words when they were first presented but a more detrimental effect on their recall following learning, as in the previous experiment. This result indicates that with multiple presentations, attention-demanding encoding processes play an increasingly substantial role in word recall. Reaction times on the CRT task were consistent with this pattern as they increased for words but not sentences when the material was repeated. Repetition of the words may have allowed participants to improve their recall through the formation of associations between unrelated items in WM: this is envisaged as attention demanding by several theoretical perspectives (e.g., [Baddeley, 2000](#); [Cowan, 1995](#)). It seems likely that the dissimilar nature of the words and the rapid presentation speed would have discouraged the formation of these links on the first trial.

As in Experiment 1, semantic errors were more frequent in sentence than word recall. In contrast, phonological errors were more common in word recall, supporting the view that this task relied more heavily on a phonological code. The error analysis also suggested that divided attention specifically impaired order memory for unrelated words but not sentences. This finding concurs with the hypothesis that encoding in WM is more attention demanding when participants are required to form new associations between previously unrelated LTM representations. The order of unrelated words is arbitrary and consequently might require attention to learn, whereas there are pre-existing linguistic constraints on word order for sentences that might operate relatively automatically. It will, however, be important to establish the generality of this finding in future research.

As the amount of material to be remembered was increased, percentage recall declined, although less rapidly for sentences than for word lists. The greater linguistic support in the sentence span task may have made memory for this type of material more robust to supra span increases in length. Word span, on the other hand, may have been more reliant on the limited capacity phonological loop and consequently unable to withstand supra span increases in length. The retention of long lengths of material was relatively unaffected by conditions of divided attention, suggesting

that the contribution of linguistic factors depended principally on automatic rather than controlled processes under these conditions. Although we did not specifically predict this outcome, participants may have been more likely to abandon attention-demanding encoding strategies when the task itself was very taxing.

Experiment 3

The experiments discussed so far have suggested that LTM support for WM relies on both attention-demanding and more automatic encoding processes, with their relative importance varying across tasks. The recall of novel sentences was found to place greater demands on controlled processing than the recall of unrelated word lists; however, it is not clear which aspects of the sentence recall task were effortful. There are at least two possibilities, both of which are broadly consistent with the proposals of [Baddeley \(2000\)](#) and [Cowan \(1995\)](#). First, it seems likely that sentence recall requires the integration of different linguistic codes as well as diverse propositions/concepts in order to derive a mental model of the overall meaning of the sentence, and this process of semantically/syntactically guided chunking may require attention. Alternatively, as the sentential material was composed of several short sentences that were unrelated in meaning, it is possible that divided attention interfered with the binding of unrelated sentence chunks (or reduced the number of unrelated sentences that participants were able to maintain simultaneously), rather than interfering with the process of within-sentence chunking per se. This experiment allowed a comparison of these possibilities.

The second alternative is perhaps coherent with [Cowan's \(2000\)](#) suggestion of a basic capacity limitation of four unrelated chunks that can be held in the focus of attention at any one time: conditions of divided attention might be expected to reduce this capacity, preventing the formation of links between unrelated sentences and causing participants to forget entire sentences. The distinction between these two possibilities may also be relevant to the LT-WM model ([Ericsson & Kintsch, 1995](#)): although the creation of retrieval structures may be effortful in some circumstances, as in the case of digit span experts, it is unclear whether mental model construction in sentence comprehension/recall draws on exactly the same processes ([Miyake & Shah, 1999](#)). Consequently, the ability to forge links between several unrelated sentences may be more effortful than the parsing and comprehension of a single sentence.

Although the items in the word lists were unrelated in meaning, attention-demanding binding processes may

not have been crucial in supporting their recall on the first trial because the specific conditions in the experiments (e.g., fast presentation speed and relatively high proportion of function words with few semantic associations) would have made it difficult for participants to adopt such strategies, forcing them to fall back onto a phonological code. Consequently, chunk size would have been rather limited and a dual task decrement in the ability to bind together unrelated chunks would have had relatively little impact on overall recall. It seems likely that participants would have started to construct larger multiword chunks with repeated presentations of the unrelated word list, producing a larger dual task decrement on later trials. In addition, how can we account for the finding that the secondary task substantially impaired sentence recall at shorter but not longer list lengths? At longer list lengths, the number of unrelated sentences would have greatly exceeded participants' WM capacity as envisaged by Cowan (2000); as a result, pre-established linguistic associations operating relatively automatically would have been more critical in determining recall than the ability to bind together unrelated sentences.

In Experiment 3, the recall of unrelated words and unrelated sentences was examined as before, but in addition, the design incorporated related sentences that formed a story. If the construction of mental models through semantically guided chunking is attention-demanding, a concurrent task might be expected to have the greatest impact on story recall, which presumably involves the construction of complex mental models spanning several sentences. In contrast, if divided attention specifically makes it difficult to bind together unrelated chunks, the concurrent task should have the greatest impact on the retention of unrelated sentences.

Method

Design

A repeated-measures design was used. Recall of unrelated words, unrelated sentences, and stories was tested with and without a concurrent secondary task. The length of test sequences was 50% greater than each participant's span for the three materials.

