Lexicality and interference in working memory in children and in adults

Juliet A. Conlin, Susan E. Gathercole *

Department of Psychology, Science Laboratories, University of Durham, Durham, DH1 3LE, UK

Received 10 June 2005; revision received 24 May 2006
Available online 14 July 2006

Abstract

Four experiments investigated the impact of the lexical status of memory and processing stimuli on complex memory performance, with the aim of exploring mechanisms of interference in working memory. In a complex memory task, participants recalled words or nonwords while either monitoring words or nonwords for phonological content, or suppressing articulation. In groups of 9- and 10-year-old children and adults, word recall was significantly more impaired by monitoring words than nonwords. A converse disturbance of nonword recall by nonword monitoring was consistently found for adults, but was less marked in the child groups. It is proposed that interference in complex memory tasks reflects the operation of two distinct processes: a lexical-semantic process of either interference between memory and processing stimuli or redintegration, and the strategic use of lexical status to discriminate potential target from non-target items. Whereas the former process is invariant with age, the latter strategy is robust in adults but in the early stages of emergence with the younger participants.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Working memory; Lexicality; Interference; Short-term memory

Introduction

Working memory is generally viewed as a limited capacity system responsible for the temporary storage and processing of information (e.g., Baddeley, 1986; Just & Carpenter, 1992). Working memory capacity is widely measured using complex memory paradigms, in which participants are required to combine memory for sequences of items whose presentation is interleaved by processing activities. The number of items to be remembered is increased until the maximum length at which memory accuracy is maintained. For example, in the reading span task developed by Daneman and Carpenter (1980), participants read a series of sentences while attempting to remember the sentence-final word (or some other target item) for later serial recall. There is comprehensive evidence that performance on such tasks is linked in both children and adults with key cognitive skills such as language comprehension (e.g., Daneman & Merikle, 1996), reasoning (e.g., Süß, Oberauer,
Wittmann, Wilhelm, & Schulze, 2002), general fluid intelligence (e.g., Engle, Tuholski, Laughlin, & Conway, 1999), and with learning abilities in the areas of both literacy (e.g., Swanson, Ashbaker, & Lee, 1996) and mathematics (e.g., Geary, Hoard, Byrd-Craven, & DeSoto, 2004).

The underlying cognitive processes that support working memory performance remain open to debate. According to the working memory model advanced originally by Baddeley and Hitch (1974) and developed subsequently by Baddeley and colleagues (Baddeley, 1996, 2000; Baddeley & Logie, 1999), working memory reflects multiple resources associated with distinct capacity-limited sub-systems. This model incorporates the central executive, which is associated with attentional control, high-level processing activities, and the coordination of activities within working memory. Other components of working memory include two modality-specific slave systems responsible for the storage of verbal (the phonological loop) and visuo-spatial material (the visuo-spatial sketchpad). A fourth component is the episodic buffer (Baddeley, 2000), responsible for integrating representations both within working memory and across the cognitive system more generally. Within this model, the storage demands of complex memory span are suggested to depend on appropriate sub-systems, with processing supported principally by central executive resources (Baddeley & Logie, 1999; Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002).

A contrasting theoretical perspective is that performance on complex span tasks is constrained by a limited resource that is shared between processing and storage (Daneman & Carpenter, 1980, 1983; Just & Carpenter, 1992). By this account, individuals vary not in the total amount of resource available, but in the resource demands imposed by processing activities; thus, individuals with inefficient processing will have less resource available to support storage and so will have relatively poor complex memory spans (e.g., Just & Carpenter, 1992; King & Just, 1991). There have in recent years been substantial challenges to this theoretical account of working memory. Towse and colleagues have demonstrated that increasing the duration over which information must be retained whilst holding task difficulty constant has a deleterious effect on complex memory span (Cocchini et al., 2002). More recent research has explored the conditions under which interference between storage and processing can arise in complex span tasks. In general, performance is disrupted when processing and storage stimuli are drawn from the same general content domain; for example, greater decrements in complex memory span for verbal material are found when the interleaved processing involves verbal than visuo-spatial processing (Bayliss, Jarrold, Gunn, & Baddeley, 2003; Shah & Miyake, 1996; Turner & Engle, 1989). Such data fit well with multiple resource accounts such as the working memory model that incorporates domain-specific storage systems, according to which verbal recall is mediated by the phonological loop, and visuo-spatial processing is handled by either or both the visuo-spatial sketchpad and central executive. It is less clear how much data could be handled by the notion of an undifferentiated working memory resource supporting both storage and processing (e.g., Daneman & Carpenter, 1980).

Evidence that similarity within content domains also disrupts complex memory span performance is more problematic for the multiple resources account in its current form. In a recent study, we compared complex span performance in children under conditions in which both verbal storage and processing items were either numerical or non-numerical stimuli (Conlin, Gathercole, & Adams, 2005a). Recall of digits was lower when the processing activity involved calculating arithmetic operations than processing the meaning of sentences. In contrast, sentence processing had a disruptive effect relative to arithmetic processing on the recall of words that were not digit names. Turner and Engle (1989) also report data showing a similarity decrement across a series of conditions that combined digit- and word-based processing tasks with recall of digits and words, although no statistical comparisons of the conditions were reported. The working memory model cannot readily accommodate these findings, as the recall of verbal stimuli would be expected to be mediated by the phonological loop, regardless of whether they were digit names or not. The crossover nature of the interaction in this study also rules out explanation in terms of differential
Several accounts of interference in working memory have been advanced that could potentially account for such findings. Engle et al. (1999) proposed that the ability to activate and maintain memory representations in the face of interference or distraction underpins individual differences in working memory capacity. By this account, complex span tasks require controlled attention to prevent distracting secondary information from interfering with the maintenance of target memory items. It does not, however, elucidate the specific mechanisms underpinning interference, nor does it explain why the greatest disruptions in memory performance arise when the processing and storage stimuli are drawn from common representational domains (e.g., Conlin et al., 2005a; Shah & Miyake, 1996).

A more detailed account advanced by Oberauer and colleagues (Lange & Oberauer, 2005; Oberauer & Kliegl, 2001; Oberauer, Lang, & Engle, 2004) is that interference results from partial overwriting of memory representations, such that the extent of overwriting is determined by the degree of overlap between the representations of target (to be remembered) and non-target (processing) items. Saito and Miyake (2004) advanced a similar model in which stimuli generate a variety of representations (e.g., phonological, semantic, visual, as appropriate), and interference arises as a consequence of high degrees of featural similarity within representational domains. The major problem for this account is that, apart from the Conlin et al. (2005a) findings discussed above, there is little evidence that similarity between processing and storage material within a content domain disrupts complex memory span. The most relevant study is that of Oberauer et al. (2004), which manipulated similarity within both the spatial and verbal domains. Memory for spatial stimuli (either visual patterns or lines drawn in a grid) was tested with processing that involved symmetry judgments of stimuli presented either in the same or different form (patterns, lines) as the memory items. Two forms of similarity were tested in the verbal domain, in which the items to be recalled were always animate nouns (either animals or plants), and the processing activity involved reading aloud three words interpolated between each memory item. The degree of semantic similarity between storage and processing items was manipulated by employing stimuli in the interleaved processing task that were either animate (high similarity) or inanimate (low similarity). The degree of phonological similarity between storage and processing items was manipulated by using processing stimuli that either rhymed with memory items or had minimal phonological overlap. No consistent performance decrements were found in any of the high similarity conditions, clearly providing a substantial challenge for feature-based interference accounts of working memory.

