Cognition 144 (2015) 1–13

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Questioning short-term memory and its measurement: Why digit span measures long-term associative learning

Gary Jones [a,*], Bill Macken [b]

[a] Nottingham Trent University, UK
[b] Cardiff University, UK

ARTICLE INFO

Article history:
Received 23 December 2014
Revised 13 July 2015
Accepted 15 July 2015
Available online 23 July 2015

Keywords:
Digit span
Associative learning
Short-term memory
Long-term memory
Sequence learning
Computational modelling

ABSTRACT

Traditional accounts of verbal short-term memory explain differences in performance for different types of verbal material by reference to inherent characteristics of the verbal items making up memory sequences. The role of previous experience with sequences of different types is ostensibly controlled for either by deliberate exclusion or by presenting multiple trials constructed from different random permutations. We cast doubt on this general approach in a detailed analysis of the basis for the robust finding that short-term memory for digit sequences is superior to that for other sequences of verbal material. Specifically, we show across four experiments that this advantage is not due to inherent characteristics of digits as verbal items, nor are individual digits within sequences better remembered than other types of individual verbal items. Rather, the advantage for digit sequences stems from the increased frequency, compared to other verbal material, with which digits appear in random sequences in natural language, and furthermore, relatively frequent digit sequences support better short-term serial recall than less frequent ones. We also provide corpus-based computational support for the argument that performance in a short-term memory setting is a function of basic associative learning processes operating on the linguistic experience of the rememberer. The experimental and computational results raise questions not only about the role played by measurement of digit span in cognition generally, but also about the way in which long-term memory processes impact on short-term memory functioning.

1. Introduction

A short-term, limited capacity system for the temporary maintenance and manipulation of information has formed an integral component of cognitive architectures since the foundations of cognitive science. It is typically construed as a mode of processing separate from, but interactive with long-term memory and the particular form this relationship takes has been theorized, broadly, in two ways; short-term memory (STM) is either seen as a set of processes and representations discrete from long-term memory, with the interaction involving both formation of new long-term memory representations and support from existing ones (e.g., Baddeley, Gathercole, & Papagno, 1998; Hulme et al., 1997; Page & Norris, 2009) or STM is that aspect of long-term memory that is currently activated by some sort of limited-capacity attentional process (e.g., Cowan, 1995; MacDonald & Christiansen, 2002; Martin & Saffran, 1997).

For both of these broad approaches, a key focus is on the influence of long-term linguistic knowledge on performance in verbal STM tasks. In general terms, STM for verbal material that corresponds to the rememberer's linguistic knowledge is better than that for material that deviates from it: serial recall for sequences of words is better than that for nonwords (Hulme, Maughan, & Brown, 1991) and recall of sequences of high frequency words is better than that for sequences of low frequency words (Hulme et al., 1997); sequences of nonwords constructed to conform to the phonotactic regularities of the rememberer's own language sustain better serial recall than those that do not (Majerus, van der Linden, Mulder, Meulemans, & Peters, 2004); sequences of words in the rememberer's first language are recalled better than those from the second language (Messer, Leseman, Boom, & Mayo, 2010), and so on.

The two approaches also share a general explanatory orientation in that they both attribute (albeit in different ways) the advantage for linguistically familiar material to enhanced integrity of the

---

* Corresponding author at: Division of Psychology, Nottingham Trent University, Chaucer Building, Nottingham NG1 4BU, UK.

E-mail address: gary.jones@ntu.ac.uk (G. Jones).

http://dx.doi.org/10.1016/j.cognition.2015.07.009
0010-0277/© 2015 The Authors. Published by Elsevier B.V.

items that form the memory sequences. For those accounts of STM that regard it as a mode of representation discrete from long-term memory, a process of redintegration is argued to take place (at encoding, storage or retrieval) such that long-term lexico-phonological representations may be used to counteract the effects of decay and/or interference that have degraded the corresponding volatile representations in short-term storage (e.g., Hulme et al., 1997; Schweickert, 1993). Those accounts that view STM as the currently activated portion of long-term memory attribute the superior performance for linguistically familiar phonological material to additional, sustained activation that accrues from the mutual connections that exist between the lexical and semantic features of words and their phonological features, as well as the increased integrity – due to frequent co-activation – of the phonological features of familiar words compared to novel or infrequent ones (e.g., Jefferies, Frankish, & Noble, 2009).