Participants

A total of 24 participants were paid to take part in the experiment. They were a mixed group of undergraduate and postgraduate students and volunteers. Two subjects with exceptionally slow reaction times (more than 3.5 *SD* away from the mean) were excluded from the analysis and replaced. The order of conditions was pseudo-randomly assigned across subjects, so that each condition appeared in each position an equal number of times.

Materials

The unrelated sentences and words in both the span tests and the experimental trials were derived from the story material. The stories were composed of short sentences containing between five and eight words, each with a unique content that built on the previous sentences. For the experimental trials, eight stories were constructed, each containing six sentences. Several versions of each sentence that contained slightly different numbers of words were produced, so that material for the full range of spans could be constructed. Two of the stories were selected to be the test stories, and versions were constructed for each span level. The remaining six stories formed the unrelated sentence material. Sentences were selected from these stories so that participants never heard the same sentence twice, and so that sequences of sentences never contained more than one sentence from a story. The unrelated test words were drawn pseudo-randomly from the story and unrelated sentence test material. Words were not repeated within a trial.

Some additional controls were employed to ensure that the three types of material were equivalent. The words used in the stories and sentences were divided into four groups: (1) nouns, (2) verbs, (3) adjectives and adverbs, and (4) function words. Each of these groups was divided into four frequency categories, using estimates of written word frequency from the Celex database (Baayen, Piepenbrock, & van Rijn, 1993). The proportions of words in each of these categories were calculated for the stories and sentences, and the composition of both materials was adjusted so that the proportions in the stories and sentences, and in the span and test materials were as similar as possible. Words were selected pseudo-randomly from the stories and sentences to form the unrelated word material. The words were chosen so that the overall proportion of words in each frequency category for each word type matched those obtained for the story and sentence materials. Some repetition of common function words was inevitable in the stories and sets of sentences as a whole, but this was kept to a minimum, and the number of repeats was equal in the two materials.

The span tests were constructed from 144 short stories, varying in length from two to four sentences (10 to 28 words). Half of the stories, selected at random, were used in the story span test and half were used in the sentence span test. The sentence span test was constructed by selecting sentences pseudo-randomly from the stories assigned to this condition, so that the material had little overall meaning. List length was increased two words at a time in the sentence and story span tests. The word span test was constructed by drawing unrelated words pseudo-randomly from the story and sentence span tests. Length was increased one word at a time. Other methodological details were as described above for the experimental trials.

Two versions of the test stories, unrelated sentences, and unrelated word lists were formed. One was used in conjunction with the secondary CRT task, while the other was used under single task conditions. The two versions were assigned randomly to the single and secondary task conditions, but so that each occurred in the single task condition for half the participants. The secondary task was the choice reaction time task, employed in Experiments 1 and 2.

Procedure

Spans for unrelated words, unrelated sentences, and stories were assessed at the start of the testing session. Participants heard four strings at each length, and had to repeat two of them correctly for a span of that length. Participants then received practice at the CRT task alone, followed by six blocks of practice at the recall task while performing the CRT task concurrently. Participants received equal practice for each of the materials.

The practice and test trials had the same structure as in Experiment 2. In dual task trials, there was a 10s period of CRT before the words began to play, which was used to indicate baseline reaction times, and participants were required to carry out the CRT task while the verbal materials were playing. When a tone sounded, participants stopped the CRT task and attempted to recall what they had heard in correct serial order. Trials without the concurrent CRT task had the same structure, but participants were only required to try to remember the words they heard. As in Experiments 1 and 2, each sen-

tence or word list was played three times in successive trials.

Results

Span measures

Mean span was 5.8 ($SD = 0.85$) for unrelated words, 15.4 ($SD = 2.60$) for unrelated sentences, and 15.4 ($SD = 2.60$) for stories. Word span correlated significantly with story span ($r = .52, p < .01$) but not sentence span ($r = .32, ns$). The strongest correlation was between sentence span and story span ($r = .67, p < .001$). Word span was significantly lower than both sentence and story span ($t(23) = 18.95, p < .001$ and $t(23) = 20.47, p < .001$, respectively). There was no significant difference between sentence and story span ($t(23) < 1$), suggesting that the presence of a story structure did not increase the number of words that participants could remember without error. It seems likely that this is because the span test examined recall of surface structure; in contrast, gist recall may have been superior for the stories (Johnson-Laird & Stevenson, 1970; Sachs, 1967).

Recall accuracy

Figs. 3A–C show the mean percentage of words recalled in any order for unrelated word lists, unrelated sentences, and stories. To facilitate comparison between experiments, participants' recall of the unrelated words and sentences will be described first, followed by their recall of the stories. In the experiment as a whole,