A contrasting account of these findings is proposed here, according to which there is no dynamic process of interference between representations of processing and storage items in working memory. Instead, we suggest that the disruptive consequences of both types of stimulus tapping a common domain may arise from failure to discriminate between target and non-target representations at the later stage of retrieval. Processing and storage items that are drawn from different domains or highly distinct categories generate cues that can facilitate this discrimination and hence lead to improved recall accuracy. What, though, constitutes an effective discrimination cue? It is argued here that to be effective, the cues will need to be highly salient, such as distinctions between storage and processing items on the basis of modality (spatial, verbal etc.) that correspond to gross intrinsic features of item representations. Membership of a small and highly familiar category such as digit names (Conlin et al., 2005a) will also be useful cues for selecting and rejecting potential target items. In contrast, the within-domain manipulations of similarity employed by Oberauer et al. (2004) yielded less useful cues for discriminating potential target from non-target responses in the dissimilar conditions. In the spatial conditions, all memory and processing stimuli were presented in a 3 × 3 grid and thus were likely to generate visual and spatial attributes that were to some extent overlapping, irrespective of the designated similarity of the experimental condition. In the verbal conditions, target and processing items were differentiated in the semantically dissimilar condition by animacy of the nouns. The animacy category is extremely broad, incorporating the full range of concepts corresponding to plants and animals. Differentiation of items on the basis of animacy is neither highly familiar nor practiced in the same way as, for example, digit and non-digit names, and for this reason may not provide an effective cue for discrimination at retrieval. Finally, there was no reliable cue for selecting potential target responses in the phonologically dissimilar condition, as neither the memory nor the processing items shared any common physical features. Thus in all three cases, the manipulations of similarity between memory and processing items in the Oberauer et al. study were not implemented in a way likely to support easy discrimination of target items at retrieval.

Further support for a cue-based explanation comes from a series of experiments conducted by Humphreys and Tehan (1999), who explored complex span and fixed list recall performance under conditions in which the processing activity involved either random or meaningful items. They proposed that meaningful sentences would provide more effective cues for isolating the memory items than either ill-formed sentences or processing of random words, which would be expected to produce interference for the memory words. An interference
effect was observed across experiments in the random processing conditions, explained by these authors in terms of cues and degraded memory representations.

Reports that processing items do indeed frequently intrude in working memory tasks, particularly in individuals with poor working memory spans, are entirely consistent with this discrimination cue hypothesis (DeBeni & Palladino, 2000; DeBeni, Palladino, Pazzaglia, & Cornoldi, 1998). It has been argued that such intrusions arise from failures to inhibit irrelevant representations in working memory (e.g., Palladino, Cornoldi, De Beni, & Pazzaglia, 2001). It is important to note that in contrast, the account advanced here does not require an assumption that non-target responses are actively inhibited for accurate performance, although target and non-target activation levels may be differentiated prior to retrieval by strategic maintenance processes such as rehearsal. Findings that proactive interference builds up across trials within complex span tasks (Lustig, May, & Hasher, 2001; May, Hasher, & Kane, 1999) provide further evidence for response confusion in working memory, in this case resulting from increases in the number of activated representations from which target memory items must be selected.

The present series of experiments provide a detailed investigation of the influences of similarity between processing and storage items on complex memory performance in children and adults, with the aim of advancing theoretical understanding of mechanisms of interference in working memory. The principal focus was on the extent to which the disruptive consequences of similarity between memory and processing items in the verbal domain operate at phonological and lexical-semantic levels, and whether they result from interference between representations during storage or a failure to differentiate target from non-target representations at retrieval. Stimuli in the three experiments were monosyllabic items with a consonant–vowel–consonant structure. Memory items were either words or nonwords presented visually, and the key processing conditions involved monitoring for phonemic content a stream of spoken words or nonwords presented between memory items. Control conditions across experiments involved either articulatory suppression or no processing activity.

The logic of the experiments is as follows. If interference is a consequence of overlap between semantic features activated by items presented for storage and processing (Oberauer et al., 2004; Saito & Miyake, 2004), word recall should be disrupted to a greater extent when words rather than nonwords are processed because the latter stimuli have no associated semantic attributes to generate interference. By the same token, nonword recall should be uninfluenced by the lexical status of the processing stimuli, as there are no semantic representations to be overwritten. Interference between processing and memory items at a phonological level, in line with the featural analysis accounts of Oberauer et al. and Saito and Miyake, should be equivalent for all storage and processing conditions because the degree of phonological overlap between stimuli is independent of lexical status. Recall should, however, be impaired following any phoneme monitoring activity compared with articulatory suppression interpolated between the presentation of memory items, on two grounds. First, the amount of phonological material generated during suppression (the single word “the” repeated throughout the experiment) is minimal, generating fewer phonological representations and hence a lower degree of phonological overlap with target items than the monitoring tasks. Second, the attentional demands of articulatory suppression are minimal (Barr-Outillet & Camos, 2001), whereas phoneme monitoring is likely to be more demanding and hence disruptive of recall.

Contrasting predictions are generated by the discrimination cue hypothesis, according to which disruptive effects of similarity between memory and processing stimuli result from the absence of a cue to differentiate representations of potential target from non-target items at retrieval. Lexical status of memory items appears to be an effective cue for such discrimination. Error analysis of serial recall for words and nonwords has established that the proportion of total error responses that are words rather than nonwords is considerably greater when participants recall lists of words than nonwords, and vice versa (Gathercole, Pickering, Hall, & Peaker, 2001). Thus, participants appear to use the lexical status of the target items to guide the selection of items for output, even in error. If the same lexical consistency strategy for selecting responses can be employed in the present complex memory span task, participants should be able to differentiate potential processing from storage items under those conditions in which the lexical status of memory and processing items differs but not when they are the same. The crucial distinction between the two sets of theoretical predictions therefore concerns the influences of word and nonword processing on nonword recall. According to the featural overlap view (e.g., Oberauer et al., 2004), performance should be equivalent in both conditions. In contrast, the retrieval-based confusion account predicts a decrement in nonword recall with nonword compared with word monitoring.

Note that the type of experimental manipulation used here differs from that used in several other studies that have investigated factors limiting working memory span (e.g., Saito & Miyake, 2004; Towse et al., 1998, Towse, Hitch, & Hutton, 2000). In the current study, the types of processing materials in a whole list were manipulated, in contrast to studies in which the same
processing materials were used across conditions in different orders to test for influences of the amount of information processing (Saito & Miyake, 2004) or processing duration (Towse et al., 1998, 2000). Thus, the experimental paradigm employed here allows for a direct test of confusability between processing and storage stimuli.

These experiments also provide one of the first systematic examinations, to our knowledge, of the effects of the lexicality of memory items in a complex memory span task. In serial recall, the presence of a recall superiority for lists of words over nonwords is well-established (e.g., Gathercole et al., 2001; Hulme, Maughan, & Brown, 1991), and is typically explained by the use of lexical knowledge at retrieval to reconstruct complete memory representations, a process termed redintegration (Gathercole, Frankish, Pickering, & Peaker, 1999; Hulme et al., 1997; Schweickert, 1993; Thorn, Gathercole, & Frankish, 2005). According to this view, redintegration will be effective for memory items with lexical representations (words), but not for those lacking such representations (nonwords). The present question concerns whether lexicality exerts a corresponding beneficial influence on complex memory performance. Such a finding would lend weight to accumulating evidence that serial recall and complex memory span paradigms tap some common cognitive processes (e.g., LaPointe & Engle, 1990; Lobley, Gathercole, & Baddeley, 2005).

The final issue addressed in these experiments concerns potential developmental changes in the mechanisms underpinning interference effects in working memory. Working memory function has been extensively researched in children as well as adults, with much of the theoretical analysis in the field being driven by both experimental and individual differences analyses of children’s performance (e.g., Bayliss et al., 2003; Towse, Hitch, Hamilton, Peacock, & Hutton, 2005). It is notable that by and large, developmental continuities rather than discontinuities have been established, with children’s working memory performance showing similar influences of key variables to that of adults (Bjorklund & Harnishfeger, 1990; Case, Kurland, & Goldberg, 1982; Kail, 1992; Swanson, 1999). It is, however, at least possible that the use of knowledge-based cues such as lexical status to discriminate potential target from non-target responses develops across the childhood years, in which case children may be less sensitive to the lexical similarity of processing and storage items than adults. The first experiment in the present series investigates complex memory span performance in 9- and 10-year-old children. The remaining two experiments involved both child and adult participant groups, in order to test the extent to which key findings generalize across age.

