Working memory and binding in sentence recall

A.D. Baddeley a, G.J. Hitch a,*, R.J. Allen b

a Department of Psychology, University of York, York Y010 5DD, United Kingdom
b Institute of Psychological Sciences, University of Leeds, Leeds, LS2 9JT, United Kingdom

Article info

Article history:
Received 17 December 2008
revision received 13 May 2009
Available online 14 July 2009

Keywords:
Working memory
Binding
Episodic buffer
Sentence advantage

A B S T R A C T

A series of experiments explored whether chunking in short-term memory for verbal materials depends on attentionally limited executive processes. Secondary tasks were used to disrupt components of working memory and chunking was indexed by the sentence superiority effect, whereby immediate recall is better for sentences than word lists. To facilitate comparisons and maximise demands on working memory, materials were constrained by re-sampling a small set of words. Experiment 1 confirmed a reliable sentence superiority effect with constrained materials. Experiment 2 showed that secondary tasks of concurrent articulation and visual choice reaction impaired recall, but did not remove or reduce the sentence superiority effect. This was also the case with visual and verbal n-back concurrent tasks (Experiment 3), and with concurrent backward counting (Experiment 4). Backward counting did however interact with mode of presenting the memory materials, suggesting that our failure to find interactions between concurrent task and materials was not attributable to our methodology. We conclude that executive processes are not crucial for the sentence chunking advantage and we discuss implications for the episodic buffer and other theoretical accounts of working memory and chunking.

Introduction

The modular, multi-store view of human memory raises important and unresolved questions about how different subsystems interact to enable the coherent operation of the system as a whole. A central problem is how information held in working memory interacts with more permanent long-term knowledge. One aspect concerns the role of working memory in acquiring new knowledge. For example, phonological short-term storage appears to be important in learning new word forms but not new associations between familiar words (Baddeley, Gathercole, & Papagno, 1998), suggesting some degree of specialisation in the links from working memory to long-term memory. A different aspect concerns effects of previously acquired knowledge on the operation of working memory. For example, short-term memory span is greatly expanded when the input can be recoded into higher-order units or chunks using previously acquired knowledge (Miller, 1956). In the present paper we focus on short-term memory for sentences, which can be two or three times better than for random words (Brener, 1940), thus far outweighing the capacity limits normally associated with working memory. We investigate how working memory takes advantage of long-term knowledge of linguistic constraints on words and their order in sentences and in particular, whether forming chunks depends on executive processes.

The theoretical significance of Brener’s (1940) finding only began to emerge a decade later with the development of information theory by Shannon and Weaver (1949), and its application to immediate memory. Miller and Selfridge (1950) used Shannon’s guessing game technique to produce word sequences that ranged from random, through first order approximations (in which words reflected their frequency in natural language), through second and third up to sixth order of approximation, each increasing the similarity of the sequences to natural text. Miller and
Selfridge (1950) and later Marks and Jack (1952) found that immediate recall increased systematically with order of approximation.

In his classic paper, Miller (1956) offered an interpretation of these and related results in terms of his concept of chunking. He suggested that short-term memory has a capacity that is limited in terms of the number of chunks it can hold. Redundant material such as text allowed several words to be combined into a single chunk, hence allowing more words to be recalled. Tulving and Patkau (1962) provided further evidence for the chunking hypothesis in a study that used free recall, defining a chunk as a sequence of words recalled in the same order as their presentation. They found that, whereas number of words recalled increased with order of approximation to English, number of chunks remained constant.

The chunking hypothesis gained further support from analysis of the serial recall of letter sequences. Miller, Bruner, and Postman (1954) generated sequences of letters that approximated more or less closely to English word structure. When presented tachistoscopically they observed that higher orders of approximation to English led to more accurate report, a result they interpreted in perceptual terms. Baddeley (1964) showed that the redundancy effect was present even with lengthy exposure, interpreting the effect in terms of memory rather than perception. McNulty (1966) demonstrated that, as in the case for words, the number of chunks recalled remained constant, while the number of letters per chunk increased with order of approximation. Furthermore, memory for sequences of unrelated consonants is strongly influenced by phonological similarity between the names of the letters, suggesting that sequences are remembered in terms of individual subvocalised letters. However, when order of approximation increases and vowels are included, the correlation with phonological similarity becomes insignificant ($r = .05$) while letter-based predictability becomes an important factor ($r = .66$) (Baddeley, 1971). There is therefore abundant evidence that sequential redundancy within language has a powerful impact on immediate recall, and that it operates through the process of chunking.

Our present interest in chunking based on language habits was motivated by the need to address some of the limitations of the model of working memory proposed by Baddeley and Hitch (1974; see also Baddeley, 1986). According to this model, working memory consists of a limited capacity attentional system (the central executive) that interacts with temporary stores for different kinds of information (the phonological loop and the visuo-spatial sketchpad). The phonological loop is a store for speech-coded information that decays in the order of 2 s but can be refreshed by subvocal rehearsal. It provides a simple and coherent account of the limit on memory span for random sequences of verbal items and its sensitivity to factors such as word length, phonemic similarity and the suppression of articulation (see Baddeley, 2007). However, despite successfully explaining a range of phenomena in the immediate recall of lists, the vastly superior immediate recall of sentences is clearly beyond the time-based capacity of the phonological loop. In itself this is not surprising, in that the original model of working memory did not attempt to specify links with long-term memory. It nevertheless serves to illustrate the point that the model must be elaborated if it is to describe how previously acquired knowledge interacts with information in the phonological loop in processes such as chunking.

A related problem for the Baddeley and Hitch (1974) model of working memory was raised by reading span, the measure of working memory capacity devised by Daneman and Carpenter (1980). There is copious evidence that this measure predicts individual differences in cognitive performance across a range of activities from language comprehension to reasoning and acquiring skills in logic (Daneman & Merikle, 1996; Kyllonen & Christal, 1990; Kyllonen & Stephens, 1990; Turner & Engle, 1989). Reading span requires the subject to process a series of sentences and subsequently recall the final word of each. Given that the sentences grossly exceed the capacity of the phonological loop, the question again arises as to how the material is maintained within working memory. The need to provide an explanation for these examples involving natural language materials, taken together with a range of related problems, prompted Baddeley (2000) to propose a fourth component of working memory, the episodic buffer (see Fig. 1).

The episodic buffer is assumed to be a limited capacity store in which information from the short-term stores and long-term memory can be integrated into episodic chunks. It differs from long-term episodic memory in being temporary and being dependent upon attentional control by the central executive. The buffer is assumed to be accessible through conscious awareness, and to be multi-dimensional, reflecting its integrative function of drawing together related information from the senses, from working memory and from long-term memory. The capacity of the buffer is assumed to be limited in terms of the number of multi-dimensional chunks it can hold at any one time. It accounts for the sentence superiority effect in immediate recall in terms of the greater size of chunks for sentences as compared with word lists.

The episodic buffer thus emphasises the need for working memory to integrate information, in contrast to the focus of our previous research which was principally concerned with isolating and studying the subcomponents.
of the system. The proposed new component thus made the working memory model more compatible with other approaches such as those of Cowan (2005) and Engle (2002) who have tended to emphasise the executive and integrative aspects of working memory, rather than its subsystems. However, if the episodic buffer concept is to prove more than a convenient label for an area of ignorance, it is necessary to demonstrate its capacity to generate empirical data that will enrich and extend our understanding of working memory. The experiments that follow are part of a program that attempts to do this.

A key assumption of the revised model was that the central executive mediates access of information from the phonological loop and the visuo-spatial sketchpad to the episodic buffer (see arrows marked a in Fig. 1). The model also included a direct link between the phonological loop and long-term linguistic knowledge, justified by the data on learning novel word forms (Baddeley et al., 1998) and by reciprocal effects such as the superior immediate serial recall of words compared with non-words (Hulme, Maughan, & Brown, 1991) (see arrow b in Fig. 1). Thus, according to the model, memory span for sentences is enhanced by direct interactions between language knowledge and the phonological loop and by attention-demanding binding processes. The combination of these processes is assumed to result in the formation of chunks in the episodic buffer.

Both Cowan’s “embedded processes” model (Cowan, 1995, 1999) and the episodic buffer model (Baddeley, 2000) assign an important role to the interaction between attention and long-term memory. Cowan assumes an attentional focus comprising the currently active representations in long-term memory, surrounded by regions that have recently been activated but are not in the current focus of attention. Baddeley’s (2000) episodic buffer interpretation differs in assuming that the buffer comprises a system in which representations are both stored and manipulated; the buffer is assumed to have links to representations in long-term memory, but also to be an active system that allows the creative manipulation and restructuring of long-term representations by interacting with the central executive. At an empirical level however, it is currently hard to distinguish between this and Cowan’s embedded processes hypothesis. In both cases, the precise question of how attention and memory interact offers an important problem. The experiments that follow address this issue, with the aim of identifying constraints that may help the further development of these models, neither of which is currently specified in sufficient detail as to allow unequivocal predictions to be made.