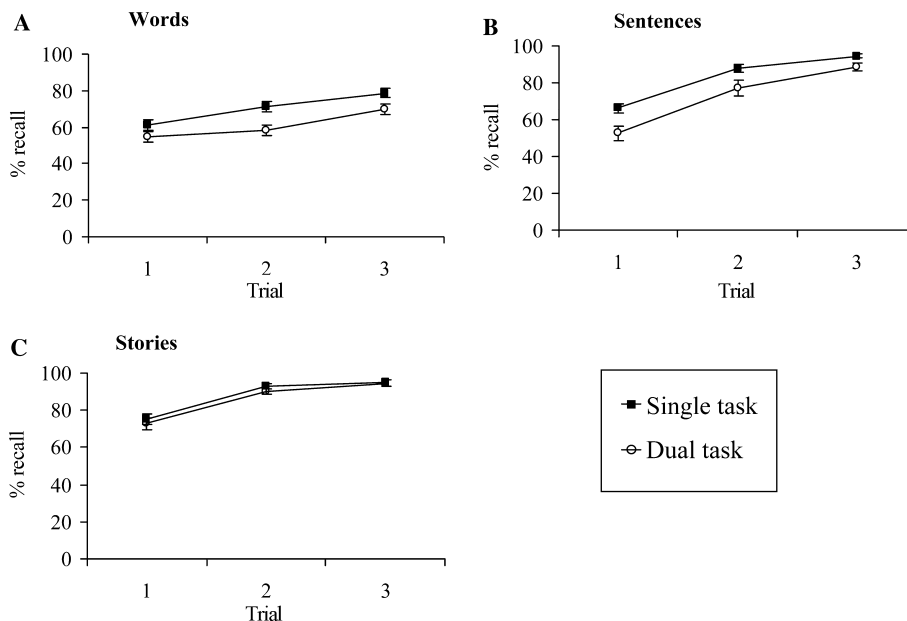


Fig. 3. Recall of words, sentences, and stories in Experiment 3 (50% above span). Figures show percentage of items recalled in any order. Error bars show standard error of mean.

there were significant main effects of material ($F(2, 46) = 45.69, p < .001$), dual task ($F(1, 23) = 17.73, p < .001$), and trial ($F(2, 46) = 208.01, p < .001$) and an interaction between material and trial ($F(4, 92) = 13.13, p < .001$).

Unrelated words and sentences. The sentences were recalled more accurately than the word lists in the experiment as a whole ($F(1, 23) = 21.91, p < .001$). There was an interaction between material and trial ($F(2, 46) = 20.52, p < .001$), which indicated that the sentences were learned more rapidly than the word lists, as in the previous experiments. Bonferroni: t tests showed that words and sentences were recalled at an equivalent level on Trial 1 ($t(23) < 1$); in contrast, the sentences were recalled more accurately than the words on Trial 2 ($t(23) = 5.85, p < .001$) and Trial 3 ($t(23) = 8.39, p < .001$). The three-way interaction between material, dual task and trial also approached significance but was weaker than in the two previous experiments ($F(2, 46) = 2.70, p = .08$). There was no significant material by dual task interaction for the first trial ($F(1, 23) = 1.67, ns$), apparently because divided attention had a somewhat larger effect on the retention of novel lists of unrelated words in this experiment. Nevertheless, planned comparisons indicated that the secondary task significantly impaired recall of sentences ($t(23) = 3.52, p < .01$) but not word lists ($t(23) = 1.56, ns$) on Trial 1, in line with the results of Experiments 1 and 2. On later trials, divided attention had a significant impact on the recall of both word lists (Trial 2: $t(23) = 2.85, p < .01$; Trial 3: $t(23) = 2.31, p < .05$) and sentences (Trial 2: $t(23) = 2.65, p < .05$; Trial 3: $t(23) = 2.65, p < .05$).

Unrelated sentences and stories. Unrelated sentences and stories. The stories were recalled more accurately than the unrelated sentences in the experiment as a whole ($F(1, 23) = 25.15, p < .001$). The material by trial interaction was again significant ($F(1, 23) = 9.32, p < .001$), although this was apparently the result of ceiling effects for story recall in later trials. Bonferroni: t tests indicated that story recall was more accurate than sentence recall on every trial (Trial 1: $t(23) = 4.78, p < .001$; Trial 2: $t(23) = 3.86, p < .01$; and Trial 3: $t(23) = 2.38, p = .08$). Therefore, the method of linking list length to span for each material failed to equate the difficulty of recalling unrelated sentences and stories that were 50% above span length—presumably because the presence of a story structure did not facilitate the recall of surface structure (and so did not allow story span to exceed sentence span) but did enhance gist recall at supra-span lengths.

There was a significant material by dual task interaction ($F(1, 23) = 4.24, p < .05$), which indicated that the secondary task impaired recall of the unrelated sentences ($t(23) = 3.54, p < .01$) but not the stories ($t(23) < 1$).

Although the interpretation of this interaction is complicated by the fact that task difficulty was not equated for the two materials, the effect remained significant ($F(1, 11) = 12.29, p < .01$) for a group of subjects who did not show better story than sentence recall under single task conditions ($t(11) < 1$). This group comprised the 12 participants who, under single task conditions, showed the smallest difference between story and sentence recall. The three-way interaction between material, dual task, and trial did not reach significance in this experiment ($F(2, 46) < 1$).