However, the influence of long-term linguistic knowledge has also been shown to operate on sequence-level factors that transcend the particular characteristics of the individual items making up those sequences. For example, sequences of alternating pairs of adjectives and nouns are better remembered if the adjective–noun ordering corresponds to that found in the rememberer's language (Acheson & MacDonald, 2009; Perham, Marsh, & Jones, 2009). Also, sequences within which the coarticulatory transitions between successive items are relatively fluent sustain better serial recall than sequences containing more complex or unfamiliar transitions, even when the items making up those sequences are equivalently familiar (Murray & Jones, 2002; Woodward, Macken, & Jones, 2008). The influence of such sequence level factors is not typically integrated into models of STM (Burgess & Hitch, 1999; Henson, 1998; Nairne, 1990; Page & Norris, 1998), within which the ordering of successive items is accomplished by some sort of order or positional cue (e.g., a primacy gradient or an oscillating context signal) which is implemented separately from the items making up the sequence.

An exception to this picture relates to the influence of pre-test exposure to particular sequential regularities of material that is subsequently subjected to serial recall or the repetition within an experimental block of particular orders of items (e.g., Burgess & Hitch, 2006; Hebb, 1961; Majerus, Martinez Perez, & Oberauer, 2012). The focus in such settings has typically been on the transmission of information from short-term processing to long-term memory – for example, with respect to the learning of new multisyllabic words (e.g., Baddeley et al., 1998; Page & Norris, 2009). The effect of recent exposure to novel transitional regularities on the subsequent serial recall of sequences corresponding to such regularities raises the question of the impact of long-term sequence learning within the short-term setting (e.g., Botvinick & Bylsma, 2005; Majerus et al., 2012). However, conceptual orientations that propose bespoke short-term memory processes distinct from other learning mechanisms seek to ensure that such factors are nullified within STM methodologies by deliberately excluding sequences which might correspond to the linguistic repertoire of the rememberer (e.g., common acronyms, canonical runs of letters or digits). In such a way, the study of STM appears to immunize itself from certain aspects of long-term linguistic knowledge (i.e., those that pertain to the level above the ostensible item), while embracing the influence of others (those that pertain to the item). In particular, it methodologically seeks to immunize itself from potential influences of what we would conceive of as long-term associative influences (other than those obtained within the experimental setting, such as the Hebb effect), which is to say that in a setting that examines rememberers' ability to retain and manipulate sequences of verbal information, the question of whether or not those sequences (as opposed to the items they are drawn from) are familiar is excluded from the analysis.

Along with the generative linguistics with which it shares conceptual and historical origins, contemporary theorizing about verbal STM, therefore, posits a fundamental distinction between the processes whereby a set of verbal information may be instantiated sequentially, and the actual elements that make up the content of that sequence. In this way, the typical short-term serial recall task is conceived of a setting in which the rememberer must deal with a novel verbal event by applying an item-independent ordering process to a set of *a priori* items. Here we propose a different way of construing the task setting, and concomitantly, a different way of construing STM. In particular, we propose that the way in which novelty is dealt with in instance- or exemplar-based theories of language may be applied also to the typical STM setting. Such approaches (e.g., Bybee, 2010; Goldberg, 2003; Pierrehumbert, 2003; Tomasello, 2005) account for novel verbal behaviour, broadly, in terms of analogy, rather than a generative approach that involves the application of a general ordering rule or process to a novel set of items. So, the readiness with which novel verbal events may be processed is related not only to the individual items involved but also to the extent to which similar events that may serve as analogies may be retrieved and applied from the participant's previous long-term linguistic experience. Such exemplar-based approaches to language have been shown to provide more ready and parsimonious explanations for an increasingly wide range of linguistic phenomena that are not well accounted for by more traditional structural or generative accounts of language (see e.g., Beckner et al., 2009 for an overview).