To summarise, our concern is to investigate the processes underpinning the superior immediate serial recall of sentences as compared to random word lists. We have noted that sentence recall exceeds the capacity of the phonological loop and benefits from chunking based on linguistic knowledge. According to the revised model of working memory (Baddeley, 2000), chunks are stored in a limited capacity episodic buffer that requires attention for access. Similarly, Cowan (2005) assumes that attention is required to form new links in working memory in the process of chunking. Thus both accounts suggest that a secondary task requiring general attention should disrupt the chunking process and hence remove or at the very least reduce the sentence superiority effect. Support for the alternative hypothesis that chunking does not require attention would imply that these theoretical accounts require substantial revision.

We should perhaps make it clear that at the present stage we are not attempting to analyse the psycholinguistic processes such as syntax and semantics that underpin sentence comprehension. Interesting and important though these processes are, we assume that the role of working memory in chunking can be tackled as a general question without becoming embroiled in the microanalysis of sentence processing. We begin by discussing the problem of how to minimise the role of long-term semantic and episodic memory in immediate verbatim recall of sentences, so as to focus on the role of sequential redundancy in working memory. We propose as a tool, a new measure we term constrained sentence span. This involves a comparison between the retention of constrained sentences and the same words presented in random order. Dual task methods are then used to assess the extent to which the binding advantage conferred by the sentence form is dependent upon the various components of working memory.

### Constrained sentence span

The fact that span for natural sentences is much greater than for arbitrary word lists gives rise to problems of scaling when comparing the effects of experimental manipulations on these two types of materials. Moreover, memory for natural sentences is likely to vary greatly depending on sentence form and content and the knowledge base of the individual. We therefore decided to try to create a paradigm in which language-based sequential redundancy played an important role, but where sentence length and the variable contribution from semantic memory would be better controlled. We achieved this by making our sentence task much more like the standard memory span procedure in which a small number of items are repeatedly used in different orders. We assume that, by making the set of items readily available, we would emphasise the need to rely on order information. Furthermore, by selecting a set of items that were constantly re-used, we assumed that proactive interference would minimise the contribution of long-term episodic and semantic memory, forcing participants to rely upon temporary binding in working memory to underpin their recall. Experiment 1 tested the effectiveness of this material in reducing the difference between memory for sentences and non-sentential word sequences while maintaining a clear sentences advantage.

### Experiment 1

**Method**

**Participants**

Twelve undergraduate and postgraduate students from the University of York took part, receiving course credit or
payment. There were five males and seven females, with an age range of 21–36 years.

Materials
A single set of words was used to construct both sentences and lists. This consisted of 15 words, of which there were four adjectives (tall, sad, old, fat), four nouns (teacher, soldier, waiter, bishop), four verbs (meets, insults, follows, helps), and three function words (not, and, or). Function words were present in both constrained sentences and lists, and were classed as test words both in terms of sequence length and recall scoring. Each word was digitally recorded by a male English speaker for auditory presentation. Words in both conditions were presented individually, thus removing cues from intonation or co-articulation present in continuous speech.

A group of 16 sequences was created for each of the constrained sentence and list sets. Each sequence consisted of eight words (three nouns, three adjectives, one verb, one function word). The sentences were modelled on a telegraphic structure so that, while not strictly grammatically correct, they did have clear meaning and internal redundancy (e.g. tall soldier follows waiter and old sad teacher). Lists were constructed by adjusting the order of the words in each sentence by at least two levels of approximation (e.g. soldier tall waiter teacher follows and old sad), to ensure minimal meaning and redundancy.

Design and procedure
Each participant was tested in a single session lasting approximately 15 min. Presentation of stimuli was performed on a Macintosh computer, using a SuperCard program. Sentences and lists were presented in two counterbalanced blocks of two practice trials and 14 test trials.

Participants pressed the space bar to begin presentation of each sequence, following a 2 s delay. Auditory presentation occurred through headphones at a rate of one word per second, using the same digital sound files for sentences and lists. As soon as each sequence ended, participants attempted to recall as many words as they could in serial order. Responses were recorded on a Sony Mini-disc.

Results
Recall of each word was scored as correct if it was produced in the appropriate position relative to an adjacent recalled word. Absolute serial position was only taken into account for first and last words in a sequence; these were scored as correct if produced in those positions in the recall sequence. For example, if the words ABCDE were recalled as CBADE, only words D and E would be scored as correct (as, in the presented sequence, the first word was not C, B did not directly follow C, and A did not follow B). If the sequence BCDAE was instead recalled, all words except A would be scored as correct (as in this case, B and C were recalled in their original relative order).

The mean proportion of words correctly recalled as a function of serial position is plotted in Fig. 2. A 2 × 8 ANOVA revealed significant effects of item set, $F(1, 11) = 76.97$, $MSE = .04$, $p < .001$, $\eta_p^2 = .88$, serial position, $F(7, 77) = 26.99$, $MSE = .02$, $p < .001$, $\eta_p^2 = .71$, and the item set by serial position interaction, $F(7, 77) = 12.9$, $MSE = .01$, $p < .001$, $\eta_p^2 = .54$. Thus, there was a highly significant advantage for sentences over lists, particularly at serial positions 3–7. Further analyses (not reported here for brevity) showed that the difference between conditions was largely due to order errors.

Discussion
The results show a large sentence superiority effect despite the use of constrained materials and the absence of prosodic cues and co-articulation. However, from a methodological perspective, using the same length of sequence for the two conditions clearly risks floor and/or ceiling effects when combined with a demanding secondary task. For that reason, we opted in subsequent experiments to attempt to match overall probability of recall by reducing the sequence length for unrelated lists.

We began by using a memory span procedure to gain a rough estimate as to likely performance on three types of material. One comprised our constrained sentences, a second involved the same words presented in random order, while a third involved sentences of increasing length selected from newspapers. As expected, span for these natural or open sentences proved to be highly variable, presumably reflecting the wide range of difference in content and syntactic structure. Overall, however, our pilot study suggested that we might obtain a roughly equal probability of correct sequence recall if we chose random lists of six words, constrained sentences of eight words, and natural sentences of 16 words.

Experiment 2 combined immediate retention of each of these three item sets with concurrent tasks designed to disrupt different components of working memory. While accepting that problems might arise in comparing 16-item open sentences with six-item word lists, the experiment did allow us to test our assumption that our constrained sentences were qualitatively similar to more naturalistic material. In order to enhance the ecological validity of our constrained sentences, we moved away from the somewhat artificial telegraphic structure of those used in Experiment 1 by including more function words. We also made our word lists more list-like by not allowing them to contain function words (excluding function words from the scoring of sentence recall in order to maintain the matching of items across sets).

Finally, in order to gain some information on the probable contribution of gist-based episodic long-term memory, we included a delayed recognition test in which participants attempted to identify sequences they had previously encountered. We predicted that delayed recognition would be substantially greater for open sentences than for lists, and hoped that our constrained sentences would resemble lists in this regard, rather than open sentences, suggesting a much-reduced role of semantic gist.

Our study thus focused on the immediate recall of open or natural sentences, constrained sentences, and lists comprising an arbitrary re-ordering of the content words used in constrained sentences. We varied sequence length across materials in an attempt to equate probability of
correct recall and studied the impact on recall and error patterns of disrupting working memory using a secondary task procedure. The secondary tasks were (i) articulatory suppression, assumed to disrupt the phonological loop while making minimal demands on attention, (ii) a demanding four-choice continuous reaction time task based on that used successfully by Craik, Govoni, Naveh Benjamin, and Anderson (1996), assumed to disrupt the central executive and the visuo-spatial sketchpad, and (iii) the combination of suppression and choice reaction time.

We predicted that all three concurrent task conditions would interfere with memory for each type of material, consistent with involvement of the phonological loop and the executive in the immediate recall task. We assumed that participants would form chunks for lists as well as sentences, the principal difference being that chunks would be larger, multi-item units for sentences. At issue was whether concurrent tasks loading the executive would reduce or even remove the sentence advantage in recall. The hypothesis that chunking processes depend on controlled attention predicts that an executive load will cause more items to be lost in sentence than list recall because of the difference in presumed chunk size for the two types of materials.

**Experiment 2**

**Method**

**Participants**

Twenty-four undergraduate and postgraduate students from the University of York took part, receiving course credit or payment. There were seven males and 17 females, with an age range of 18–34 years.

**Materials**

Four groups of eight sequences were created for each of the open sentence, constrained sentence, and constrained list item sets, one for each of four concurrent task conditions. Open sentences were drawn from the www.guardian.co.uk and www.independent.co.uk online news resources. They consisted of 16 test words, plus a selection of definite and indefinite articles ‘the’, ‘a’, ‘an’, prepositions ‘of’, ‘at’, ‘in’, ‘for’, ‘on’, ‘by’, ‘to’, and the conjunction ‘and’. Open sentences were selected to sample a broad range of commonly known subject matter. Constrained sentences and constrained lists were constructed from a limited pool of twelve nouns, twelve adjectives, four verbs, and four adverbs. All words were of medium–high Kucera and Francis (1967) frequency (mean 199.55). Constrained sentences contained eight test words comprising three nouns, three adjectives, one verb and one adverb. As for open sentences, they also contained function words that were not included in the word count. Constrained lists were six words long, and had a similar distribution of nouns, adjectives, verbs, and adverbs to the constrained sentences but were non-sensical, lacked function words, and had no syntactic or semantic structure. Examples of each of the three item sets are provided in Table 1 with details of the sentence generation procedure given in the Appendix. All sequences were recorded digitally in a male English speaking voice using natural intonation.