Recall errors

Errors were coded as for Experiments 1 and 2. Table 1 shows the mean number of omission, morphological, phonological, semantic, repetition, and unrelated errors for words, sentences, and stories as a percentage of the number of items presented. Order errors are also shown, expressed as a percentage of total item recall. A series of t tests, corrected for multiple comparisons, was used to compare errors on words/sentences and sentences/stories. There were more omission ($t(23) = 3.58, p < .05$), morphological ($t(23) = 3.54, p < .05$), phonological ($t(23) = 4.80, p < .01$) and order errors ($t(23) = 4.11, p < .01$) for words compared with sentences. In contrast, semantic errors were more common in sentence than word recall ($t(23) = 7.26, p < .001$), possibly reflecting the greater role played by semantic coding in this task. There were more omission errors for sentences compared with stories ($t(23) = 5.49, p < .001$), but no other significant error differences between these two conditions. All of these effects reached significance in non-parametric tests (Wilcoxon's signed ranks).

Reaction times

RT was recorded during two periods in each trial: before the words began to play (baseline) and while they were playing (encoding). Mean RT is shown in Table 4. To facilitate comparison between experiments, RT during the encoding of unrelated words and sentences will be described first, followed by an analysis of RT during the sentences/stories tasks. There was a main effect of time period in the experiment as a whole ($F(2, 46) = 147.33, p < .001$), reflecting the RT cost of concurrent encoding.

Unrelated words and sentences. There was a main effect of material ($F(1, 23) = 28.64, p < .001$), indicating that in the experiment as a whole, reaction times were faster for sentences than for word lists. The material by trial interaction approached significance ($F(2, 46) = 2.59, p = .09$), apparently because RT increased across the three trials for word lists but not sentences. Bonferroni: t tests indicated that there was no significant RT difference between sentences and words on the first trial ($t(23) < 1$), whereas on Trials 2 and 3, RT was slower for the unre-

Table 4
Mean (*SD*) reaction times in milliseconds in Experiment 3

	Words		Sentences		Stories	
	Baseline	Encoding	Baseline	Encoding	Baseline	Encoding
Trial 1	430 (86)	643 (188)	453 (89)	588 (151)	424 (86)	548 (139)
Trial 2	439 (103)	701 (215)	418 (64)	590 (168)	431 (68)	535 (91)
Trial 3	432 (69)	740 (203)	413 (62)	596 (151)	426 (85)	555 (104)
Average	434	695	428	591	427	546

lated words ($t(23) = 2.87$, $p < .05$ and $t(23) = 4.03$, $p < .01$, respectively). This pattern is suggestive of a greater role for attention-demanding encoding processes in word recall following learning and is consistent with the larger dual task decrement in retention observed for words on later trials. The overall reaction time cost of encoding was greater for words than for sentences ($F(1, 23) = 30.07$, $p < .001$). In addition, the interaction between time period and trial in the data set as a whole was significant, reflecting the pattern observed for unrelated words ($F(2, 46) = 3.66$, $p < .05$).

Unrelated sentences and stories. There was a main effect of material ($F(1, 23) = 7.61$, $p < .05$) and a material by time period interaction ($F(1, 23) = 8.68$, $p < .01$), indicating that the RT cost of encoding was greater for unrelated sentences compared with stories. Bonferroni: t tests indicated that there was no difference in baseline RT between the sentences and stories ($t(23) < 1$). In contrast, RT during encoding was greater for the sentences relative to the stories ($t(23) = 3.20$, $p < .01$), suggesting that there may have been a trade-off between memory and RT performance. Participants' recall of the unrelated sentences was poorer than that of the stories despite their slower reaction times, suggesting that the unrelated sentences relied more heavily on attention-demanding processes. No other interactions reached or approached significance.

Discussion

This experiment compared the impact of a secondary CRT task on the recall of unrelated words, unrelated sentences, and stories composed of related sentences. In the two previous experiments, it was found that divided attention substantially impaired the recall of novel unrelated sentences but only impaired the recall of unrelated words following learning. The current experiment aimed to determine which aspects of the sentence recall task involved controlled processing. If the construction of mental models through semantically guided chunking is attention-demanding, then a concurrent task might be expected to impair story recall to a greater extent than sentence recall; if in contrast, divided attention specifically interferes with the binding of unrelated chunks, then a secondary task should have the greatest impact

on the retention of unrelated sentences. The results supported the second of these two possibilities. As in the previous experiments, divided attention was found to substantially impair the recall of unrelated sentences but this effect did *not* generalise to the recall of stories. The RT data were consistent with this pattern: the RT cost of encoding was larger for unrelated sentences than for stories.

The previous finding that divided attention impaired the recall of novel sentences to a greater extent than word lists did not quite reach significance in this experiment, although there was a trend in this direction. It is not clear why this effect was weaker in the current experiment. However, if divided attention specifically interferes with the binding of unrelated items, it might be expected that a concurrent task would have some effect on word list recall. Moreover, it is clear from Experiments 1 and 2 that divided attention *can* impair the recall of unrelated words in some circumstances (i.e., following learning). It seems likely that although the recall of unrelated sentences is more vulnerable to disruption by a secondary task than the recall of unrelated words, the size of this interaction may be sensitive to the nature of the words in the lists. This possibility, which is raised again in the General discussion, warrants further investigation.