On the face of it, it might appear that the STM setting is immune from such an analysis, given the typical approach, discussed above, of deliberately excluding sequences likely to have been encountered previously by the participant. However, the question of prior encounters with a particular sequence (or part thereof) presented to a participant in a STM experiment is not simply a matter of familiarity versus novelty; rather, it will always be a matter of degree. Indeed, the influence of the type of sequence level linguistic characteristics discussed above (e.g., syntactic and phonotactic regularity) could be construed as just such an influence on short-term serial recall, via a process of analogy, of the degree of similarity between the given sequence and the content of the participant's natural linguistic experience (e.g., Goldberg, 2003). Here we provide a detailed analysis of another robust effect in STM – superior recall for sequences of digits over sequences of other types of items – which, we argue, speaks to just this issue. At first glance, such an effect might appear readily amenable to the typical item-oriented account – for example, perhaps digits have some inherent characteristics that make them easier to remember than other items. However, it turns out instead to provide proof of principle for the overlooked influence of long-term associative factors on STM performance, even in those settings in which such factors have been ostensibly eliminated. This analysis has implications, therefore, not only for how digit span performance is to be interpreted, but also for how STM itself is to be theorized; and in particular, since the assessment of STM plays a key role in many accounts of other, higher cognitive functions, for how its role in those functions is to be construed.

Digit span is the standard test of verbal STM performance that is routinely used in psychological studies, either as a stand-alone test or as part of a number of psychological assessment batteries (e.g., Elliot & Smith, 2011; Kaufman & Kaufman, 2004; Wechsler, 2008, 2009). The task involves progressively longer sequences of digits being presented, the goal being to recall them in their correct order until two sequences of a particular length are recalled incorrectly. Span is usually taken to be the number of sequences accurately recalled. Although various other measures of verbal STM capacity exist, the digit span task outnumbers all of them by a factor of at least 16:1.[1]

The widespread use of digit span within psychology has led to the task having major theoretical import across a number of areas. For example: digit span increases with age (e.g., Dempster, 1978; Hale, Bronik, & Fry, 1997; Salthouse, Mitchell, Skovronek, & Babcock, 1989), providing a foundation for theorists to incorporate STM capacity as a mechanism of development (e.g., Baddeley et al., 1998; Halford, 1993; Pascual-Leone, 1970); people who score highly on intelligence tests have larger digit spans than people who score poorly on intelligence tests, suggesting that intelligence is heavily influenced by STM capacity (e.g., Bachelder & Denny, 1977; Hornung, Brunner, Reuter, & Martin, 2011; Kyllonen, 1996); developmental disorders such as dyslexia and specific language impairment involve poor digit span performance, linking atypical development to impairments in STM capacity (e.g., Archibald & Gathercole, 2006; Baddeley et al., 1998; Helland & Asbjørnsen, 2004); and deficits in digit span have largely contributed to hypotheses that suggest involvement of verbal STM across numerous neuropsychological conditions from schizophrenia to Duchenne muscular dystrophy (e.g., Cherry, Buckwalter, & Henderson, 1996; Hinton, De Vivo, Nereo, Goldstein, & Stern, 2001; Mathias & Wheaton, 2007; Stone, Gabrieli, Stebbins, & Sullivan, 1998). As such, the measurement of digit span plays a central role in broad cognitive architectures that posit limited capacity, STM processes as fundamental building blocks for the accomplishment of higher cognitive functions.