**Experiment 1**

**Method**

**Participants**

Eighteen children were drawn from year 5 of a local primary school in Stockton-on-Tees, England. They were all native English speakers and their ages ranged from 9 years 9 months to 10 years 8 months (mean age 10 years 3 months).

**Design and materials**

A set of 144 words and 144 nonwords, all of which had a one-syllable consonant–vowel–consonant structure, were used as processing and memory stimuli. The words were taken at random from the MRC Psycholinguistic Database, with the constraint that the mean age-of-acquisition for each word was under 5 years (from the norms of Gilhooly & Logie, 1982). This was to ensure a high degree of lexical familiarity with the word stimuli. The nonwords were drawn from the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002). Of the 144 items in each set, 18 items had the onset phoneme /k/ (e.g., cap, keb). The word and nonword sets were used to construct 42 lists for the processing task, each comprising three items. Each 3-item list contained zero, one, or two items with the onset phoneme /k/, unpredictably within the list. The consonant composition of the remaining items within each list was as distinctive as possible, i.e. within each processing sequence, the items contained different consonants. Recall stimuli were also drawn from the word and nonword pool, but did not include any of the items with the onset phoneme /k/. There were three lists each of two, three, four, and five memory items (a total of 42 items). There was no phonological overlap between memory items within a single list.

A two-way within-subjects design was employed with type of processing activity (word processing, nonword processing, articulatory suppression) and memory item (word, nonword) as independent variables, and memory span as the dependent variable. The recall conditions were blocked; half of the participants completed the word-recall conditions first, the other half completed the nonword-recall conditions first. The order of processing activities was counterbalanced across groups of participants.

**Procedure**

Each child was tested individually in a quiet area of the school. The experimental stimuli were presented on a laptop computer. In the word processing and nonword processing conditions, the sequence of three processing items interpolated between memory items was presented auditorily (read aloud by the experimenter) at a rate of approximately one item per second. The memory items
were presented auditorily and visually (items appeared in print on the screen and were read aloud by the experimenter). A sequence of processing items preceded the first memory item. The recall task was to remember the memory items displayed on the screen in the same order as presented. Children were also required to tap the table when they heard an item with the onset phoneme /k/ in the list of processing items.

In the articulatory suppression condition, children looked at the blank screen for 3 s while repeatedly saying the word ‘the’ aloud. A metronome was set to pace the children to say one ‘the’ every 750 ms. After 3 s, a memory item appeared on the screen and was read aloud by the experimenter. The children were instructed to suspend articulation while the item was on the screen. The memory item remained visible for 1 s; then the screen went blank. Again, children were requested to recall, in order, the items that had appeared on the screen. The experimenter recorded on a response sheet whether responses were correct or incorrect.

Testing began with three trials of two lists (i.e. two items for recall), followed by three trials of three lists, and so forth. The number of lists increased (to a maximum of five lists) until a child failed to recall correctly the memory items of all three trials at a particular level. Testing was discontinued at this point. Each child practiced the monitoring task, the articulatory suppression, and then one trial of processing plus recall, prior to testing.

Span was scored as follows: starting from a baseline score of one (in cases where none of the items from the two-list trials were correctly recalled), each correctly recalled memory item counted as one third; the total number of thirds was then added up to provide a span score. For example, the correct recall on all the trials of two items, of two lists of three items and two lists of four items yielded a span score of 1 + (3 + 2 + 2) × 1/3 = 3.33. Hence, the minimum score was 1.0, and the maximum score was 5.0.

Results

Table 1 presents the mean span scores for word and nonword recall across processing conditions. Recall of words was superior to nonwords in the nonword monitoring and articulatory suppression conditions, but not in the word monitoring condition. Word monitoring appeared to impair word recall, but nonword recall was uninfluenced by the lexicality of the processing material.

A 2 (memory item) × 3 (processing activity) within-subjects analysis of variance (ANOVA) was conducted on the span scores. All three terms were significant: memory item, $F(1,16) = 93.20$, $MSE = 34.85$, $p < .001$, partial $\eta^2 = 0.85$; type of processing, $F(2,32) = 122.01$, $MSE = 14.48$, $p < .001$, partial $\eta^2 = 0.88$; and the interaction between memory item and type of processing, $F(2,32) = 27.72$, $MSE = 6.12$, $p < .001$, partial $\eta^2 = 0.63$. Planned pairwise comparisons were conducted to explore differences in word recall and nonword recall across processing activities. Memory for words in the nonword processing condition was superior to that in the word processing condition, $t(17) = 11.71$, $p < .05$, $d = 2.44$, $95\% CI = -1.70$ to $-1.18$. The articulatory suppression condition produced significantly higher word spans compared to the word processing condition, $t(17) = 16.21$, $p < .05$, $d = 0.60$, $95\% CI = -2.26$ to $-1.74$, and to the nonword processing condition, $t(17) = 3.84$, $p < .05$, $d = 0.77$, $95\% CI = -0.86$ to $-0.25$. With regard to the recall of nonwords, articulatory suppression resulted in significantly higher spans compared to the word processing condition, $t(17) = 3.63$, $p < .05$, $d = 1.33$, $95\% CI = -0.82$ to $-0.22$, and the nonword processing condition, $t(17) = 4.00$, $p < .05$, $d = 0.95$, $95\% CI = -0.68$ to $-0.21$. There was no significant difference in nonword recall between word and nonword processing conditions, $t(17) = 0.50$, $p > .05$, $d = 0.18$, $95\% CI = -0.39$ to $0.24$.

Discussion

There were three key findings from Experiment 1. First, recall accuracy was markedly superior for words than for nonwords, establishing a lexicality effect in complex memory span that parallels that found in serial recall (e.g., Hulme et al., 1991). Second, word recall was selectively impaired by word processing. This finding is consistent with featural accounts of interference (Oberauer et al., 2004; Saito & Miyake, 2004), according to
which overlap between the semantic features of words encountered as memory items and as processing items will lead to loss of information in the word processing but not the nonword processing condition. Third, nonword recall was disrupted to an equivalent extent by both word and nonword processing relative to articulatory suppression. As the phonological content of the articulatory suppression activity was minimal compared with the two processing conditions, this result is entirely consistent with the view that interference in working memory can result from overwriting of shared features within the phonological domain. Equally, this could reflect the increased attentional demands of the phoneme monitoring conditions relative to articulatory suppression (Barrouillet & Camos, 2001).

The selective effect of word processing on word recall also fits well with a retrieval cue hypothesis, according to which participants’ knowledge of the lexical status of the memory items should allow them to discriminate effectively representations of target from non-target items. The plausibility of this account is, however, weakened by the lack of a corresponding decrement in nonword recall with nonword processing. If lexical status can be used to select likely target responses and reject non-target ones, nonword recall should be (but was not) disrupted most by monitoring nonword stimuli.

Experiment 2 was designed to replicate the findings from Experiment 1 using a computer-controlled stimulus presentation format. This experiment also included an adult group of participants as well as a further group of 9- and 10-year-old children, in order to establish whether the patterns of interference observed in Experiment 1 with children generalize to adults. Finally, a no-processing control condition was included in this experiment, in order to test whether suppressing articulation had a detrimental effect on span.

**Experiment 2**

**Method**

**Participants**

Sixteen children were drawn from year 5 of a local primary school in Stockton-on-Tees, UK. They were all native English speakers and their ages ranged from 9 years 10 months to 10 years 7 months (mean age 10 years 4 months). None of the children had participated in Experiment 1. The adult sample comprised 16 postgraduate students, with an age range of 23 years 10 months to 44 years 3 months (mean age 27 years 2 months).