For the delayed recognition test, two sequences were selected from each group of eight. One of the two was used as a ‘true’ item, while the other was altered and served as a ‘false’ item. Open sentences were altered by exchanging between one and three content words for new words, in order to change the underlying meaning of the sentence. Constrained sentences and lists were altered by replacing two content words with other experimental words from the same class (i.e. noun for noun, verb for verb, etc.). This procedure provided four true and four false items for each of the three item sets, distributed evenly across the four concurrent task conditions.

**Design and procedure**

Each participant was tested in a single session lasting approximately 50 min. The main part of the experiment examined prose and word list recall under different
concurrent task conditions. A $4 \times 3$ repeated measures design was used, manipulating concurrent task condition (undivided attention; articulatory suppression; visuo-spatial continuous reaction time task; suppression + visual task) and item set (open sentences; constrained sentences; constrained lists). Auditory presentation of memory stimuli and presentation and response recording in the visuo-spatial task was performed on a Macintosh computer, using a program written in HyperCard.

Pre-test of visuo-spatial task. A measure of baseline performance in the visuo-spatial task was obtained by giving 25 practice trials followed by 200 test trials. In each trial, four squares each measuring 6 × 6 cm were presented horizontally across a white background, with a black circle (diameter 1 cm) in the centre of one of these spaces. Participants responded by pressing one of four keys on the keyboard (‘z’, ‘x’, ‘c’, ‘v’), corresponding to the square in which the target circle was positioned. Immediately upon the participant’s response, the visual display was reset and the circle appeared in one of the three remaining squares, selected at random. The computer recorded reaction times and error rates, and participants were instructed to be both as quick and as accurate as possible. No feedback was provided at any point.

Memory for verbal sequences. Completion of the visuo-spatial pre-test was followed by the central part of the experiment. Conditions were blocked so that eight sequences from each of the three item sets were successively presented under each concurrent task. The same blocking was repeated for each concurrent task condition in turn. The orders of both concurrent task condition and item set were counterbalanced using Latin square designs. The four groups of sequences for each type were used in the same order for all participants, ensuring that each group featured in each concurrent task condition an equal number of times.

Undivided attention (No task). Each trial commenced with participants pressing the space bar. After a 1 s delay, the memory stimulus was presented through speakers positioned on either side of the computer. Participants were instructed to listen to the words and upon completion repeat them back in the order in which they were presented. Responses were recorded using a Sony minidisk recorder.

Articulatory suppression (AS). The same procedure was followed with the addition that participants repeatedly articulated ‘1-2-3-4’ from the point of pressing the space bar to the end of stimulus presentation. They were instructed to speak at a constant rate demonstrated by the experimenter (approximately two utterances per second), and at a suitably modest volume to ensure that perception of the auditory input was not disrupted.

Visuo-spatial task. The visuo-spatial continuous reaction time task was implemented as described earlier. Participants performed four trials and the fourth response triggered presentation of the memory sequence. They were instructed to continue responding in the visual task as quickly and as accurately as possible throughout presentation and recall of the memory sequence and to halt immediately upon completion of recall. As in the other conditions, participants began each trial at their own pace by pressing the ‘s’ key (to reset the experimental program). Reaction times and error rates from the visual task were recorded, and coded by the computer as having occurred during the introductory trials, memory stimulus presentation, or recall.

Articulatory suppression + visuo-spatial task (AS + V). The final condition was a combination of those previously described. Participants listened to and repeated back a sentence or list while simultaneously performing the visuo-spatial task. In addition, they were also required to repeatedly articulate the sequence ‘1-2-3-4’ from the first visual trial through to completion of sentence/list presentation. Thus, participants performed both suppression and the visual task during presentation of the memory stimuli, with only the latter also continuing through recall.

Post-test of visuo-spatial task. A second baseline measure of performance on the visual task was obtained following completion of all the verbal recall conditions. This was a repeat of the pre-test with the 25 practice trials removed, thus leaving 200 test trials. Pre- and post-test data were then combined to provide an overall estimate of baseline performance.

Delayed recognition test. In the final segment of the experiment, participants were given a two-page A4 booklet containing 24 sentences or word lists. Twelve of these were used in the experiment, with the remaining twelve being other experimental items with a few constituent words altered (see Materials). The same 24 items were presented to all participants, in the same randomised order. Participants were instructed to read each item and record a yes/no decision as to whether it had been presented in the experiment in the exact form displayed together with a confidence rating of 1–5.

Results and discussion

Sentence and word list recall

Recall was scored using the procedure described in Experiment 1, with the modification that function words were ignored. For example, for the sentence LUCY the OLD PILOT RAPIDLY BORROWED the SMALL RED BOOK, a response of “Lucy the wealthy pilot rapidly stole the red old book” would score four for the italicised items. The mean number of words correctly recalled, collapsed across concurrent tasks, was 9.88 ($SD = 1.77$) for open sentences, 6.74 ($SD = .57$) for constrained sentences, and 4.82 ($SD = .62$) for lists. Thus, many more words were correctly recalled for open sentences than for constrained sentences, $t(23) = 10.52, p < .001$, effect size (Cohen’s $d$) = 1.79, or for lists, $t(23) = 16.65, p < .001, (d = 3.07)$, with absolute recall

Table 1

<table>
<thead>
<tr>
<th>Condition</th>
<th>Example of items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open sentence</td>
<td>CAR HEADLIGHTS THAT CAN HELP MOTORISTS SEE ROUND CORNERS WILL FINALLY BE INTRODUCED SOMEWHERE IN THE NEXT YEAR</td>
</tr>
<tr>
<td>Constrained sentence</td>
<td>LUCY the OLD PILOT RAPIDLY BORROWED the SMALL RED BOOK</td>
</tr>
<tr>
<td>Constrained List</td>
<td>JOHN CAR INSTANTLY LARGE BORROWED WHITE</td>
</tr>
</tbody>
</table>
also higher for constrained sentences than for lists, \( t(23) = 31.98, p < .001, (d = 3.20) \). Having confirmed our assumption that natural sentences would be easier to recall than constrained sentences, which would in turn be easier than lists, analysis of the effects of the concurrent tasks focused on proportion correct, taking advantage of our attempt to match for overall item difficulty by varying sequence length.

The mean proportions of words correctly recalled for each item set and concurrent task condition are displayed in Fig. 3. A 4 × 3 repeated measures ANOVA on probability of correct recall showed a significant effect of item set, \( F(2, 46) = 108.88, \text{MSE} = .01, p < .001, \eta^2_p = .83 \), indicating that we had underestimated the difficulty of the open sentences. The concurrent task effect was also significant, \( F(3, 69) = 41.26, \text{MSE} = .01, p < .001, \eta^2_p = .64 \), but the interaction between item set and concurrent task was not, \( F(5, 115) = 1.61, \text{MSE} = .003, p = .149, \eta^2_p = .07 \). In separate analyses of each of the concurrent tasks, the effect of visuo-spatial task, \( F(1, 23) = 11.97, \text{MSE} = 1.92^{–03}, p < .01, \eta^2_p = .34 \), and articulatory suppression, \( F(1, 23) = 77.16, \text{MSE} = 3.58^{–03}, p < .001, \eta^2_p = .77 \), were both significant. Furthermore, the interaction between articulatory suppression and item set was significant, \( F(2, 46) = 3.73, \text{MSE} = 1.87^{–01}, p < .05, \eta^2_p = .14 \), reflecting a slightly larger effect of suppression on the two constrained item sets, relative to open sentences. However, a significant interaction was not observed between item set and visuo-spatial task, \( F(2, 46) = .54, \text{MSE} = 1.35^{–03}, \eta^2_p = .02 \). Thus there were clear interference effects in the various dual-task conditions but, importantly, none of them had a selective effect on memory for sentences as compared with memory for lists.

Error analysis

If, as we suggest, our constrained sentences gain their advantage from the same processes as benefit natural sentence recall, then we might expect this to be reflected in the pattern of errors. Despite being closer in span size to lists, constrained sentences should show closer similarities to the longer natural sentences.

Errors were categorised as order errors in which items from the sequence were reported in the incorrect position, omissions, intra-experiment intrusions, and semantic errors. The distribution of these various error forms is shown in Table 2, with order errors also displayed in Fig. 4. Derivational, phonological and extra-experimental intrusions were noted, but were considered too infrequent for analysis, with rates of <.03 in all conditions.

Order errors. To avoid confounding differences in order memory with differences in levels of overall item memory, the rate of order errors is reported as a proportion of the total number of words recalled, regardless of order (Murdock, 1976). A 4 × 3 repeated measure ANOVA revealed significant effects of item set, \( F(2, 46) = 27.29, \text{MSE} = 3.76^{–01}, p < .001, \eta^2_p = .54, \) concurrent task condition, \( F(3, 69) = 5.45, \text{MSE} = 9.66^{–04}, p < .01, \eta^2_p = .19 \), and a significant set by task interaction, \( F(6, 138) = 2.97, \text{MSE} = 8.90^{–04}, p < .01, \eta^2_p = .11 \).