However, although digit span is the archetypal test of verbal STM capacity (e.g., Bunting, 2006; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Gathercole & Pickering, 2000; Hansson, Juslin, & Winman, 2008) the fact that span size for digit stimuli is consistently greater than span size for other stimuli such as words or letters is something that has hitherto received little consideration (e.g., Bachelder & Denny, 1977; Crannell & Parrish, 1957; Dempster, 1981; Jacobs, 1887). This may be because, as we suggested above, the effect does not appear on the face of it to constitute a fundamental challenge to the broad conceptualizations of STM that accompany its measurement, but rather may be taken to fall into that class of effects that can be attributed to characteristics of different types of items.

From this perspective, there are a number of possible reasons for digit span superiority. For example, digits are usually sampled from a smaller pool than other stimuli such as letters or words. As a consequence, unknown items at recall are substantially easier to guess when they are digits than when they are not.[2] Another possibility is that, because digits are rarely matched to other stimuli for characteristics such as frequency, syllabic length and phonemic length, any differences may be caused by differences on these parameters. Furthermore, the exclusion from memory lists of well-known, canonical sequences (e.g., *1, 2, 3*; *6, 5, 4*; *2, 4, 6*; etc.) represents the deliberate attempt to eliminate the possibility that the digit span advantage stems from the possibility that sequences of digits may contain a greater number of familiar sequences than lists of other stimuli (e.g., Bachelder & Denny, 1977; Woods et al., 2011). However, there is evidence that the different behaviour of digit sequences within the short-term memory setting may be due to factors that transcend the nature of the individual items themselves. For example, while pre-test exposure to particular item co-occurrence regularities leads to enhanced serial recall when the material comprises individual words or nonwords, no such advantage accrues when the material involves digits (Majerus et al., 2012), possibly reflecting the already high level of exposure to random sequences of digits within the typical adult's linguistic experience. Rather than testing this possibility by within-experiment manipulation of particular transitional regularities, we adopted a corpus-based approach to understand the basis of the digit span superiority effect. If it is indeed due to the fact that a typical participant has had greater prior exposure to digit sequences compared to sequences of other types of material, then the effect should be evident even when the material is matched on critical item-level factors. Furthermore, it should only be evident in recall of sequences, rather than in enhanced recall of individual isolated items, and if the effect is genuinely one that resides in long-term associative processes, as opposed to the type of item- or set-level processes by which linguistic influences on STM are typically explained, then it should also be possible to show that some digit sequences sustain better serial recall than others as a function of their occurrence within the corpus. In the experiments reported here, we seek to evaluate these possibilities, as well as provide a corpus-based simulation to test the viability of our account.

2. Experiment 1: digit span versus word span

Characteristics of the items making up sequences, such as frequency of occurrence, phonological distinctiveness, articulatory duration and complexity all have empirically well-established influences on STM performance (but see Macken, Taylor, & Jones, 2014; Taylor, Macken, & Jones, 2014, for critical evaluation of how such effects are manifest), so in Experiment 1 we sought to replicate the digit span advantage in a setting where digit and non-digit materials were matched on potentially critical dimensions. In order to establish the generality and robustness of the digit span superiority effect within our setting, we compared digit span with span for words on two occasions (Experiments 1a and 1b) each using a different set of words. We also used a closed set of words in both cases in an attempt to address any recall advantage that accrues to such stimulus sets over open sets, and tested serial recall using order reconstruction, to further attempt to minimize any influence on recall of knowledge of the items making up the sequences (see e.g., Jalbert, Neath, Bireta, & Surprenant, 2011; Lovatt, Avons, & Masterson, 2002; Neath, 1997).

2.1. Method

2.1.1. Design

In both Experiments 1a and 1b, the within subjects independent variable was stimulus type (digits or words). The dependent variables were span size (length of longest list accurately recalled) and span total (number of lists accurately recalled).