**Design and materials**

The processing and storage stimuli were identical to those used in Experiment 1. In this experiment, however, the task was extended to include a no-processing control condition with a list of storage items only. As in the previous experiment, a two-way within-subjects design was employed with type of processing activity (word processing, nonword processing, articulatory suppression, control) and memory item (word, nonword) as independent variables, and span as the dependent variable. The recall conditions were blocked; half of the participants completed the word-recall conditions first, the other half completed the nonword-recall conditions first. The order of processing activities was counterbalanced across participants.

**Procedure**

The procedure was similar to that of Experiment 1. In this experiment, however, task duration and presentation of stimuli were computer controlled. In the word processing and nonword processing conditions, participants were instructed to look at a blank computer screen while a list of three items was presented auditorily, from a recording, at a rate of one item per second. As in Experiment 1, participants were instructed to tap the table whenever a presented item had the onset phoneme /k/. Following presentation of the final item in each list, the memory item appeared on the computer screen, and was also played aloud. The memory item remained on the screen for 1 s; then the screen went blank again. At the end of a trial, a question mark appeared on the screen, prompting participants to recall in serial order the items that had appeared. The articulatory suppression condition was almost identical to that in Experiment 1, except that here, the memory items were presented via an audio recording. In the control condition, participants were required to look at a blank screen for 3 s, after which a memory item appeared on the screen and was presented auditorily from a recording.

**Results**

Table 2 presents the mean span scores for word and nonword recall across the different processing activities for the two age groups. In both groups, there was a sizeable recall advantage for words over nonwords in the nonword monitoring, articulatory suppression and control condition that was eliminated with word monitoring. In adults but not in children, nonword recall was impaired when the monitoring task involved nonwords rather than words.

A 4 (processing activity) × 2 (memory item) by 2 (age group) analysis of variance was performed on the span scores. There were significant main effects of processing activity, $F(3,90) = 281.61$, $MSE = 42.58$, $p < .05$, partial $\eta^2 = 0.90$, and memory item, $F(1,30) = 280.92$, $MSE = 111.24$, $p < .05$, partial $\eta^2 = 0.90$, but not of age group, $F(1,30) = 2.99$, $MSE = 2.60$, $p > .05$, partial $\eta^2 = 0.09$. The processing activity by memory item
interaction was significant; $F(3, 90) = 120.59$, $MSE = 16.70$, $p < .05$, partial $\eta^2 = 0.80$, but none of the remaining interaction terms reached significance: processing activity by age, $F(3, 90) = 2.23$, $MSE = 0.34$, $p > .05$, partial $\eta^2 = 0.07$, memory item by age, $F(1, 30) = 0.44$, $MSE = 0.17$, $p < .05$, partial $\eta^2 = 0.01$, or processing activity by memory item by age, $F(3, 90) = 0.29$, $MSE = 0.04$, $p > .05$, partial $\eta^2 = 0.01$.

Planned pairwise contrasts were conducted in order to explore differences in word recall and nonword recall across processing activities for each age group. In the adult group, memory for words in the nonword processing condition was superior to that in the word processing condition, $t(15) = 11.01$, $p < .05$, $d = 1.97$, $CI = -1.03$ to -0.70. The control condition produced significantly higher word spans than the word processing condition, $t(15) = 25.61$, $p < .05$, $d = 7.70$, $CI = -3.29$ to -2.78, the nonword processing condition, $t(15) = 19.82$, $p < .05$, $d = 5.50$, $CI = -2.40$ to -1.94, and the articulatory suppression condition, $t(15) = 12.52$, $p < .05$, $d = 3.60$, $CI = -1.66$ to -1.18. In nonword recall, the control condition yielded significantly higher spans than the articulatory suppression condition, $t(15) = 7.45$, $p < .05$, $d = 1.03$, $CI = -0.72$ to -0.40, the word processing condition, $t(15) = 6.61$, $p < .05$, $d = 1.24$, $CI = -0.88$ to -0.45, and the nonword processing condition, $t(15) = 10.16$, $p < .05$, $d = 1.91$, $CI = -1.25$ to -0.81. In addition, there was a significant difference in nonword recall between word and nonword processing conditions, $t(15) = 3.93$, $p < .05$, $d = 0.58$, $CI = 0.17$ to 0.56, reflecting the lower levels of performance in the nonword processing condition. Comparisons between word recall and nonword recall across processing activities in the children’s data revealed that memory for words in the nonword processing condition was superior to that in the word processing condition, $t(15) = 6.20$, $p < .05$, $d = 1.88$, $CI = -1.47$ to -0.72. The control condition produced significantly higher word spans than the word processing condition, $t(15) = 26.03$, $p < .05$, $d = 6.08$, $CI = -3.41$ to -2.89, the nonword processing condition, $t(15) = 10.14$, $p < .05$, $d = 3.97$, $CI = -2.49$ to -1.62, and the articulatory suppression condition, $t(15) = 5.76$, $p < .05$, $d = 2.28$, $CI = -1.62$ to -0.74. In nonword recall, the control condition yielded significantly higher spans than the articulatory suppression condition, $t(15) = 4.54$, $p < .05$, $d = 0.41$, $CI = -0.38$ to -0.14, the word processing condition, $t(15) = 3.86$, $p < .05$, $d = 0.93$, $CI = -0.91$ to -0.26, and the nonword processing condition, $t(15) = 4.24$, $p < .05$, $d = 1.13$, $CI = -1.05$ to -0.38. There was, however, no significant difference in nonword recall between word and nonword processing conditions, $t(15) = 1.42$, $p > .05$, $d = 0.37$, $CI = -0.07$ to 0.33.

**Discussion**

The findings of Experiment 1 were replicated in Experiment 2: although children’s recall of words was selectively impaired by word processing compared with nonword processing, their nonword recall was disrupted to an equivalent extent by word and nonword processing. However, a contrasting pattern of findings was obtained for the adult participants. In both word and nonword recall, adults were disrupted to a greater extent when the processing stimuli shared the same lexical status as the memory items: by nonword monitoring in the case of nonword recall, and word monitoring in the case of word recall. Although the nonword decrement with nonword processing was not as substantial in magnitude as the corresponding word–word decrement, it was
nonetheless highly significant. In both age groups, processing led to lower levels of performance than articulatory suppression, which itself depressed accuracy relative to a no activity control condition.

These results establish a developmental change in similarity-based interference in children and adults. It was argued that the pattern of results obtained for the child group in Experiment 1 (and now also in Experiment 2) could readily be accommodated in terms of overwriting of overlapping semantic features of processing and storage items (Oberauer et al., 2004; Saito & Miyake, 2004). The selective impairment in nonword recall by nonword processing in the adult data, however, cannot be explained by such an account, as the degree of phonological overlap between the nonword memory items and both words and nonwords in the processing tasks was equivalent. Instead, these data are consistent with the proposal that similarity effects result from confusion between target and non-target representations at retrieval, due to the absence of a cue to discriminate potential target from non-target responses. The data from Experiment 2, therefore, suggest that knowledge of the non-lexical status of the memory stimuli was used by the adult but not the child participants to distinguish word representations generated in the processing condition from the target nonwords.

There is, however, another possible explanation for differences in the patterns of interference for the child and adult participants, and in particular why the disruptive effect of nonword processing in the recall of nonwords was absent in the younger age group. Memory spans for nonword lists in the child groups in both experiments were very low: in Experiment 2, for example, the mean span score in the word processing condition was 2.16 for the children, compared with 2.40 for the adults. A potential decrement in this condition in the children’s data may therefore have been masked by a floor effect in the data. A further problem with low scores from a span procedure is that very few trials are tested in total so that, for example, an individual with a span of two will have been tested only on six trials. Low measurement sensitivity may therefore have been the cause of the absence of an impairment in children’s nonword span scores when the processing activity involved processing nonwords, and may also have contributed to the reduced effects of similarity of lexical status on the recall of nonwords than words in the adult group. In addition, the lack of an age effect is somewhat surprising, given the findings of moderate though significant developmental increases in working memory capacity across adolescence in previous research (e.g., Gathercole, Pickering, Ambridge, & Wearing, 2004). It is possible that the low measurement sensitivity of the span paradigm used here contributed to the results.