The effect of a concurrent task on sentence recall: A closer look

Experiment 3 showed that a concurrent task specifically interfered with the maintenance of sentences that were unrelated in meaning, indicating a special role for attention in the binding of arbitrary combinations. The secondary task effect did not generalise to the recall of related sentences that formed a story, suggesting that the processes which elevated sentence span above word span (for example, the integration of different linguistic codes; semantically and syntactically guided chunking) were not themselves disrupted by divided attention. If this interpretation is correct, the concurrent task might be expected to cause forgetting of entire sentences in the absence of a story structure. To address this issue, each sentence that was presented to an individual subject

was classified as having been recalled entirely correctly, partially or not at all. The secondary task reduced the number of correctly recalled sentences (Experiment 1: $F(1,23) = 4.55$, $p < .05$; Experiment 2: $F(1,23) = 8.23$, $p < .01$; and Experiment 3: $F(1,23) = 3.02$, $p < .1$) apparently due to an increase in the number of sentences that were entirely forgotten (Experiment 1: $F(1,23) = 3.26$, $p = .08$; Experiment 2: $F(1,23) = 14.83$, $p < .001$; and Experiment 3: $F(1,23) = 2.70$, $p = .11$). There was no influence of the secondary task on the percentage of sentences that were partially recalled (Experiment 1: $F(1,23) = 2.26$, $p = .15$; and Experiment 2: $F(1,23) < 1$; and Experiment 3: $F(1,23) = 2.02$, $p = .17$). In the third experiment, the unrelated sentence and story conditions did not differ in terms of either completely correct or partial sentence recall ($F(1,23) = 2.80$, $p = .11$ and 2.05 , $p = .17$); however, unrelated sentences were more likely to be completely forgotten than those from stories ($F(1,23) = 5.82$, $p < .05$). These findings are consistent with the view that the recall of unrelated sentences is particularly sensitive to the effects of a concurrent task because divided attention disrupts the maintenance of several unrelated chunks.

General discussion

Three experiments examined the effect of a concurrent attention-demanding visual choice reaction time (CRT) task on the recall of auditorily presented stories, sentences, and lists of unrelated words, in order to examine the relative importance of controlled and automatic encoding processes for retention of these materials. List length was based on span for each participant and for each type of material, with the aim of equating recall under single task conditions. The pattern across the experiments was that on Trial 1, retention of words and unrelated sentences was equivalent when attention was undivided; when participants performed the CRT task concurrently however, the retention of sentences was markedly impaired whereas the retention of the unrelated words was relatively unaffected. In contrast to the findings for unrelated sentences, the secondary task had little effect on the recall of related sentences that formed a story (Experiment 3), suggesting that the processes that underlie the span advantage for sentences relative to words (for example, semantically/syntactically guided chunking and the integration of different linguistic codes) can operate relatively automatically. These processes are likely to play a larger role in the recall of stories than unrelated sentences (given that the meaningfulness of stories should encourage the formation of larger chunks) and yet the secondary task only impaired recall of the latter material. The dual task decrement for sentential material (composed of several short sentences) was affected by the semantic relatedness of the

sentences, suggesting that divided attention interfered with the ability to form links between arbitrary combinations of sentences (reducing the number of unrelated sentence chunks that participants were able to maintain simultaneously) rather than interfering with the process of within-sentence chunking per se.

Different types of errors occurred on the word and sentence span tasks, lending support to the notion of coding differences between the materials. Phonological errors were more frequent on the word lists and semantic errors were more frequent on the sentences, suggesting a greater involvement of phonological coding in word span and semantic coding in sentence span (Baddeley, 1966b; Conrad, 1964). There were also fewer order and morphological errors on the sentences, presumably because the sentence structure constrained both the order and the morphological marking of the words. In Experiment 2, the concurrent task influenced order memory for unrelated words but not sentences, in line with the view that controlled processing comes into play when arbitrary associations must be maintained, but this finding did not replicate across the experiments.

Over three trials in which the same material was repeated, the sentences were learned more quickly than the word lists. In addition, the concurrent reaction time task impaired the retention of unrelated words to a greater extent after they had been presented several times, suggesting that attention-demanding encoding processes made a more substantial contribution to word recall as learning increased. With multiple presentations, participants may have switched from using predominantly phonological encoding (which is likely to be relatively ineffective but have a low attentional cost) to the formation of multiword chunks; this would have enhanced recall considerably in the single task condition but would have increased participants' vulnerability to the effects of divided attention.

As the amount of material to be remembered was increased above span (Experiment 2), percentage recall declined less rapidly for sentences than for word lists. The greater involvement of long-term semantic/syntactic representations in the sentence span task presumably made WM for this material more robust to supra span increases in length. Word span may have been less robust because of its greater reliance on the limited capacity phonological loop. The retention of longer lengths of material was relatively unaffected by conditions of divided attention, possibly because participants were less likely to use attention-demanding strategies when the task was very taxing. In addition, the semantic/syntactic factors which underpinned the more substantial sentence recall advantage at longer list lengths were apparently operating largely automatically; consequently, at longer lengths, controlled processing may have played a relatively minor role in sentence recall.