As Fig. 4 suggests, both open and constrained sentences showed a lower proportion of order errors than lists, for open sentences, \( t(23) = 4.78, p < .001, (d = 1.10) \), and for constrained sentences, \( t(23) = 5.99, p < .001, (d = 1.17) \), with order errors being slightly more frequent for open than constrained sentences, \( t(23) = 2.44, p < .05, (d = .45) \). This pattern suggests that retention of order is facilitated by language-based redundancy, with this effect being at least as great for the constrained as for the open sentences, hence supporting the view that the constrained and open sentences reflect a similar serial order mechanism. The task by item set interaction reflects a clear effect of the two conditions involving articulatory suppression that is most marked for the lists, suggesting that the capacity for subvocal rehearsal is particularly important in the absence of language-based sequential redundancy.

Omissions. These comprise instances when a presented item is absent from recall, without an alternative response being made. As shown in Table 2, omissions are
particular prominent for open sentences. However, as Fig. 3 shows, proportional recall of sentences was relatively low, suggesting that this observation should be interpreted with caution. Analysis indicated an overall effect of item set, $F(2, 46) = 137.12$, $MSE = .01$, $p < .001$, $\eta^2_p = .86$, and concurrent task, $F (3, 69) = 11.67$, $MSE = 3.50^{-03}$, $p < .001$, $\eta^2_p = .34$, but no significant interaction, $F(6, 138) = 1.27$, $MSE = 2.06^{-03}$, $p = .278$, $\eta^2_p = .05$. The main effect of concurrent task again appears to indicate principally a reduction of performance under articulatory suppression, although in this case the effect is broadly similar across conditions.

*Semantic errors.* These comprise intrusions with a similar meaning but distinct phonology to the target (e.g. police for detectives, or removed for withdrawn). As Table 2 shows, semantic errors are far more common in the case of open sentences. This could reflect a greater reliance on semantic gist with open sentences, but might also occur because the limited item set in the constrained conditions allow subjects to edit out such errors from their responses. ANOVA indicated a significant effect of item set, $F(2, 46) = 113.91$, $MSE = 1.20^{-03}$ $p < .001$, $\eta^2_p = .83$, and a small effect of concurrent task, $F(3, 69) = 3.37$, $MSE = 2.83^{-04}$ $p < .05$, $\eta^2_p = .13$, which again seems to reflect a slightly higher error rate under concurrent articulatory suppression.

*Intra-experimental intrusions.* These comprise intrusions from sequences presented previously in the experiment. They were slightly more frequent for constrained sentences than open sentences or lists, particularly for the suppression and visual + suppression conditions. ANOVA indicated a significant effect of item set, $F(2, 46) = 23.49$, $MSE = 2.50^{-03}$ $p < .001$, $\eta^2_p = .51$, concurrent task, $F(3, 69) = 3.69$, $MSE = 1.09^{-03}$ $p < .05$, and the interaction, $F(6, 138) = 4.99$, $MSE = 7.63^{-04}$ $p < .001$, $\eta^2_p = .18$. Again, this interaction is difficult to interpret due to the generally low rate of this error type.

Finally, it is notable that while articulatory suppression clearly impaired performance, the concurrent visual task had little effect on either the frequency or pattern of errors, suggesting that the effects observed were mainly attributable to disruption of the phonological loop. Overall, the error analysis supports the proposal that constrained sentences, although much shorter, have much in common with open sentences particularly in respect to the impact of redundancy on order errors, while the limited response set tends to reduce the frequency of semantic errors. A similar pattern of errors was also found in our subsequent experiments, but in the interests of brevity these data are not reported.

*Visuo-spatial task performance.*

Performance on the visuo-spatial task is shown in Table 3. Outliers were removed from the reaction time data by excluding any points over three standard deviations above the mean in a given condition. Reaction times were collapsed across pre- and post-test sessions, resulting in a mean reaction time of 387 ms ($SD = 44$), and an error rate of 5.6% ($SD = 3.86$). Visuo-spatial task latencies, when

---

**Table 2**

<table>
<thead>
<tr>
<th></th>
<th>Open sentences</th>
<th>Constrained sentences</th>
<th>Semantic errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>No task</td>
<td>.05 (.05)</td>
<td>.01 (.01)</td>
<td>.06 (.04)</td>
</tr>
<tr>
<td>Visual CRT</td>
<td>.05 (.03)</td>
<td>.01 (.01)</td>
<td>.06 (.03)</td>
</tr>
<tr>
<td>AS</td>
<td>.05 (.02)</td>
<td>.01 (.01)</td>
<td>.07 (.04)</td>
</tr>
<tr>
<td>Visual CRT + AS</td>
<td>.06 (.05)</td>
<td>.01 (.01)</td>
<td>.08 (.04)</td>
</tr>
</tbody>
</table>

Fig. 4. Proportional order error rates for each item set and concurrent task condition in Experiment 2 (AS = Articulatory Suppression).
performed as a concurrent task are displayed in Table 3. It is apparent that, relative to the mean baseline latency, reaction time was slowed when encoding the to-be-remembered sequences, and was even more impaired during recall. Articulatory suppression slowed reaction time further, during both encoding and recall, with one exception, namely in the open sentence recall condition. A $3 \times 2 \times 2$ repeated measures ANOVA confirmed these observations, with significant effects of item set, $F(2, 46) = 16.94$, $MSE = 2560$, $p < .001$, $\eta_p^2 = .42$, suppression, $F(1, 23) = 11.21$, $MSE = 3029$, $p < .01$, $\eta_p^2 = .33$, and task stage, $F(1, 23) = 153.03$, $MSE = 3854$, $p < .001$, $\eta_p^2 = .87$.

This pattern of results is basically as expected based on studies of LTM by Craik et al. (1996) and Naveh-Benjamin, Kilb, and Fisher (2006) who found effects on a concurrent RT task to be greater during retrieval than encoding. Further slowing occurs when an additional task, articulatory suppression, is required. These effects do not differ between list and constrained sentence recall, with the one anomaly occurring for the retrieval phase with open sentences. It is unclear whether this reflects the greater sentence length, or possibly a different pattern when gist becomes important.

Delayed recognition of sentences and lists

This final task was included to throw light on the episodic LTM contribution to performance across the three sets of items. Non-parametric A’ was used as the recognition measure; mean scores were $.90$ ($SD = .09$) for open sentences, $.50$ ($SD = .28$) for constrained sentences, and $.56$ ($SD = .28$) for constrained lists. Recognition was significantly higher for open sentences than either constrained sentences, $t(23) = 7.30$, $p < .001$, $(d = 1.73)$, or lists, $t(23) = 5.62$, $p < .001$, $(d = 1.61)$, but did not significantly differ between the latter two sets, $t(23) = .65$. Indeed, further analysis revealed that delayed recognition was not significantly above chance (.50) on constrained sentences or lists, $t(23) = .03$, $p = .976$, and $t(23) = 1.05$, $p = .306$, respectively.

While these results should be interpreted with caution, they are consistent with our initial assumption that constrained sentence recall would be much less likely to reflect a major contribution from episodic LTM than would open sentences. This may well reflect pro-active interference from successive sequences of highly similar material, an effect that is marked in the Peterson and Peterson (1959) short-term forgetting paradigm (Keppel & Underwood, 1962; Wickens, 1970).

To summarise, Experiment 2 showed that performance was greatest for the open sentences and least for the lists, with constrained sentence recall at an intermediate level, as would be expected given chunking based on language habits. There was a small but significant effect of the visual RT task and a more substantial effect of articulatory suppression. However, these effects did not consistently interact with material type, suggesting that the factors responsible for the advantage attributable to sentential redundancy were not influenced differentially by the demand imposed by our simultaneous tasks.

The pattern of errors for the open sentences showed a preponderance of omissions, together with a significant number of semantic intrusions, presumably reflecting a reliance on gist in this condition. Such errors were extremely rare in the other conditions, possibly reflecting a shallower level of semantic coding. The contrast in the final recognition performance between the good recognition of open sentences and chance performance on the constrained and random sets is consistent with this interpretation. The use of a closed word pool for constrained sentences and lists may also have enabled the exclusion of extra-experimental phonological forms from recall.

Overall, our results offer further support for the use of constrained sentences as a tool for analysing the effects of language-based redundancy on serial recall. Constrained sentences once again show a clear advantage over lists, indicating a robust effect of sentential redundancy in this case when sequence lengths were unequal and chosen to give broadly similar proportions of correct recall. However, given that level of performance on constrained sentences and lists in baseline conditions was approximately 90%, it could be argued that our results may have been constrained by ceiling effects. In addition, our selection of concurrent tasks did not manipulate the load on different components of working memory very systematically.