2.1.2. Participants

84 undergraduate and postgraduate psychology students at Nottingham Trent University (age: *M* = 22.07, *SD* = 7.07, 67 female) participated in Experiment 1a as part of a laboratory class in return for course credit. As with all experiments reported here, participants were treated in accordance with British Psychological Society ethical principles and the research received ethical approval from the Nottingham Trent University Social Sciences ethics committee. A further 25 undergraduate and postgraduate psychology students (age: *M* = 26.32, *SD* = 5.65, 23 female) were recruited on a voluntary basis to participate in Experiment 1b.

---

[1] Based on Google Scholar 04/30/2014 for the following searches within an article (approximate number of hits in parentheses): "digit span" (83,900), "forward span" (1230), "backward span" (1290), "reading span" (5080), "complex span" (2590), "letter span" (1050), "word span" (3550). Note also that forward span and backward span tasks commonly involve digits.

[2] Although Crannell and Parrish (1957) found no effect of sample size, they did not display items at test and therefore relied on participant's memory of the item pools.

2.1.3. Materials

Since the Wechsler scales dominate psychological assessment using digit span, the digit span task from the Wechsler Adult Intelligence Scale, 4th edition (WAIS, Wechsler, 2008) was used. The task has two span lists at each length, from two items to nine items. None of the lists contain well-known sequences (e.g., *1, 2, 3*; *3, 6, 9*). In order to be as similar as possible to the word stimuli, digits were displayed in word form rather than numeral form (e.g., 'one' as opposed to '1'). For Experiment 1a, a word was selected to match each digit for phonemic length and number of syllables. Where possible (eight of nine cases), words also matched digits for number of letters. Frequency of occurrence averaged over one-third greater for words over digits (Kučera & Francis, 1967) to attempt to account for the fact that digits exist in both word form and numeral form ($M = 1052.44$, $SD = 1649.57$ for words; $M = 719.67$, $SD = 1050.66$ for digits). In order to implement this matching process, it was not possible to select word stimuli that were all drawn from the same syntactic category, therefore, arguably, the word sequences formed a more heterogeneous set of verbal material than the digit sequences. In order to address the possibility that such set-level factors might be at play in any differences in span for digit and word sequences and to further test the generality of the digit superiority effect, in Experiment 1b, we therefore selected a set of non-digit stimuli, all of which were nouns that were matched on length with the digits. However, in order to constrain this set of materials to a single syntactic class, we had to relax the frequency-matching constraint. We return to this issue in the analyses of the performance data. Table 1 shows the digits and words used together with their respective frequencies. Word span lists were created by replacing each digit in the digit span lists with its corresponding matched word.

2.1.4. Procedure

This was identical for Experiments 1a and 1b. Tasks were automated using Macromedia Authorware 6.5 and the order of presentation of the type of test was randomized on a participant-by-participant basis. As per the WAIS, testing began at span lists of two items in length and only increased in length when at least one of the two sequences was correctly recalled.

For each list, items were presented individually on the computer screen for a period of .7 s with a .3 s interval between items. After presentation of each span list, a set of clickable buttons appeared at the bottom of the screen (one for every item from which lists were drawn e.g., the written form of the digits 1–9 for the digit span task) in a randomized order and a set of 'XXXX's also appeared to indicate how many items had been presented for the span list. Participants were instructed to click on the buttons in the order in which the original list had been presented.

2.2. Results and discussion

Experiment 1a Table 2 shows the span scores for digits and words. Despite matching digits and words on a number of metrics, using a closed set and order reconstruction in both cases and ensuring that digit lists did not contain well-known sequences, span remained significantly greater for digits than for words for both span size ($t(83) = 7.51$, $p < .001$, Cohen's $d = .96$) and for span total ($t(83) = 7.51$, $p < .001$, Cohen's $d = .86$). Subsequent analyses therefore use span size only.