Reaction times on the CRT task were consistent with this pattern. RT was similar for the words and unrelated sentences on the first trial, suggesting that the larger dual task decrement in retention observed for sentences did not result from a trade-off in memory and RT performance. The dual task decrement for unrelated sentences manifested itself in memory performance rather than in RT, in line with the results of Craik et al. (1996) who found that performing an RT task during encoding had a substantial effect on recall but a smaller effect on RT. RT was slower for unrelated sentences than for stories, even though stories were recalled more accurately, suggesting that the sentences were more effortful to encode. Over three trials in which the same material was repeated, RT increased for words but not sentences/stories. This finding is consistent with the greater dual task decrement in retention observed for word lists on later trials and again suggests that attention-demanding encoding processes played a greater role in the word span task as learning increased.

In summary, we have obtained evidence for a large role of automatic linguistic knowledge effects in sentence/prose recall, as well as a significant attentional contribution to the retention of unrelated chunks of information. The finding that the dual task decrement for unrelated sentences did not generalise to the recall of stories revealed that sentences/stories activate pre-existing semantic and syntactic representations in LTM relatively automatically as they are processed; word lists cannot activate these representations to the same extent, accounting for the span advantage for prose. This type of automatic support for sentences is envisaged by connectionist models that view verbal STM as arising directly from the representations involved in language processing (MacDonald & Christiansen, 2002; Martin & Saffran, 1992, 1997; Martin, Saffran, & Dell, 1996).

Divided attention interfered with the ability to rapidly form links between arbitrary combinations of sentences, reducing the number of unrelated sentences that participants were able to maintain simultaneously, but did not interfere with processes such as (1) syntactic parsing or (2) the creation of mental representations of sentence meaning. There is some convergent evidence to suggest that these processes operate relatively automatically. (1) Waters, Caplan, & Rochon (1995) found that divided attention impaired comprehension of sentences that contained a large number of propositions, but did not impair syntactically complex sentences to a greater extent than simple ones. In addition, patients with Alzheimer's disease who have executive deficits are largely unimpaired at comprehending syntactically complex sentences but show greater difficulty for sentences containing multiple propositions (Kempler, Almor, Tyler, Andersen, & MacDonald, 1998; Rochon, Waters, & Caplan, 1994, 2000; Waters et al., 1995;

Waters & Rochon, 1998). It seems, therefore, that syntactic parsing is relatively automatic, whereas the maintenance/binding of several different propositions is effortful. Although we did not systematically manipulate number of propositions in the current study, the concepts in the unrelated sentence condition were more diverse than those in the stories. (2) Recent research suggests that the construction of models of sentence meaning proceeds relatively automatically. Radvansky & Copeland (2004) found that demanding WM span tasks did not predict participants' abilities to process the meaning of prose as measured by, for example, the detection of narrative inconsistencies. Interestingly, WM span did correlate with the ability to remember information that was poorly integrated into the story as a whole. In addition, mental model construction does not decline with age, despite older adults' poorer attentional control (Radvansky, 1999; Radvansky, Copeland, Berish, & Dijkstra, 2003; Radvansky, Zwaan, Curiel, & Copeland, 2001). However, older adults do show specific difficulty in retrieving nonintegrated facts (Radvansky, Zacks, & Hasher, 1996). These studies are consistent with our results suggesting that attention is critical for the encoding of unrelated sentences but not stories.

The differential role of attention in sentence and story recall is coherent with Cowan's (2000) suggestion of a basic capacity limitation of four unrelated chunks that can be held in the focus of attention at any one time: conditions of divided attention might be expected to reduce this capacity, preventing participants from forming new associations between unrelated sentences, and causing them to forget entire sentences. In contrast, retention of the stories may have been less reliant on the capacity to maintain several unrelated chunks because the meaningfulness of this material would have encouraged the formation of larger chunks, potentially spanning several sentences. It is also interesting to note the implications of these findings for the LT-WM model of Ericsson & Kintsch (1995). In some situations in which LT-WM is used, such as digit recall by memory experts, meaning is effortfully imposed on arbitrary strings of items in order to increase their memorability: this may be analogous to our participants' efforts to remember sequences of unrelated sentences. However, it appears that the recall of coherent prose does not rely on exactly the same processes despite the involvement of LT-WM (see Miyake & Shah, 1999; for a similar suggestion).

Why did the secondary task have little impact on word recall on the first trial, since these items were also unrelated in meaning? One possibility is that the specific conditions in the experiments (fast presentation speed and a relatively high proportion of function words with few semantic associations) encouraged the participants to rely primarily on a phonological code to retain the word lists; in contrast, span for the sen-

tence material extended far beyond the capacity of the phonological loop. As the words in the lists were unrelated in meaning, chunk size is likely to have been small and thus recall may not have been profoundly affected by chunk capacity. Although across the experiments, dual task interference effects broadly followed the pattern unrelated sentences > unrelated words > stories, it is important to note that in Experiment 3, the difference between the first two materials did not reach significance: the concurrent task may therefore have had a marginal effect on the retention of unrelated words. It would be interesting to establish whether divided attention has a more substantial effect on the recall of novel word lists that contain several semantically related clusters of words. Semantic grouping is expected to increase word span beyond the capacity of the phonological loop (Poirier & Saint Aubin, 1995) and in these circumstances, recall may be more reliant on the ability to maintain several unrelated chunks (in this case, groups of semantically related words). Whatever the role of attention in word span tasks, our findings suggest that for sentential material, attention-demanding processes are involved in the formation of links between, or simultaneous maintenance of, unrelated concepts/propositions but that there is also a major contribution from syntactic/semantic factors that operates relatively automatically.