Experiment 3 therefore used a more systematic design, employing an adaptation of the n-back technique (Jonides et al., 1997) to devise a balanced set of four secondary tasks. The n-back method involves tracking a series of stimuli at a lag specified by the parameter n, allowing us to manipulate size of lag and type of stimulus orthogonally in a $2 \times 2$ design. Thus, we tested at 0-back (essentially, simple shadowing) and 2-back (shadowing at a lag of two items), and we did this separately for verbal and visuo-spatial stimulus sequences. Two-back has been shown to be a highly demanding executive task that neuroimaging studies have found to depend on both the frontal regions associated with executive control, and the more posterior cortical regions associated with storage (Owen, McMillan, Laird, & Bullmore, 2005; Smith & Jonides, 1999). Using the n-back as a concurrent task allows a

### Table 3

<table>
<thead>
<tr>
<th></th>
<th>Visual only</th>
<th>Visual + AS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Presentation</td>
<td>Recall</td>
</tr>
<tr>
<td>Open sentences</td>
<td>424.09 (48.29)</td>
<td>568.60 (96.12)</td>
</tr>
<tr>
<td>Constrained sentences</td>
<td>415.46 (53.86)</td>
<td>481.41 (69.40)</td>
</tr>
<tr>
<td>Constrained lists</td>
<td>411.79 (47.07)</td>
<td>486.39 (77.09)</td>
</tr>
</tbody>
</table>
separation of the functioning of the three components of the original working memory model, with the difference between baseline and 0-back conditions reflecting the impact of the phonological and visuo-spatial subsystems, and the difference between 0- and 2-back conditions indicating the further influence of the central executive. We used the recall variant of the n-back task, which Kane, Conway, Miura, and Colflesh (2007) have indicated as being a more appropriate method of tapping working memory processing than recognition n-back.

In this and subsequent studies we used only constrained sentences and constrained word lists. This choice avoids the large differences in levels of performance associated with open sentences, allowing us to detect a clear effect of redundancy based on language habits without introducing complexity due to sequence length, varied semantic content and an additional substantial contribution from episodic LTM.

As before we expected firstly a general pattern of secondary task interference common to sentences and lists, reflecting the role of working memory in immediate serial recall of verbal materials. Thus, memory for both types of materials should depend principally on the phonological loop and the central executive and much less so the visuo-spatial sketchpad, generating predictions that verbal n-back would more interfering than visuo-spatial n-back, and 2-back more than 0-back. As before, we expected sentences to be recalled better than lists and we were interested in whether loading the central executive would remove or reduce this sentence superiority effect.

Experiment 3

This experiment compared immediate serial recall of constrained sentences and random word lists under baseline conditions and combined with 0-and 2-back versions of our verbal and visuo-spatial tasks.

Method

Participants

Twenty-four undergraduate and postgraduate students from the University of York took part, receiving course credit or payment. There were 10 males and 14 females, with an age range of 18–31 years.

Materials

Five sets of 10 sequences were created for each of the constrained sentence and constrained list sets, using the same word pool as Experiment 2. Sentences and lists were constructed in the same way as before, with the exception that they were nine and seven test words long respectively (as opposed to eight and six), to avoid possible ceiling effects.

Design and procedure

Each participant was tested in a single session lasting approximately 1 h. Stimulus presentation and recording of concurrent task responses was performed on a Macintosh computer, using a HyperCard program. Each session started with 60 practice trials on each of the four n-back tasks (visuo-spatial and verbal 0- and 2-back), performed in a counterbalanced order, before moving on to the sequence recall task.

Visuo-spatial n-back. A 3 × 3 grid was presented on screen, each grid square measuring 4 × 4 cm. On each trial, a 1 cm diameter black circle appeared in one of the nine locations and remained present until a response was made or the trial timed out (after 3 s). The next trial followed immediately, with the circle appearing in one of the eight remaining locations, chosen at random. In the 0-back condition, participants had to respond by pressing a key on the 3 × 3 numerical keypad corresponding to the on-screen location presently occupied. In the 2-back condition, the task was to respond to the location occupied two trials previously. In this condition, participants were presented with two preparatory trials before each practice and test block, to enable a 2-back response on the first trial proper. In the preparatory trials, the word ‘Wait’ appeared below the 3 × 3 grid. Participants were instructed to let each of these trials time out and not respond. On the third trial, they attempted to respond with the location occupied in trial one, and so on for subsequent trials.

Verbal n-back. On each trial a random digit was presented at the centre of the screen in 30 pt Arial bold font. In the 0-back task, participants were instructed to read aloud the presented number, while in the 2-back task, they were to articulate the number presented two trials previously. Again, two extra ‘Wait’ trials were added before each block, with participants not responding until the third trial. As soon as a verbal response was produced, the experimenter pressed a key, moving the program on to the next trial and immediately prompting presentation of the next digit.

Sequence recall. Each of the five sets of ten sentences and lists were used equally often in each concurrent task condition. These conditions were blocked so that each concurrent task was performed on both sentences and lists, before moving on to the next task. The order of item set and concurrent task conditions was counterbalanced across the experiment, with the exception of the baseline (undivided attention) condition. Five sentences and five lists were presented for recall under baseline conditions at the start of the memory phase, with the remaining five from each set presented following completion of the concurrent task conditions, to allow an estimation of any practice effects. Each of the four concurrent task conditions was presented in between the two halves of the baseline condition (visuo-spatial 0-back/2-back, and verbal 0-back/2-back), with 10 sentences and 10 lists for each.

On each trial in the baseline condition, a mouse click by the participant triggered presentation of the verbal sequence through headphones, following a 2-s delay. As soon as the sequence had finished, participants attempted to recall as many words as they could in serial order, before moving on to hear the next trial. In the remaining conditions, presentation of the verbal sequence was preceded by four concurrent task trials (plus two ‘Wait’ trials for the 2-back conditions), with response on the fourth trial triggering presentation. Participants were instructed to continue performing the visuo-spatial or verbal concurrent
task as quickly and accurately as possible, while listening to the item being presented. On completion of sentence or list presentation the screen turned blank and recall was attempted. Thus, no additional tasks were performed during recall of the items. Participants clicked a mouse to start the next set of concurrent task trials. Responses in the primary and secondary tasks were recorded on a Sony Mini-disc recorder.

Results

Sentence and word list recall
Recall was scored using the procedure described in Experiment 2. As expected, sentential redundancy had a beneficial impact on the mean numbers of words recalled. In the baseline condition these were 7.20 (SD = 1.04) for sentences and 5.18 (1.22) for lists, a highly significant difference, t(23) = 14.25, p < .001 (d = 1.74). The mean proportions of words correctly recalled for each item set and concurrent task condition are displayed in Fig. 5. A 2 × 5 repeated measures analysis of variance on proportion correct revealed significant effects of item set, F(1, 23) = 43.65, MSE = .01, p < .001, ηp² = .66, and concurrent task, F(4, 92) = 84.52, MSE = .01, p < .001, ηp² = .79, but importantly, no interaction, F(4, 92) = .71, MSE = .01. Further analyses revealed that recall in the visuo-spatial 0-back condition was significantly poorer than in the baseline condition, F(1, 23) = 10.30, MSE = 3.08, p < .01, ηp² = .31, but significantly better than the visuo-spatial 2-back condition, F(1, 23) = 8.20, MSE = .01, p < .01, ηp² = .26. Recall in the verbal 0-back condition was also significantly poorer than in the baseline condition, F(1, 23) = 175.52, MSE = 3.66, p < .001, ηp² = .88, and significantly better than in the verbal 2-back condition, F(1, 23) = 45.32, MSE = .02, p < .001, ηp² = .66.

Error analyses
The different error types were scored and analysed using the method described in Experiment 2. Derivational, semantic, phonological, and extra-experimental intrusions all occurred at proportional rates of less than .01 and so were insufficiently frequent for analysis.

ANOVA on the rate of order errors revealed significant effects of item set, F(1, 23) = 93.58, MSE = .01, p < .001, ηp² = .81, and concurrent task, F(4, 92) = 34.92, MSE = 3.56, p < .001, ηp² = .35, while crucially, the set by task interaction was not significant, F(4, 92) = 1.17, MSE = 2.85, p = .328, ηp² = .03 again suggesting that the nature of secondary task disruption was not qualitatively different for sentential material.

Visuo-spatial and verbal n-back performance
Concurrent n-back performance was based only on trials recorded during sentence or list presentation (i.e., excluding the four pre-presentation trials). Error rates and mean reaction times are displayed in Table 4.

As expected, accuracy was substantially poorer in the 2-back versions of each concurrent task. Moreover, there were no differences associated with materials, and no interactions between materials and concurrent task conditions. However, while reaction times were faster in the visual than verbal n-back tasks, accuracy tended to be lower. We return to this latter observation in the discussion.

Discussion
We again obtained a clear advantage of constrained sentences over lists, and a marked overall effect of concurrent task. Considering first the 0-back tasks which are assumed to load principally on the visuo-spatial and phonological storage systems with minimal executive demand, we found a small effect of the visuo-spatial task and a much larger effect of articulatory suppression. In the 2-back tasks, where more substantial storage and executive processing is required, further impairment occurred, particularly in the case of verbal 2-back condition. However, the absence of significant item set by concurrent task interactions indicates that a range of manipulations that have a marked effect on performance influence sentences and word lists to an equivalent extent, suggesting once
again that sentential redundancy effects are not dependent on the contribution of working memory to encoding.