We also analysed this digit span advantage (digit span minus word span) as a function of span length for participants whose digit span size was between 4 and 7 (extreme span scores were excluded due to very few people [$N = 5$] achieving them). As can be seen in Table 3, this advantage increases as span increases, $F(3,75) = 20.82$, $p < .001$, $\eta_p^2 = .45$.

Table 1
Digit and word frequencies for stimuli used in Experiments 1 and 2.

Digit	Frequency	Word (Expt. 1a)	Frequency	Noun (Expt. 1b)	Frequency
one	3292	had	5133	house	591
two	1412	out	2096	water	442
three	610	might	672	door	312
four	359	high	497	car	274
five	286	need	360	money	265
six	220	best	351	book	193
seven	113	heavy	110	letter	145
eight	104	earth	150	game	123
nine	81	date	103	seed	41

Table 2
Means and 95% confidence intervals for digit and word span scores in Experiments 1a and 1b.

	Task type	
	Digit span	Word span
Span size	Expt. 1a: 5.43 [5.17, 5.69] Expt. 1b: 5.68 [5.15, 6.21]	Expt. 1a: 4.45 [4.28, 4.63] Expt. 1b: 4.96 [4.52, 5.40]
Span total	Expt. 1a: 7.87 [7.40, 8.34] Expt. 1b: 8.94 [7.35, 9.13]	Expt. 1a: 6.15 [5.86, 6.45] Expt. 1b: 6.96 [6.19, 7.73]

Table 3
Means and 95% confidence intervals for span size differences between digits and words, for participants having a digit span size between 4 and 7 inclusive.

	Digit span size			
	4 ($N = 16$)	5 ($N = 22$)	6 ($N = 30$)	7 ($N = 11$)
Digit span size – (minus) word span size	−.13 [−.45, .20]	.86 [.55, 1.18]	1.47 [1.19, 1.74]	1.82 [1.16, 2.48]

Experiment 1b Table 2 shows the span scores for digits and nouns. Using words of the same syntactic class did not eliminate digit span superiority ($t(24) = 2.98$, $p = .007$, Cohen's $d = .61$). However, as noted earlier, digits were of a higher frequency than nouns. In order to attempt to control for this, the proportion of correctly recalled items was calculated for digit lists while excluding recall of the two most frequent digits (*1* and *2*) and for noun lists excluding recall of the two most infrequent nouns (*game* and *seed*). To avoid potential confounds relating to the length of the lists being recalled across stimuli, only recall of stimuli at the same list lengths were considered (e.g., if a participant proceeded to a list length of 7 for nouns but only 5 for digits, list lengths up to 5 were considered). In addition, only the last 4 lists were analysed (i.e., at span) so that the proportion correct was not artificially inflated by accurate performance at short list lengths. Even after excluding the two most frequent digits and the two least frequent nouns, a greater proportion of digits than nouns were accurately recalled ($M = 83.07$, $CI = 76.92–89.21$ for digits, $M = 71.94$, $CI = 66.66–77.22$ for nouns, $t(24) = 3.18$, $p = .004$, Cohen's $d = .80$).

There is clear evidence here, as elsewhere, then, that not only is there a robust STM advantage for digits over words that is not readily accountable for in terms of item type, length, frequency of occurrence or set size, but also that advantage is greater as span length increases. This effect, in principle, does not necessarily reflect a challenge to the typical approach of accounting for differences in STM for different classes of material by reference to the processes supporting the maintenance and retrieval of item-level representations. For example, it has been argued that item redintegration is expected to have increasing effects later than earlier in recalled sequences and so recall of longer sequences will show greater evidence of effects operating via such processes (e.g., Hulme et al., 1997). However, this relationship between span length and the digit span advantage is also congruent with an account of the effect that focuses on the long-term linguistic experience of the participant. Digit sequences are very frequent in the natural linguistic environment and while word sequences occur more often than digit sequences, the syntactic processes that constrain word order in the linguistic environment means that random word sequences are unlikely to occur with any great frequency, whereas digit order is largely unconstrained and random digit sequences therefore occur very often.