In order to investigate this possibility, a further experiment was conducted that compared recall of words and nonwords under conditions of either word processing or nonword processing in adults and children. Experiment 3 employed a fixed list length procedure designed to overcome the differential sensitivity of the span procedure to age and experimental conditions. Measurement sensitivity was also boosted by increasing the number of trials tested at each individual list length from 3 to 6. All participants were tested on lists of two, three, and four recall items; adults were also tested on 5-item sequences in recognition of their greater memory spans.

One anticipated consequence of including supra-span list lengths in Experiment 3 was that substantial numbers of errors would be generated, providing the opportunity to test some specific predictions of the discrimination cue hypothesis. The first prediction is that error responses should typically show consistency with the lexical status of the memory items. Second, if the absence of a clear cue to differentiate potential target from non-target responses underlies the poor performance in conditions in which the memory and storage items share the same lexical status, there should be an increase in the frequency of incorrect recall of items encountered during the processing activity in these conditions. Note that although these predictions do not necessarily run contrary to interference-based theories of working memory, they are central to the cue discrimination hypothesis.

**Experiment 3**

**Method**

**Participants**

For the child sample, 16 children were drawn from year 5 of a local primary school in Stockton-on-Tees, UK. They were all native English speakers and their ages ranged from 9 years 10 months to 10 years 7 months (mean age 10 years 4 months). The adult sample in this experiment comprised 16 undergraduate and postgraduate students. They were all native English speakers, and their ages ranged from 19 years 11 months to 44 years 3 months, with a mean age of 23 years 4 months.

**Design and materials**

The processing and memory stimuli were taken from the same pool of items as Experiments 1 and 2. The lists for the processing task contained five items (words or nonwords), of which either zero, one, or two items began with the phoneme /k/. As in the previous experiments, each list had an associated memory item that was presented at the end of the list, but was not part of the processing task. None of the memory items had the onset phoneme /k/. The number of memory items to
be recalled in serial order varied in length: six trials each of two, three and four items; for adults, the number of items to be remembered included further trials of five items. The recall conditions were blocked; half of the participants completed the word-recall conditions first, the other half completed the nonword-recall conditions first. The order of processing activities was counterbalanced across participants.

Procedure

Each participant was tested individually. Presentation of the experimental stimuli was controlled by a laptop computer. The sequence of five processing items interpolated between memory items was presented auditorily at a rate of one item per second; memory items were presented both auditorily and visually (in print on the computer screen) one second after the fifth item in each sequence of processing stimuli. A sequence of processing items preceded the first memory item. The recall task was to remember the memory items displayed on the screen in the same order as presented. Participants were also required to tap the table when they heard an item with the onset phoneme /k/ in the list of processing items. Testing began for children and adults with six trials of lists containing two memory items. List length increased by one item over successive blocks of six trials, with testing ceasing for children at list length four, and for adults at list length five. Responses were recorded manually by the experimenter at the time of testing.

Results

A strict serial recall criterion was adopted, according to which an item was only scored as correct if it was recalled in its original position in the sequence. Recall responses were further subclassified into the following categories: an order error was recorded when a memory item was recalled in a different position in the list at output than at the original presentation. A memory intrusion was recorded when an item from another list in the same experimental condition was recalled. A processing intrusion was recorded when an item encountered in processing tasks was recalled; this error category was further subclassified as either a processing item from the same trial or from another trial in the same experimental condition. A novel intrusion occurred when an item that was not present in the same experimental condition was recalled; these errors were further subclassified as either word or nonword responses. The final error category was blank response, occurring when the participant did not recall any item at a particular list position.

Recall accuracy in each condition for both age groups is summarized in Table 3, which shows the mean proportion of items correctly recalled at each list length as a function of lexicality of processing item, lexicality of recall item, and list length. In both age groups, recall accuracy declined with increasing list length. The lexicality effect (superior recall of words over nonwords) was present only in the nonword processing condition for both age groups. In the word processing condition, this effect was eliminated in the child group (word and nonword recall at .48 in both conditions), and reversed in the adult group, where a higher proportion of nonwords (.52) was recalled than words (.46) when preceded by word processing. Nonword recall was lower with nonword than word processing in both age groups at all list lengths.

A four-way analysis of variance compared adults' and 10-year-olds' accuracy scores as a function of lexicality of memory item, lexicality of processing item, and list length. Necessarily, this analysis only included data from the list lengths (2, 3 and 4) completed by both participant groups. All four main effects were significant: age, \( F(1,30) = 67.95, \quad MSE = 1.01, \quad p < .05 \); partial \( \eta^2 = 0.69 \); lexicality of processing item, \( F(1,30) = 8.40, \quad MSE = 0.15, \quad p < .05 \); partial \( \eta^2 = 0.22 \); lexicality of memory item, \( F(1,30) = 97.95, \quad MSE = 2.00, \quad p < .05 \); partial \( \eta^2 = 0.77 \); and list length, \( F(2,60) = 434.18, \quad MSE = 9.04, \quad p < .05 \); partial \( \eta^2 = 0.94 \). These terms reflect, respectively, the greater recall accuracy of the adults than the children, of memory items following nonword processing than word processing, of word than nonword lists, and of short than long sequences. Significant interactions were obtained between list length and age, \( F(2,60) = 19.93, \quad MSE = 0.42, \quad p < .05 \); partial \( \eta^2 = 0.40 \); between processing item and memory item, \( F(1,30) = 202.33, \quad MSE = 2.69, \quad p < .05 \); partial \( \eta^2 = 0.87 \); between processing item, memory item, and age, \( F(1,30) = 11.80, \quad MSE = 0.16, \quad p < .05 \); partial \( \eta^2 = 0.28 \); between processing item, memory item, and list length, \( F(2,60) = 16.31, \quad MSE = 0.13, \quad p < .05 \); partial \( \eta^2 = 0.35 \); and between processing item, memory item, list length, and age, \( F(2,60) = 4.52, \quad MSE = 0.04, \quad p < .05 \); partial \( \eta^2 = 0.13 \).

Further analyses were conducted to investigate detailed interactions between the lexicality of processing and memory items. Collapsed across age group, nonword recall was significantly lower with nonword than word processing, \( t(31) = 6.90, \quad p < .05 \); \( d = 2.35 \); \( CI = 0.09 \) to 0.17; and word recall was significantly lower following word than nonword processing, \( t(31) = 9.82, \quad p < .05 \); \( d = 7.70 \); \( CI = -0.25 \) to -0.16. In order to explore the significant three-way interaction found between lexicality of processing and memory items with age, two further analyses were performed. First, a 2 (age) \( \times \) 2 (processing item) analysis of variance on word recall was conducted, yielding a main effect of processing item, \( F(1,30) = 137.79, \quad MSE = 0.69, \quad p < .05 \); partial \( \eta^2 = 0.82 \) that reflected the higher scores in nonword processing condition (\( m = 0.71 \) than in the
word processing condition \( (m = 0.51) \). There was also a main effect of age, \( F(1,30) = 87.84, \ RMSE = 0.23, \ p < .05, \ partial \eta^2 = 0.75, \) with adults recalling more \( (m = 0.67) \) than children \( (m = 0.55) \). The interaction between processing and age was significant, \( F(1,30) = 14.28, \ RMSE = 0.07, \ p < .05, \ partial \eta^2 = 0.32; \) this was due to the superior recall of adults in the nonword processing condition only. A corresponding 2 (age) × 2 (processing item) analysis of variance performed on the nonword recall data yielded a significant main effect of processing item, \( F(1,30) = 46.96, \ RMSE = 0.26, \ p < .05, \ partial \eta^2 = 0.61, \) due to superior recall in the word processing condition \( (m = 0.53) \) than the nonword processing condition \( (m = 0.40) \). There was a main effect of age, \( F(1,30) = 12.58, \ RMSE = 0.11, \ p < .05, \ partial \eta^2 = 0.30, \) with adults recalling more \( (m = 0.51) \) than children \( (m = 0.43) \) The interaction between processing and age was nonsignificant, \( F < 1. \)