The overall pattern of errors mirrors that observed for proportion correct in the case of omissions, order errors and intrusions. As in Experiment 2, however, order errors were more frequent in lists than sentences consistent with the assumption that sentence-based redundancy enhances the storage of order information.

Performance on the secondary tasks was virtually perfect in the 0-back conditions. The visual 2-back task showed an error rate of approximately 20%, whereas the error rate for the verbal equivalent was around 14% when encoding sentences or lists (see Table 4). This pattern has implications for interpreting the verbal memory data where, it may be recalled, concurrent verbal activity was more disruptive than was visuo-spatial. This could conceivably reflect a differential trade-off with subjects required to combine sentence memory with the visual task tending to bias attention in the direction of memory, while those performing the verbal task might conceivably have shown an opposite trade-off. However, of more immediate significance is the lack of interaction between speed or accuracy and type of memory material, suggesting that whatever the nature of the impact of secondary tasks on performance, it is broadly equivalent for sentences and lists.

### Experiment 4

Our final study was concerned to test the generality of our previous observations by extending them to visual presentation of the memory materials. It seemed possible that reading, being a less automatic way of processing language than hearing, might be more likely to show differential disruption under concurrent task load. Our final study therefore compared retention of visually and verbally presented sentences and lists. In order to avoid peripheral visual interference with the visual memory task, we switched from concurrent tasks involving ongoing stimulus presentation to a more traditional pair of concurrent tasks, namely counting backwards to disrupt both phonological and executive processing, coupled with articulatory suppression to disrupt phonological alone (Allen, Baddeley, & Hitch, 2006; Peterson & Peterson, 1959).

Visual presentation in the absence of subvocal rehearsal inevitably removes any presentational cues from co-articulation or prosody. We therefore reverted to the materials and method used in Experiment 1 in which we initially recorded all spoken words independently and presented them in a telegraphic form including function words in word lists as well as sentences. Experiment 4 was therefore an attempt to replicate our previous findings with auditory presentation and explore their generalisation to visual presentation, using more tightly constrained stimulus material and modified concurrent tasks. Thus, for auditory presentation we expected counting backwards to cause additional interference over and above articulatory suppression and to do so to the same extent when recalling sentences and lists. This outcome would indicate that the presence of intonation cues was not responsible for participants’ ability to chunk sentences without requiring controlled attention in Experiments 2 and 3. Assuming that reading is more demanding of executive processes than listening, we expected the additional interference associated with counting backwards to be greater for visual than we had found with auditory presentation. Moreover, if the chunking processes underlying sentence recall are especially demanding of executive processes in the case of visual presentation, then counting could have a bigger effect on memory for visual sentences than visual lists.

### Method

#### Participants

There were 32 participants (11 male and 21 female) in this experiment. All were undergraduate or postgraduate students at the University of York, taking part for financial payment or course credit.

#### Materials

Stimulus sets and scoring were the same as in Experiment 1. Four sets of sentences and four sets of lists were created, with 11 sequences (one practice and 10 test) in each set. Sentences consisted of seven words (two adjectives, three nouns, one verb, one function), for example, **tall soldier follows waiter not old teacher**, while lists were five words (one adjective, two nouns, one verb, one function) in length, e.g., **follows old not waiter teacher**. As in Experiments 2 and 3, sequence lengths were chosen so as to try and match difficulty in terms of proportion of items correctly recalled. Lists were constructed by removing a noun and an adjective from each sentence, and adjusting the order of the words by at least two levels of approximation, to ensure minimal meaning and redundancy.

#### Design and Procedure

A $2 \times 2 \times 2$ repeated measures design was implemented, manipulating item set (sentences, lists), modality...
(auditory, visual), and concurrent task (articulatory suppression, backward counting). There were therefore eight experimental blocks in total, alternating between sentence and list blocks and grouping by modality, and at the higher level, by concurrent task. In other words, all suppression (or backward counting) conditions were performed together, with modality conditions grouped within these blocks. Counterbalancing was applied at each level of grouping. There was one practice trial and 10 test trials within each individual sequence block. Each condition implemented a different set of sequences (counterbalanced across the experiment), with sequences drawn from a set in a random order for each participant.

The experimental session began with a baseline measure of backward counting performance. Participants were presented with three digits by the experimenter (e.g. “2-7-6”), and, treating it as a three-digit number, attempted to count backwards from there (e.g. “two-seven-five, two-seven-four, two-seven-three…”1). There were five practice trials at the start of the experiment, each lasting 8 s, with another five trials at the end of the experimental session (to control for practice effects).

Participants pressed the space bar to begin presentation of each sequence, following a 2 s delay. Auditory presentation occurred through headphones at a rate of one word per second, using the same digital sound files for sentences and lists. With visual presentation, words were presented serially at upper centre screen in black 18-pt lower case Arial, at a rate of one per second. Following auditory or visual sequence offset, there was a 1 s blank screen delay, a further 1 s delay containing a visual mask (subtending a visual angle of approximately 19° by 6.5°), and then the serial recall phase.

In the articulatory suppression conditions, participants repeated the digit sequence “1-2-3” out loud during sequence presentation, from the point of pressing the space bar through to presentation of the visual mask. For backward counting, a three-digit number was presented for 2 s at upper centre screen in black 18-pt Arial at the start of the trial, with counting proceeding from this point through to mask presentation.

Results

Sentence and word list recall

Recall was scored using the procedure described earlier. As before, structure and redundancy had a beneficial impact on recall. The mean number of words recalled in the articulatory suppression condition was, with auditory presentation, 5.0 (SD = 1.1) for sentences and 3.5 (0.8) for lists, and with visual presentation, 4.1 (SD = 1.1) for sentences and 3.1 (1.0) for lists. Hence the sentence superiority effect for number of words recalled was significant, F(1, 31) = 98.13, MSe = .43, p < .001, ηp² = .76. There was a significant effect of presentational modality, F(1, 31) = 60.41, MSe = .78, p < .001, ηp² = .66, but no interaction between modality and materials, F(1, 31) = 2.71, MSe = .59, p = .110, ηp² = .08. Thus, the recall advantage for sentences over lists remains even with serial visual presentation of words.

As previously, our principal analysis was based on mean proportion of words correct. Recall scores for each item set and concurrent task condition are displayed in Fig. 6. A 2 × 2 × 2 repeated measures analysis of variance revealed no effect of item set, f(1, 31) = 0.18, MSe = .01, reflecting successful titration of sequence lengths in matching for difficulty level. There were significant effects of concurrent task, F(1, 31) = 378.35, MSe = .01, p < .001, ηp² = .92, and modality, F(1, 31) = 58.74, MSe = .01, p < .001, ηp² = .66, with as expected, worse performance under backward counting and with visual presentation. In addition, there was a significant interaction between concurrent task and modality, F(1, 31) = .01, p < .05, ηp² = .14, indicating a larger impact of executive load on encoding during visual presentation. However, the concurrent task by item set interaction was not significant, F(1, 31) = 1.25, MSe = .01, p = .27, ηp² = .04. Thus, backward counting had a much greater effect on recall than did suppression, but such disruption was no greater for sentences than for lists. The modality by item set and three-way interactions were not significant (F < 1).

Articulatory suppression and backward counting performance

The total number of counts achieved in each trial was divided by the length of the trial in seconds (to allow for the different trial lengths used for sentences and lists), to obtain a measure of the mean number of counts per second. The mean suppression and backward counting scores for each concurrent task condition, along with the baseline score in backward counting, are displayed in Table 5.

A 2 × 2 × 2 repeated measures analysis of variance on the concurrent task conditions revealed significant effects of task type, F(1, 31) = 86.74, MSe = .01, p < .001, ηp² = .74, and sequence type, F(1, 31) = 63.21, MSe = .01, p < .001, ηp² = .67. Backward counting was performed more slowly than articulatory suppression, consistent with our assumption that counting is a more demanding task. In addition, both concurrent tasks were slower with sentences relative to lists. The main effect of modality was not significant, F(1, 31) = .46, MSe = .01, though there was a significant interaction between modality and task type, F(1, 31) = 4.21, MSe = .01, p < .05, ηp² = .12, with backward counting being slightly slower with visual, relative to auditory presentation of sequences. There were no other significant interactions (F < 1). Finally, performance of backward counting as a concurrent task was compared with the baseline measure of counting performance. A series of planned comparisons revealed that backward counting was significantly slower, relative to baseline during auditory sentence presentation, t(31) = 4.84, p < .001, visual sentence presentation, t(31) = 6.04, p < .001, and visual list presentation, t(31) = 2.78, p < .01, while the effect for auditory list presentation was not significant, t(31) = 1.57, p = .128.