Acknowledgments

Part of this work was carried out at the University of Bristol, UK, while the first author was supported by an Economic and Social Research Council studentship. The work was also supported in part by Medical Research Council Grant G9423916 awarded to Alan Baddeley.

Appendix

Examples of materials used in Experiment 1

Word list—50% above span
(9 words, presented to participants with word spans of 6)
Essay, marmalade, is, lots, clowns, wine, spaces, often, a

Sentences—50% above span
(20 words, presented to participants with sentence spans of 13)
Railway stations are noisy places. Guns can cause serious injuries. Water is boiled in kettles. Pink roses are pretty flowers.

Examples of materials used in Experiment 2

Word list—50% above span
(9 words, presented to participants with word spans of 6)
Are, found, plays, cream, the, from, people, is, telephone

Word list—100% above span
(12 words, presented to participants with word spans of 6)
Fine, of, cars, eggs, often, night, bark, pedestrians, kites, in, tasty, are

Word list—200% above span
(18 words, presented to participants with word spans of 6)
Like, are, is, fruit, learning, the, drive, keep, lunch, busy, from, collisions, doors, sell, contain, summer, be, laugh

Sentences—50% above span
(20 words, presented to participants with sentence spans of 13)
Planets travel around the sun. Marmalade is eaten for breakfast. Large hotels contain many rooms. Cars drive on the left.

Sentences—100% above span
(26 words, presented to participants with sentence spans of 13)
Shopping is an enjoyable activity. Water is boiled in kettles. Comics are fun to read. Film is needed to take photographs. Ancient buildings should be preserved.

Sentences—200% above span
(39 words, presented to participants with sentence spans of 13)
Muddy boots make carpets dirty. Climbing up mountains is tiring. Rubbish is placed in bins. Paper is sold in pads. Ripe fruit becomes soft. Shells are found on beaches. Farmers work in the country. Shoes are made from leather.

Examples of materials used in Experiment 3

Word list—50% above span
(9 words, presented to participants with word spans of 6)
town, into, loved, for, be, winning, on, mistakes, store

Sentences—50% above span
(18 words, presented to participants with sentence spans of 12)
The woman next door was angry. They drove to their new home. One large family had a picnic.

Stories—50% above span
(18 words, presented to participants with sentence spans of 12)
A teenage girl loved buying clothes. She went shopping with her mum. They travelled into town by bus.

References

- Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database [CD-ROM]. Linguistic data consortium*. Philadelphia, PA: University of Pennsylvania.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory?. *Trends in Cognitive Sciences*, 4, 417–423.
- Baddeley, A., & Wilson, B. A. (2002). Prose recall and amnesia: Implications for the structure of working memory. *Neuropsychologia*, 40, 1737–1743.
- Baddeley, A. D. (1966a). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, 18, 302–309.

- Baddeley, A. D. (1966b). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology*, *18*, 362–365.
- Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158–173.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.
- Baddeley, A. D., & Levy, B. A. (1971). Semantic coding and short-term memory. *Journal of Experimental Psychology*, *89*, 132–136.
- Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, *27*, 586–595.
- Baddeley, A. D., & Warrington, E. K. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *9*, 176–189.
- Bourassa, D. C., & Besner, D. (1994). Beyond the articulatory loop—a semantic contribution to serial order recall of subspan lists. *Psychonomic Bulletin & Review*, *1*, 122–125.
- Bransford, J., & Johnson, M. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 717–726.
- Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology*, *26*, 467–482.
- Britton, B. K., & Tesser, A. (1982). Effects of prior knowledge on use of cognitive capacity in three complex cognitive tasks. *Journal of Verbal Learning and Verbal Behavior*, *21*, 421–436.
- Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 141–189). Hillsdale, NJ: Erlbaum.
- Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology A*, *33*, 497–505.
- Conrad, R. (1964). Acoustic confusion in immediate memory. *British Journal of Psychology*, *55*, 75–84.
- Cowan, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.
- Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–99). Cambridge: Cambridge University Press.
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185.
- Craik, F., Govoni, R., Naveh-Benjamin, M., & Anderson, N. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*, 159–180.
- Doolling, D., & Lachman, R. (1971). Effects of comprehension on retention of prose. *Journal of Experimental Psychology*, *88*, 216–222.
- 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: Mechanisms of active maintenance and executive control* (pp. 102–134). Cambridge: Cambridge University Press.
- Epstein, W. (1961). The influence of syntactical structure on learning. *American Journal of Psychology*, *74*, 80–85.
- Ericsson, K., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.
- Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, *28*, 200–213.
- Gathercole, S. E., & Baddeley, A. D. (1990a). Phonological memory deficits in language disordered children: Is there a causal connection?. *Journal of Memory and Language*, *29*, 336–360.
- Gathercole, S. E., & Baddeley, A. D. (1990b). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology*, *81*, 439–454.
- Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*, 562–562.
- Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: An investigation of the knowledge-is-power hypothesis. *Cognitive Psychology*, *44*, 339–387.
- Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, *30*, 685–701.
- Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a reintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *23*, 1217–1232.
- Jefferies, E., Jones, R., Bateman, D., & Lambon Ralph, M. A. (in press). A semantic contribution to nonword recall? Evidence for intact phonological processes in semantic dementia. *Cognitive Neuropsychology*.
- Jefferies, E., Jones, R., Bateman, D., & Lambon Ralph, M. A. (2004). When does word meaning affect immediate serial recall in semantic dementia?. *Cognitive, Affective and Behavioral Neuroscience*, *4*, 20–42.
- Johnson-Laird, P. N., & Stevenson, R. (1970). Memory for syntax. *Nature*, *227*, 412.
- Kempler, D., Almor, A., Tyler, L. K., Andersen, E. S., & MacDonald, M. C. (1998). Sentence comprehension deficits in Alzheimer's disease: A comparison of off-line vs. on-line sentence processing. *Brain and Language*, *64*, 297–316.
- Knott, R., Patterson, K., & Hodges, J. R. (1997). Lexical and semantic binding effects in short-term memory: Evidence from semantic dementia. *Cognitive Neuropsychology*, *14*, 1165–1216.
- Kucera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short term memory. *Journal of Memory and Language*, *31*, 713–733.

- MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*, 35–54.
- Marks, L. E., & Miller, G. A. (1964). The role of semantic and syntactic constraints in the memorization of English sentences. *Journal of Verbal Learning and Verbal Behavior*, *3*, 1–5.
- Martin, N., & Saffran, E. (1992). A computational account of deep dysphasia: Evidence from a single case study. *Brain and Language*, *43*, 240–274.
- Martin, N., Saffran, E., & Dell, G. (1996). Recovery in deep dysphasia: Evidence for a relation between auditory-verbal STM capacity and lexical errors in repetition. *Brain and Language*, *52*, 83–113.
- Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, *14*, 641–682.
- Martin, R. C., & Lesch, M. (1996). Associations and dissociations between language impairment and list recall: Implications for models of STM. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 149–178). East Sussex, UK: Psychology Press.
- Martin, R. C., Shelton, J., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language*, *33*, 83–111.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits of our capacity for processing information. *Psychological Review*, *63*, 81–97.
- Miller, G. A., & Selfridge, J. A. (1950). Verbal context and the recall of meaningful material. *American Journal of Psychology*, *63*, 176–185.
- Milner, B. (1966). Amnesia following operation on the temporal lobes. In O. L. Zangwill (Ed.), *Amnesia* (pp. 109–133). London: Butterworth.
- Miyake, A., & Shah, P. (1999). Toward unified theories of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 442–481). Cambridge: Cambridge University Press.
- Patterson, K., Graham, N., & Hodges, J. R. (1994). The impact of semantic memory loss on phonological representations. *Journal of Cognitive Neuroscience*, *6*, 57–69.
- Poirier, M., & Saint Aubin, J. (1995). Memory for related and unrelated words: Further evidence on the influence of semantic factors in immediate serial recall. *Quarterly Journal of Experimental Psychology A*, *48*, 384–404.
- Potter, M. C., & Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, *29*, 265–282.
- Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, *38*, 265–282.
- Radvansky, G. A. (1999). Aging, memory and comprehension. *Current Directions in Psychological Science*, *8*, 49–53.
- Radvansky, G. A., & Copeland, D. E. (2004). Working memory span and situation model processing. *American Journal of Psychology*, *117*, 191–213.
- Radvansky, G. A., Copeland, D. E., Berish, D. E., & Dijkstra, K. (2003). Aging and situation model updating. *Aging, Neuropsychology and Cognition*, *10*, 158–166.
- Radvansky, G. A., Zacks, R. T., & Hasher, L. (1996). Fact retrieval in younger and older adults: The role of mental models. *Psychology and Aging*, *11*, 258–271.
- Radvansky, G. A., Zwaan, R. A., Curiel, J. M., & Copeland, D. E. (2001). Situation models and aging. *Psychology and Aging*, *16*, 145–160.
- Rochon, E., Waters, G. S., & Caplan, D. (1994). Sentence comprehension in patients with Alzheimer's disease. *Brain and Language*, *46*, 329–349.
- Rochon, E., Waters, G. S., & Caplan, D. (2000). The relationship between measures of working memory and sentence comprehension in patients with Alzheimer's disease. *Journal of Speech, Language, and Hearing Research*, *43*, 395–413.
- Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, *2*, 437–442.
- Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology*, *22*, 261–273.
- Tulving, E. (1962). Subjective organization in free recall of unrelated words. *Psychological Review*, *69*, 344–354.
- Vallar, G., & Shallice, T. (Eds.). (1990). *Neuropsychological impairments of short-term memory*. New York, NY: Cambridge University Press.
- Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*, 1256–1271.
- Waters, G. S., Caplan, D., & Rochon, E. (1995). Processing capacity and sentence comprehension in patients with Alzheimer's disease. *Cognitive Neuropsychology*, *12*, 1–30.
- Waters, G. S., & Rochon, E. (1998). Task demands and sentence comprehension in patients with dementia of the Alzheimer's type. *Brain and Language*, *62*, 361–397.