Finally, a series of \( t \)-tests comparing nonword recall in the word and nonword processing conditions were conducted separately for each age group and list length, in order to establish under what conditions precisely disruptive effects of nonword processing on nonword recall were found. Significantly higher recall with word than nonword processing was found in the adult group at list lengths 3, 4 and 5, and for the children at list length 3 only \( (p > .05 \) for all remaining contrasts). The frequency of error responses in each category as a function of condition and for each group, collapsed across all list lengths, is shown in Table 4. Levels of performance were similar across age groups, with average recall accuracy of 49.7% for children and 51.7% for adults. The most common category of error was a blank response, constituting 35.6% of all responses for children and 31.8% for adults. Order errors (migrations of memory items to non-target positions at recall) constituted 8.3% of responses for children and 11.2% for adults. Intrusion errors were less common, comprising 6.3% of errors for children and 5.3% of errors for adults. The lexical consistency between memory items and the errors responses was extremely high, at 100% of the errors sharing the same lexical status as the memory items for both children and adults under word-recall conditions, 96.9% for nonword recall in adults, and 92.1% for nonword recall in children.

In order to compare the distributions of errors of each kind across conditions, a series of 2 (word recall, nonword recall) × 2 (word processing, nonword processing) \( \chi^2 \) analyses were performed on the frequencies of

<table>
<thead>
<tr>
<th>Recall item</th>
<th>Word</th>
<th>Nonword</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Adults</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Processing item</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word</td>
<td>2</td>
<td>.81</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.46</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.34</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.22</td>
</tr>
<tr>
<td>Mean</td>
<td>.46</td>
<td>.13</td>
</tr>
<tr>
<td>Nonword</td>
<td>2</td>
<td>.99</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.80</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.64</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>.44</td>
</tr>
<tr>
<td>Mean</td>
<td>.72</td>
<td>.13</td>
</tr>
<tr>
<td>Total</td>
<td>.59</td>
<td>.13</td>
</tr>
<tr>
<td>10-year-old</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Processing item</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word</td>
<td>2</td>
<td>.83</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.38</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.23</td>
</tr>
<tr>
<td>Mean</td>
<td>.48</td>
<td>.10</td>
</tr>
<tr>
<td>Nonword</td>
<td>2</td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>.58</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>.34</td>
</tr>
<tr>
<td>Mean</td>
<td>.62</td>
<td>.12</td>
</tr>
<tr>
<td>Total</td>
<td>.55</td>
<td>.12</td>
</tr>
</tbody>
</table>
each of the principal error categories, for each age
group. Consider first the order errors. Their frequency
did not vary across experimental conditions for either
children, \( \chi^2 < 1 \), or adults, \( \chi^2 = 1.477, p > .05, w = 0.30 \), reinforcing previous findings from serial recall
that the lexicality of memory items influences the accura-
cy or item rather than order memory (Gathercole et al.,
2001).

Although the frequency of blank responses did not
vary as a function of experimental condition for the chil-
dren, \( \chi^2 < 1 \), it did for the adults, \( \chi^2 = 63.956, p < .001, w = 2.0 \). The latter term reflected the increased frequency of blank responses in the conditions in which the memory and processing items shared lexical status—for word as opposed to nonword processing in word recall, and for nonword compared with word processing in nonword recall. Comparisons of the distributions of errors across the two age groups established that the frequency of blank responses in word recall following word processing was significantly increased in the adults compared with the children, \( \chi^2 = 28.501, p < .001, w = 0.94 \), whereas there was no significant difference in blank responses across the two nonword-recall conditions across age group, \( \chi^2 < 1 \).

The distribution of intrusion errors varied systemati-
cally across conditions in both age groups, with many
more intrusion errors in word recall in the word than
nonword processing conditions in both age groups. In
the adult data, a corresponding increase in the frequency
of intrusion errors was also apparent in nonword recall
with nonword rather than word processing; this effect
was somewhat weaker in the child data. The frequency
of intrusion errors was investigated in a series of further
analyses. In an initial analysis, possible differences in the
frequency of total intrusion errors (collapsed across the
subtypes of memory, processing, and novel) across the
two factors of recall (word, nonword categories) and
processing (word, nonword) were explored. Significant
differences were found, in both children, \( \chi^2 = 71.775, p < .001, w = 2.12 \), and adults, \( \chi^2 = 169.497, p < .001, w = 3.25 \). Further analyses were performed for the two age groups in each of the word recall and nonword-recall conditions. No significant difference across the groups was found in the effect of the lexicality of pro-
cessing material in word recall, \( \chi^2 < 1 \), although there
was a highly significant group difference in the frequency
of intrusions across the two nonword recall conditions,
\( \chi^2 = 20.237, p < .001, w = 0.80 \). This reflects the large
increase in intrusions in the nonword processing condi-
tion in the adults, but not the children. Further 2 (mem-
ory) \times 2 (processing) analyses performed separately for
the two age groups established highly significant differ-
ences in the distributions of both memory intrusions
and processing intrusions across conditions for both
groups (\( p < .001, \) in each case). For the memory intru-
sions, the increased frequency of intrusions in word recall with word than nonword processing did not differ significantly across the groups (\( \chi^2 < 1 \)), although the increase in memory intrusions with nonword processing in nonword recall was significantly greater for adults than children (\( \chi^2 = 11.843, p < .001, w = 0.61 \)). No signif-
ificant differences across age groups were found in the
corresponding analyses of the processing intrusions.

Note that the novel intrusion data were not analyzed
separately, due to the low frequency of this category of
error.

In a final set of analyses, the frequency of error
responses that were words as opposed to nonwords
was compared across age groups, separately for the
word-recall and nonword-recall conditions. The high

Table 4
Frequency of responses in each category in Experiment 3 for children and adults, collapsed across list lengths

<table>
<thead>
<tr>
<th>Recall:</th>
<th>Children</th>
<th></th>
<th></th>
<th>Adults</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing:</td>
<td></td>
<td>Word</td>
<td>Nonword</td>
<td>Word</td>
<td>Nonword</td>
<td>Word</td>
</tr>
<tr>
<td>Correct</td>
<td>414</td>
<td>536</td>
<td>415</td>
<td>354</td>
<td>618</td>
<td>968</td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blank response</td>
<td>235</td>
<td>246</td>
<td>348</td>
<td>403</td>
<td>431</td>
<td>236</td>
</tr>
<tr>
<td>Order</td>
<td>99</td>
<td>75</td>
<td>61</td>
<td>52</td>
<td>171</td>
<td>131</td>
</tr>
<tr>
<td>Intrusion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Memory other trial</td>
<td>49</td>
<td>6</td>
<td>20</td>
<td>16</td>
<td>55</td>
<td>7</td>
</tr>
<tr>
<td>Processing same trial</td>
<td>30</td>
<td>0</td>
<td>4</td>
<td>11</td>
<td>37</td>
<td>0</td>
</tr>
<tr>
<td>Processing other trial</td>
<td>28</td>
<td>0</td>
<td>2</td>
<td>8</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>Processing total</td>
<td>58</td>
<td>0</td>
<td>6</td>
<td>19</td>
<td>61</td>
<td>0</td>
</tr>
<tr>
<td>Novel word</td>
<td>8</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Novel nonword</td>
<td>1</td>
<td>0</td>
<td>11</td>
<td>13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Novel total</td>
<td>9</td>
<td>1</td>
<td>14</td>
<td>20</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>Intrusions total</td>
<td>116</td>
<td>7</td>
<td>40</td>
<td>55</td>
<td>124</td>
<td>9</td>
</tr>
</tbody>
</table>

Note: Total responses in each condition: children 864, adults 1344.
frequency of word errors in word recall was equivalent for both groups, $\chi^2 < 1$. There was, however, a significant group difference in the nonword recall data, reflecting the greater frequency of nonword error responses in nonword recall for the adults than children, $\chi^2 = 12.060$, $p < .001$, $w = 0.61$.