Discussion

Results for the auditory presentation condition are discussed first before moving on to the comparison between...
the auditory and visual conditions. The purpose of the auditory condition was to see whether the effects of a demanding concurrent task found in Experiments 2 and 3 would replicate when stimuli were presented using discontinuous rather than continuous speech. The outcome was straightforward. Once again sentences were recalled better than word lists; there was substantial dual-task interference from an attention-demanding concurrent task; and the amount of interference was no greater for sentences than for lists. Thus we can be confident that this set of observations reflects the sequential redundancy of language and that the earlier results were not due to perceptual cues associated with continuous speech. Analyses of secondary task performance confirmed that the pattern of dual-task decrements in recall was not due to differential trade-offs between the primary and secondary tasks. It is interesting to note that secondary task performance was significantly poorer in the sentence than list conditions. However, the effect was the same for articulatory suppression and backward counting, suggesting that sentences placed greater demands on the phonological loop (involved in both secondary tasks) but not the central executive. Whether this effect was due to the difference in structure or length of the two types of sequence is impossible to say as the two were confounded.

The pattern of results for visual presentation was virtually identical to the auditory data, except that recall was generally poorer and backward counting caused significantly more interference relative to articulatory suppression. The increase in dual-task interference was reflected in slower backward counting performance with visual as compared with auditory presentation of sequences. We expected visual presentation to be more attention-demanding than auditory presentation given that participants had to direct their attention to the stimuli and carry out the extra process of phonological recoding. We were more particularly concerned to see whether the sentence superiority effect for visual sequences was vulnerable to an attention-demanding concurrent task. The results were clear-cut in showing a sentence superiority effect of the same magnitude as with auditory presentation that remained intact despite concurrent backwards counting. Although not shown here, our data once again indicated that sentences were associated with a reduction in order errors, and that serial position curves were bow-shaped for sentences as well as lists. These observations provide further evidence that the results of Experiments 2 and 3 were not an artefact of co-articulation or prosodic speech cues. The results thus add to our earlier evidence that working memory takes advantage of the structural constraints in sentences without making demands on executive processes. It seems, therefore, that the chunking process triggered by sequential redundancy in language is an automatic rather than a controlled process.

**General discussion**

To recapitulate, we have examined the effects of sequential redundancy in immediate serial recall by analysing the sentence superiority effect, whereby memory for sentences is enhanced relative to memory for word lists. We assume that memory for sentences benefits from a process in which short-term storage of the phonological input interacts with knowledge of the sequential redundancy of language to bind together groups of items into larger chunks. Our main aim was to explore whether the
integration of these different types of information involves a multi-modal episodic buffer that operates under central executive control (Baddeley, 2000). We approached this question by comparing the effects of different concurrent tasks on the immediate serial recall of sentences and word lists. As formulated, the episodic buffer hypothesis predicts that a concurrent task taxing the limited resources of the executive should interact with materials, thus reducing or even removing the sentence superiority effect.

An initial methodological problem for our approach was that memory for natural sentences is typically very much better than memory for lists. Marked differences in performance levels make experimental comparisons difficult, especially as regards the detection and interpretation of interactions. We attempted to overcome this difficulty by developing a constrained sentence span task that combines the presence of linguistic redundancy with features of the classic word span task in which items are repeatedly sampled from a limited set. We found a robust and reliable sentence superiority effect for constrained sentences in a series of four experiments encompassing major changes in the way information was presented (whether as continuous speech, disconnected speech or visually), and minor changes in other aspects (principally variations in sequence length and the role of function words).

Analyses of errors of recall shed light on the basis of the sentence superiority effect in showing lower proportions of order errors for both open and constrained sentences than for word lists (see Experiment 2). The reduction in order errors is consistent with our assumption that memory for sentences benefits from the sequential redundancy inherent in language. Open sentences tended to produce more semantic errors than either constrained sentences or word lists, consistent with the suggestion that they benefit from a cueing effect based on memory for gist. One of the aims of using constrained sentences was to reduce the involvement of memory for gist and the reduction in semantic errors suggests this aim was met. A second aim of using constrained sentences was to increase the demands on working memory by sampling from a limited word pool and using a limited range of syntactic structures. This second aim also appears to have been met in that a post-experimental recognition test demonstrated substantial long-term episodic memory for open sentences but chance performance for constrained sentences (see Experiment 2). In addition, constrained sentences led to increased pro-active interference in the form of intra-experimental intrusions. We conclude therefore that constrained sentences serve their purpose of facilitating experimental comparisons with unstructured word lists, retaining the crucial feature of sequential redundancy while being less reliant on episodic LTM and semantic coding than open sentences.

Turning to the question of whether the sentence superiority effect depends on the central executive, the answer seems clear. A remarkably consistent pattern of results was found across a range of concurrent tasks loading on executive processes (continuous choice RT in Experiment 2, 2-back recall in Experiment 3 and backwards counting in Experiment 4). Furthermore, this pattern was obtained regardless of whether the memory materials were presented orally with intonation (as in Experiments 2 and 3), or orally without intonation or visually (as in Experiment 4). In each case the concurrent task impaired recall to the same extent for sentences and lists and in no case was the sentence superiority effect removed or reduced by a concurrent task involving executive processes. This pattern of effects was primarily reflected in order errors, the tendency for sentences to generate fewer order errors than lists, which we interpret as an indicator of sequential chunking. We note also that less demanding concurrent tasks assumed to load on the visuo-spatial and phonological subsystems (articulatory suppression in Experiment 2; visuo-spatial and verbal 0-back in Experiment 3) gave rise to similar results. The amounts of interference were typically smaller and once again the sentence superiority effect remained intact, implying that while both the phonological loop and the visuo-spatial sketchpad contribute to recall, they do this no more for sentences than lists. Thus, our main conclusion is that the processes involved in chunking occur more or less automatically, with no special dependence on the central executive, inconsistent with our initial suggestion the executive plays a critical role in accessing the episodic buffer.

It could however be argued that the type of interaction we were looking for was too difficult to detect because our methods were not sufficiently sensitive. However, in the present investigation our methodology was not only consistent in indicating no differential effect of executive load on memory for sentences and lists across a range of experiments: it was also demonstrably capable of revealing this type of interaction. Thus, in Experiment 4 we were able to detect a significant differential effect of executive load for visual and auditory methods of presenting the memory materials, whilst at the same time finding no differential effect on sentences vs. lists. Whilst this is not a conclusive argument, it does at least indicate that our methodology was sensitive, suggesting that any executive involvement in chunking was at most slight.

It is important to comment on the fact that in the present investigation we have not assessed chunking directly, preferring instead to use the sentence superiority effect as an index of chunking. We took this decision because our experiments were exploring the general question of the relationship between chunking and attention and we sought a way of avoiding the well-known problems of defining and measuring chunks (and indeed detailed issues concerning language processing as discussed in the introduction). However we note that if one is prepared to make plausible assumptions about chunking it is possible to design experiments that allow sentence and list recall to be scored in terms of chunks as well as items (see e.g. Gilchrist, Cowan, & Naveh-Benjamin, 2008). Adopting this kind of approach would seem important for making further progress in understanding where and how chunking processes take place.

The most obvious way of accounting for our results in terms of the concept of the episodic buffer as a temporary store for bound information is to modify the assumption that access to the buffer is dependent on the central executive (Baddeley, 2000). That is, we would assume that chunks enter the episodic buffer where they become
available to attention, but the binding processes whereby
cunks are formed and enter the buffer occur outside
attention. Thus, at any instant the episodic buffer contains
a limited number of recently formed chunks that can enter
into the focus of attention.

This view of chunking is consistent with evidence on
sentence comprehension reviewed by Caplan, Waters,
and DeDe (2008), who conclude that the “initial, auto-
matic, on-line, obligatory, unconscious, processes that as-
sign the structure and literal meaning of a sentence”
require specialised resources and do not draw on the cen-
tral executive of working memory. We would tentatively
equate the creation of chunks during sentence presenta-
tion with these “interpretive” processes, which Caplan
et al. distinguish from “post-interpretive” processes that
do require executive processes. The latter would corre-
spond in our account to the use of focused attention to ac-
tively manipulate chunks stored in the episodic buffer. Our
interpretation is also broadly consistent with the work of
Potter and colleagues who have shown that readers can
understand and remember sentences presented at rates of
up to 12 words per second whereas only very few unre-
lated words can be remembered at such fast rates. Potter
interprets these observations in terms of rapid formation
of chunks through automatic activation of knowledge in
long-term memory, and temporary storage of such chunks
in a conceptual short-term memory store (CSTM). Our no-
tion of the episodic buffer is similar to CSTM but differs in
that the buffer is assumed to be multi-modal and capable
of binding together and integrating information from other
parts of the memory system besides long-term memory,
including sensory stores, the phonological loop and the vi-
suo-spatial sketchpad.

Our interpretation is also broadly compatible with the-
etorical approaches that identify working memory with
the focus of attention (Cowan, 2005) or controlled atten-
tion (Engle, 2002) in that we assume the contents of the
episodic buffer are available to attention. However we
would argue that the contents of the buffer are not neces-
sarily in the focus of attention at any instant, an idea that
might be mapped onto Cowan’s proposal that a limited
amount of additional information lies immediately outside
the current focus of attention in a readily accessible state.
A potentially important difference between ourselves and
Cowan is our suggestion that the limit on number of
chunks in recall may be a function of the storage capacity
of the episodic buffer, rather than as Cowan suggests,
purely a limit on the capacity of focused attention. Another
distinction is that Cowan (2008) regards the phonological
loop and visuo-spatial sketchpad as part of activated
long-term memory whereas we see them as temporary
stores, separate from the episodic buffer. Seeking to clarifying
these key issues is an important goal for future research.