Discussion

Experiment 3 replicated findings of substantial disruptions in word recall by word as opposed to nonword monitoring in children and adults from Experiments 1 and 2, extending their generality from a span paradigm to a fixed list length procedure that included supra-span sequence lengths. This pattern of results is consistent with proposals of interference between semantic features activated for the memory and processing items (Oberauer et al., 2004; Saito & Miyake, 2004). In this experiment, in contrast to both Experiments 1 and 2, a parallel disruptive influence of nonword monitoring on nonword recall was also found in both age groups. This result cannot readily be accounted for in terms of either lexically or phonologically based interference. It does, however, fit well with the hypothesis that the known lexical status of the memory items can be used as a cue to discriminate potential target from non-target responses at the time of retrieval. By this account, representations of processing stimuli and other non-target representations cannot be rejected if their lexical status corresponds to that of memory stimuli, leading to increased frequency of error responses.

A key issue concerns why children of equivalent ages showed no sensitivity to the lexical status of processing material when recalling nonwords in the two earlier experiments, but did so in Experiment 3. There are at least two factors that may be at play here. First, measurement sensitivity was greatest in this final experiment, due to the employment of a fixed list length procedure with multiple list lengths that prevented scaling effects in the data and ensured equal numbers of trials in each condition. No significant impairments in nonword recall in Experiment 3 following nonword processing were found in either age group at list length 2, at which performance approached ceiling levels, or at list length 4 for the children where performance levels were very low. Nonword processing deficits were, however, found at list lengths 3, 4, and 5 for adults, and at list length 3 for children.

Nonetheless, scrutiny of the qualitative patterns of error across nonword-recall conditions in the two age groups indicates that there were genuine developmental differences in Experiment 3 that cannot readily be accounted for by scaling factors, as recall accuracy was very similar for children and adults (50% and 52%, respectively). Some features of performance were common to both the child and adult groups. In particular, both groups showed a high degree of lexical consistency between the incorrect recall attempts and the memory items as predicted by the discrimination cue hypothesis, in both cases. Intrusions from both other memory and processing items were much higher when word recall was paired with word than nonword processing. However, in adults a corresponding pattern of increased intrusion errors and blank responses was found in nonword recall following nonword compared to word processing whereas in children, this pattern of increased vulnerability of recall when both memory and processing items are nonwords was much less marked, with no sizeable increase in the number of intrusion errors in the nonword over the word processing condition. Also, the frequency with which nonword and word errors were generated in recalling nonwords in the two monitoring conditions was equivalent for children. These results indicate that the non-lexical status of nonword memory items was not as effectively used as a cue for differentiating potential target from non-target representations in the child as the adult group.

One limitation of Experiments 1 to 3 was that performance on the processing tasks was not monitored. It is therefore not possible to determine whether the reduced interference observed when memory and processing stimuli had different rather than common lexical status arose from a tradeoff between resources allocated storage and processing. Specifically, it may have been the case that the superior memory spans found when the lexical status differed may have been mirrored by a decrease in the efficiency of processing. Such an outcome would substantially undermine claims that lexical status was used to benefit the discrimination of representations of memory from processing items. A final experiment employing adult participants was therefore conducted in which processing efficiency (the accuracy of phoneme monitoring) was monitored under conditions in which the lexicality (words, nonwords) of both the memory stimuli and processing items was orthogonally varied. Findings that memory performance is enhanced under conditions in which the lexicality of the memory and processing items differ but that processing accuracy does not vary by conditions would lend strong support for the lexical consistency hypothesis. However, such a view would be substantially challenged if improvements in memory performance are found to be accompanied by decreases in processing efficiency, indicating that the results reflect strategic tradeoffs between resources allocated to processing and storage.

Experiment 4

Method

Participants

Sixteen adults, whose ages ranged from 18 years 7 months to 22 years 5 months (mean age 20 years 1
A two-way analysis of variance was performed on the recall data as a function of the lexicality of memory and processing items. The significant main effect of memory item, $F(1,15) = 54.33$, $MSE = 0.55$, $p < .05$, partial $\eta^2 = 0.78$, reflected the superior recall accuracy of word over nonword items. Although the effect of processing item was nonsignificant, $F(1,15) = 2.88$, $MSE = 0.03$, $p > .05$, partial $\eta^2 = 0.16$, the interaction between memory and processing type was significant, $F(1,15) = 17.38$, $MSE = 0.41$, $p < .05$, partial $\eta^2 = 0.14$. Simple effects conducted to explore this interaction term. No significant difference was found between word and nonword recall following word monitoring, $F < 1$. In contrast, recall of words was significantly higher than recall of nonwords following nonword monitoring, $F(1,15) = 43.61$, $MSE = 0.95$, $p < .05$, partial $\eta^2 = 0.74$. Word recall was significantly higher following nonword processing than word processing, $F(1,15) = 19.43$, $MSE = 0.33$, $p < .05$, partial $\eta^2 = 0.56$, and recall of nonwords was significantly higher following word than nonword processing, $F(1,15) = 6.19$, $MSE = 0.12$, $p < .05$, partial $\eta^2 = 0.29$.

Experiment 4, therefore, replicated the principal findings from adult participants in both Experiments 2 and 3 that in recalling both word and nonword lists, performance was impaired to a greater extent when the processing items shared the same lexical status as the memory items. In contrast, processing accuracy did not vary across experimental conditions in Experiment 4, despite levels of performance that fell considerably below 100%. The data from this experiment therefore provide no evidence for a tradeoff between processing and storage resources.

**General discussion**

Four experiments established marked and consistent patterns of interactions between the lexicality of the memory and processing items in a complex memory paradigm, in both children and adults. In each experiment, recall of lists composed either of word or nonword stimuli, and participants performed a simple phoneme monitoring activity interleaved between successive presentation of memory stimuli. In the critical conditions, memory and processing items were either words or nonwords. Recall of words was substantially disrupted when participants monitored sequences of words rather than nonwords interpolated between memory items. Under conditions of no interpolated processing, articulatory suppression, and nonword processing, recall was superior for word than nonword lists. This phenomenon corresponds to the lexicality effect in immediate serial recall, which is widely attributed to the additional availability of lexical representations to supplement temporary phonological storage of verbal memory items.
(e.g., Gathercole et al., 2001; Hulme et al., 1991). However, this lexicality effect in complex memory performance was abolished when the processing activity involved words. This disruptive influence of word processing on word recall occurred in both memory span (Experiments 1 and 2) and fixed list length (Experiments 3 and 4) procedures, and was present in groups of 9- and 10-year-old children (Experiments 1, 2, and 3) and adults (Experiments 2, 3, and 4). The selective deficits in adults found in word recall with word monitoring, and in nonword recall with nonword monitoring, were not accompanied by shifts in monitoring accuracy, and so did not appear to reflect strategic changes in accuracy tradeoffs.

This finding is entirely consistent with feature-based theories of working memory such as those of Saito and Miyake (2004) and Oberauer et al. (2004), according to which stimuli are represented as constellations of features within a variety of dimensions that are overwritten by other stimuli with overlapping features. Encountering familiar words during the monitoring activity would be expected to activate their associated semantic features, leading to degradation in overlapping semantic representations of the items to be remembered. For example, in the word recall and word processing condition of the final experiment, participants encountered five times as many words in the processing intervals as words to be recalled, generating a substantial degree of potential semantic interference.

An alternative explanation that fits well with the unanticipated abolition of the lexicality effect under conditions of word monitoring in these experiments is that the disruptive effect of the irrelevant words arises from a disturbance of the redintegrative process suggested to underpin the lexicality effect. One possibility is that the lexical representations activated by the processing material give rise to false redintegration, a process whereby a lexical entry that was not a memory item is erroneously selected to reconstruct an incomplete phonological memory trace. Given the common CVC pattern shared by all memory and processing stimuli, such false completions seem quite likely. Note that the key difference between the two accounts of the lexical interference effect considered here concerns the stage at which the processing of words disrupts memory performance: whereas feature-based interference accounts attribute this effect to disturbance of the array of attributes that represent the memory item, the redintegration accounts locate this effect in the subsequent process of retrieval. More detailed empirical investigations are required to distinguish between these theoretical alternatives.

The converse finding that nonword recall is impaired when processing involves monitoring nonwords rather than words—consistently in adults and less robustly in children—cannot readily be accommodated by either of these accounts. This aspect of the results fits well instead with the hypothesis that selection of appropriate candidates for recall at the point of retrieval is facilitated by the availability of a salient cue that allows the effective discrimination of potential target from non-target items. In the context of the present experiments, the lexical status cue allows participants in the nonword recall and word processing condition to reject any representations with lexical status, and also to reject any non-lexical items in the condition in which word recall is accompanied by nonword processing. It is suggested that the absence of corresponding similarity effects within content domains in Oberauer et al. (2004) may have been due to the relatively low salience and discrimination power provided in the similar conditions in this study. Experimental manipulations designed to provide simple and effective cues for distinguishing potential target from non-target responses may yield similarity effects within both the phonological and semantic domains, although this will be due to the use of discrimination cues rather than interference per se. It should also be noted that the present findings do not necessarily contradict the possibility that lexical status could be used at encoding, rather than at retrieval, to mark items for later recall.

Several features of the error data from Experiment 3 are consistent with this lexical discrimination cue hypothesis. First, the vast majority of substitution errors preserved the lexical status of the memory items, whether words or nonwords. Second, in conditions in which memory and processing items were either both words or both nonwords, intrusions by non-target items were common. Interestingly, error responses corresponding to stimuli that were not present in any phase of the experiment were rare in any condition, suggesting that confusion as to which activated representations correspond to memory stimuli is the major source of error in this paradigm. In adults, performance when the memory and processing items shared the same lexical status was also characterized by a large increase in the frequency of blank responses, where the participant was unable to generate any response at a particular position in the response protocol. Such null responses could either arise from excessive numbers of activated representations in the absence of any coherent strategy for response selection, or may be consequences of intrusion errors at the preceding list position that would disrupt the use of item–item associations. However, clear interpretation of this feature of the data is confounded by the absence of a corresponding increase in the frequency of blank responses in the child group in the word recall and word processing condition, which was also associated with a decrement in recall accuracy.

Evidence that lexical status acts as a cue to discriminate potential target from non-target representations was less strong in the 9- and 10-year-old children tested in these experiments than in adults. In the complex span
tasks employed in Experiments 1 and 2, no significant differences in nonword recall were found in the nonword and word processing conditions. In Experiment 3, which used a more sensitive fixed list length procedure that included supra-span lengths, a disruptive influence of nonword processing on nonword recall was found for children with three-item lists. However, qualitative analysis of the errors indicated that intrusions in nonword recall by other nonwords were not much more common following nonword than word monitoring in children. This contrasts both with the large numbers of intrusion errors present in the word recall word processing condition in the same group, and in the corresponding nonword conditions in the adults. One possible explanation for this is that the phonological features of words have a relatively stronger cue value—and are therefore more salient distractors—for children than for adults.

We suggest that the findings of this series of experiments are best characterized in terms of two distinct cognitive processes. The first process leads to the disruption of word recall following word processing, and may arise either from interference or degradation resulting from the activation of irrelevant lexical representations (e.g., Oberauer et al., 2004; Saito & Miyake, 2004) or from false redintegration at the time of memory retrieval. This leads to the loss of strong and distinctive lexical representations that support the lexicality effect, and underlies the disruptive effect of processing irrelevant words interpolated between lexical memory items. On the basis both of lexicality effects in serial recall established in children as young as 4 years of age (Gathercole et al., 2004) and the present findings of consistent influences of word processing on word recall in children and adults, it is proposed that this process represents a fundamental property of the working memory system that is present from an early age. It is also proposed that this process underpinned the selective interference between digit and non-digit stimuli across memory and processing tasks reported by Conlin et al. (2005a), which was found to be present in 6-year-old children.

The second process involves the use of cues to discriminate target from non-target representations. This process, we argue, is strategic rather than fundamental in nature and emerges gradually across development. These conclusions fit well with the large body of metamemory research indicating that use of strategies for optimizing memory performance emerges relatively late during the childhood years (DeMarie & Ferron, 2003; Schneider & Pressley, 1997). In adults, word recall is constrained by a combination of the lexically based process and cue-based selection when memory and processing stimuli share a common lexical status. In contrast, nonword recall is constrained by a single mechanism of cue-based discrimination when processing involves nonwords. In 9- and 10-year-old children, the lexically based process is fully operational, but the cue-based selection strategy is not.

A more general issue raised by these findings concerns the extent to which they reflect mechanisms or processes that are specific to working memory, or are properties of other components of the cognitive system that are tapped in the course of complex memory paradigms. On balance, we believe that the data favor the latter position. There is already substantial evidence that experimental phenomena characteristic of verbal short-term memory such as the word length and phonological similarity effects are present also in complex memory paradigms (LaPointe & Engle, 1990; Lobley et al., 2005). The present findings indicate that the lexicality effect is also equivalent in both complex memory and serial recall paradigms, suggesting that a common redintegration process is applied in both cases. A similar case can be made regarding semantic interference between memory and processing stimuli. The lexical-semantic knowledge accessed by familiar words cannot plausibly be located in the temporary storage capacities of working memory, although their activated features may well be conceived as part of this system. Disruptive effects of semantic similarity of target and distractor stimuli have been established in a variety of paradigms, including picture naming (Damian & Bowers, 2003; Vigliocco, Vinson, & Siri, 2005), word naming (Colangelo, Buchanan, & Westbury, 2004), and cross-language translation (Bloem, van den Boogaard, & La Heij, 2004). Thus on the grounds of parsimony at least, it seems appropriate to characterize the corresponding interfering effects of exposure to irrelevant word on word recall in terms of the general processes of competition between activated conceptual representations that contribute to performance on working memory tasks as well as other language-related activities.

To what extent, then, should the complex task methodology employed in these experiments be viewed as tapping processes specific to a working memory system at all? It is suggested here that the most useful conception of working memory may be a broad one that incorporates the full range of current processing and temporary storage activities that is associated with a particular cognitive activity. Inevitably, this ‘definition’ of working memory will vary according to task. Experimental research over the past decade designed to isolate specific processes in working memory has generated many apparently contradictory findings that challenge the credibility of simple conceptualizations of working memory constraints (Barrouillet et al., 2004; Conlin, Gathercole, & Adams, 2005b; Saito & Miyake, 2004; Towe et al., 2005). Perhaps the key point of convergence across studies is the multiplicity of cognitive resources, processes and strategies that are available within the cognitive system to support the capacities to
retain material over a short period of time. Within this framework, the contribution of the present series of experiments is to demonstrate even within a tightly constrained complex memory paradigm, the operation of two distinct sources of interference between storage and processing items.

References