Our present interpretation is apparently somewhat at
odds with Ericsson & Kintsch’s (1995) theory of long-term
working memory (LTWM). Ericsson & Kintsch discuss
memory for language in terms of a model in which chunks
are formed in long-term memory. They assume working
memory holds pointers to a retrieval plan that reflects
the organisation of information in long-term memory
and is used to unpack the contents of chunks during recall.
This approach stands in contrast to our assumption that
chunks are held in a temporary episodic buffer. Ericsson
and Kintsch support their view by showing that memory
for language can survive interruptions and delays under
some circumstances. They also show more generally that
this applies to memory for any type of information where
the individual has extensive knowledge, such as a chess ex-
pert’s memory for the arrangement of pieces on a chess-
board. Interestingly, our present data on post-experimen-
tal recognition memory suggest that the type of chunk
formed may be an important factor as regards its durability
in memory. Thus, long-term recognition memory was rela-
tively good for open sentences but at chance for con-
strained sentences, an observation we attributed to the
role of gist in the former but not the latter (see Experiment
2). This observation suggests a speculative way of combing
the LTWM and episodic buffer approaches into a uni-
fied account using Hebb’s idea of short-term memory as
a set of transient reverberatory circuits whose persistence
results in the formation of more stable representations in
long-term memory (Hebb, 1949). We would suppose that
the episodic buffer holds chunks in the form of such cir-
cuits and that this activity is more readily overwritten
through similarity-based interference when it is primarily
syntactic (as in constrained sentences) than when it is pri-
marily semantic (as in open sentences). We would suppose
further that chunks that are not overwritten quickly have a
higher probability of being stored in long-term memory,
and that these circumstances correspond most closely to
those discussed by Ericsson and Kintsch.

Consideration of the role of long-term memory in recall
raises the further question whether participants in our
experiments might not have formed chunks at all, simply
using their knowledge that they had received a sentence
rather than a list to constrain guessing during recall. Thus,
in the middle of a recall protocol the participant would
have some idea what parts of speech could come next in
the case of a sentence, and this benefit of guessing alone
would improve memory for the order of the words.
Although such guessing would almost certainly have oc-
curred to some extent, it does not come close to explaining
the size of the sentence superiority effect we observed. For
example, in Experiment 2 participants recalled on average
about five list items and seven items from constrained sen-
tences, a difference of two items. However, participants
made on average very much less than one order error in
recalling lists. Moreover, given the size of the experimental
word pool (12 adjectives, 12 nouns, four verbs, four ad-
verbs), the average probability of converting an order error
into a correct response by constrained guessing is not high.
Thus, we conclude that the sentence superiority effect was
too big to be explained solely in terms of redintegrative
processes during recall, while accepting that such pro-
cesses no doubt made some contribution to recall
performance.

We should note at this point that we have reached a
broadly similar theoretical conclusion in the case of bind-
ing in visual working memory. The binding of features such
as shape and colour into objects is also unaffected by a
concurrent executive load (Allen et al., 2006), as is the
binding effect of symmetry in visual STM (Rossi-Arnaud, Pieroni, & Baddeley, 2006), suggesting the presence of a direct link between the visuo-spatial sketchpad and the episodic buffer. We have even found evidence that binding shape and colour is automatic when one feature is presented visually and the author auditorily (Allen, Hitch, & Baddeley, 2009). It could be argued that the binding of features into objects and the detection of symmetry involve a much more basic level of integrative processing than sentential chunking based on syntactic and semantic information. There is however evidence of automatic processing of visual stimuli at a much deeper and more complex level from an extensive series of studies by Potter and her associates (e.g., Potter, 1993; Potter & Lombardi, 1990; Potter, Straub, & O’Connor, 2004). For example, viewers can identify a complex visual target such as “two men drinking beer” from a sequence presented as rapidly as eight items per second (Potter, 1975). We have already noted a close degree of correspondence between Potter’s notion of CSTM and our view of the episodic buffer.

More generally still, it is important to note that the present experiments only speak to chunking processes that operate within individual sentences. Jefferies, Lambon Ralph, and Baddeley (2004) investigated the role of different components of working memory in chunking in a study that required participants to recall a series of sentences or scrambled word strings under either control conditions, or when performing an attention-demanding continuous four-choice reaction time task similar to that used in Experiment 2 (Craik et al., 1996). Number of sentences or lists presented was set at 50% above span for that type of material, participants being given three successive learning trials. The effect of the concurrent task was significant but not large, and differed between the two sets of materials. There was little initial effect on immediate recall of the word lists, but the size of the effect grew over trials, while for the sets of sentences, the trend was in the opposite direction. Jefferies et al. interpreted the result as suggesting that attentional resources were required for chunking under some circumstances. In the case of the scrambled words, this built up as subjective units developed, requiring more than one trial as observed by Tulving (1966). In the case of sentences, the principal challenge came from linking together the unrelated sentences; once they were chunked into a structure, the need for executive processing became less. This interpretation was supported by a further experiment in which a third condition was added, comprising sentences that were thematically related, hence likely to require minimal attention for inter-sentence chunking. The concurrent task had little effect on this condition, supporting our chunking hypothesis. Thus, it seems that there are limits on the generality of the conclusions from the present experiments, and chunking may involve attention when binding in memory for super-span verbal materials is made particularly difficult.

In conclusion, our investigation into the role of sentential redundancy in binding has been guided by the multi-component working memory framework, which operates by framing questions rather than making precise predictions. In answering such questions, we provide data that have potential importance for a range of alternative theoretical approaches. What then, are the implications of our results for the concept of an episodic buffer? Firstly, they demonstrate that despite its lack of precise predictions, the theoretical framework is capable of posing important questions concerning the role of attention in binding in short-term memory, producing unexpected results that will need to be accounted for by any adequate theory of binding. Secondly, they strongly suggest a modification of the initial Baddeley (2000) model, which proposed that the buffer, important for binding features into episodes, could be accessed only via the central executive. It is now clear that such binding can occur without disruption, even when the central executive is loaded to a point at which a substantial decrease in overall performance occurs. This conclusion appears to hold not only for chunking in verbal short-term memory but also for different kinds of feature binding in visual short-term memory (Allen et al., 2006, 2009; Rossi-Arnaud et al., 2006).

The earlier version of the buffer (Baddeley, 2000) could be seen as involving a minimal change from the initial Baddeley and Hitch (1974) concept of a central executive that could both process and store information, replacing the single executive with two systems that interact in a powerful but unspecified way. Our recent results suggest a more complete separation between executive control and passive episodic storage. The evidence now appears to favour direct access to the buffer through the phonological and visuo-spatial subsystems, and possibly more directly through perception and from LTM. This would imply a relatively passive store, with both binding and the manipulation of bound representations depending on processes that operate outside the buffer, the former being automatic and the latter involving executive processes. We plan to continue to investigate this and related hypotheses in future work.

Acknowledgments

We are grateful to David Turk for help in developing the constrained sentence test, to Elizabeth Littlewood for assistance with data collection, and to the Medical Research Council Grant G9423916 for financial support. We thank also several anonymous reviewers and Nelson Cowan for valuable comments on earlier versions of this paper.

Appendix A

Constrained word pool and sequence construction rules

<table>
<thead>
<tr>
<th>Nouns (1)</th>
<th>Nouns (2)</th>
<th>Nouns (3)</th>
<th>Verbs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lucy</td>
<td>pilot</td>
<td>book</td>
<td>gave</td>
</tr>
<tr>
<td>Peter</td>
<td>athlete</td>
<td>bicycle</td>
<td>borrowed</td>
</tr>
<tr>
<td>Mary</td>
<td>musician</td>
<td>suit</td>
<td>cleaned</td>
</tr>
<tr>
<td>John</td>
<td>lawyer</td>
<td>car</td>
<td>stole</td>
</tr>
<tr>
<td>red</td>
<td>large</td>
<td>frightened</td>
<td>Adverbs</td>
</tr>
<tr>
<td>black</td>
<td>small</td>
<td>wealthy</td>
<td>rapidly</td>
</tr>
<tr>
<td>green</td>
<td>old</td>
<td>angry</td>
<td>instantly</td>
</tr>
<tr>
<td>white</td>
<td>new</td>
<td>hopeful</td>
<td>cheerfully</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>easily</td>
</tr>
</tbody>
</table>
Constrained sentences

a. Each sequence contained 1–2 words from each subset of nouns and adjectives, one verb, and one adverb, sampled without replacement.
b. There were also 2–4 function words from the, to, and, from, and for, sampled with replacement. These were not included in any data analysis.
c. A variety of grammatical templates were allowed, within the bounds of acceptable linguistic structure.

Constrained lists

a. Each list was constructed by removing all function words and two test words from a constrained sentence.
b. Word order was then manipulated so that each sequence was at least two orders of approximation from an acceptable grammatical structure.

References