We have demonstrated the robustness and generality of the digit superiority effect in Experiments 1a and 1b by showing similar effects when comparing digit span to span for two different sets of words that address a number of potentially critical item- and set-level factors that are often invoked to explain different recall performance with different linguistic materials. The survival of the digit span superiority effect under these conditions provides support, therefore, for the possibility that performance in the typical STM task – such as the span task used here – is a function of the extent to which long-term memory for linguistic events can be brought to bear to support short-term processing of novel verbal sequences. If this is indeed the case, then we would also expect the more frequent encounter with random sequences of digits compared to words to lead to an increasing advantage in the STM setting as sequence length increases. For example, recall of a two-item list A–B can only benefit from such prior experience if a person has previously encountered the sequence A–B a sufficient number of times for it to be retained in long-term memory. However, a three-item list A–B–C can benefit from three potential associative encounters: A–B, B–C, and A–B–C. In Experiment 2, we examined whether the increase in digit superiority at increasing sequence length reported above is more amenable to an account based on maintenance and retrieval processes, such as redintegration, that operate at the item level, or one reflecting the role in STM of prior encounters with particular sequences of verbal material.

3. Experiment 2: recall of lists containing both digits and words

To distinguish genuinely sequence-level effects from those operating on the individual items within sequences, in Experiment 2, span was tested for mixed sequences of digits and words. This was done by removing for each digit sequence either the odd or even numbers and replacing them with their corresponding matched word. This manipulation means that, not only can recall of the individual items be compared across digits and words independently of serial position, recall of individual items can be measured as a function of whether the item occurs as part of a sequence of others of its class or in isolation, neither preceded nor succeeded by another item of the same class.

Table 4 shows how one hypothetical digit span list, *one-four-seven-nine* is changed for the 'even-numbers' (i.e., odd numbers removed and replaced with corresponding word) and 'odd-numbers' (i.e., even numbers removed and replaced with corresponding word) conditions. By having one condition where only even numbers are changed to their corresponding words and another condition where only odd numbers are changed to their corresponding words, critical comparisons can be made across digit and word recall. From Table 4, for example, recall of 'one' in serial position 1 can be directly compared to recall of its matched word 'had' in the same serial position. Similarly, recall of digit sequences and word sequences can be directly compared as they also appear in the same serial positions, as illustrated by 'heavy-date' and 'seven-nine' in Table 4.

Table 4
Example stimuli across conditions of Experiment 2.

Stimuli	Isolated digit	Isolated word	Digit pair	Word pair
one-four-seven-nine	N/A	N/A	N/A	N/A
had-four-heavy-date (even numbers condition)	four	had	N/A	heavy-date
one-high-seven-nine (odd numbers condition)	one	high	seven-nine	N/A

If digit span is greater than word span because of some general characteristic of digits over words, then recall accuracy should be greater for both isolated digits and digit sequences than for isolated words and word sequences. However, if digit span superiority reflects more frequent episodic encounters with random digit sequences than random word sequences, then recall accuracy should be greater for digit sequences than for word sequences, but there should be no difference between recall of isolated digits and isolated words.

3.1. Method

3.1.1. Design

The between subjects independent variable was list-type (even-numbers or odd-numbers) and the within subjects independent variables were sequence-type (isolated or paired) and stimulus-type (digits or words). The dependent variable was the proportion of items correctly recalled.

3.1.2. Participants

68 undergraduate and postgraduate psychology students (age: *M* = 22.13, *SD* = 6.28, 56 female) were recruited and tested in the same way as described for Experiment 1.

3.1.3. Materials and procedure

Span was measured in the same way as described in Experiment 1, with the critical difference that lists were mixed sequences of digits and words, constructed by either replacing each odd digit with its corresponding word (the even-numbers condition) or by replacing each even digit with its corresponding word (the odd-numbers condition). Since we found comparable digit span superiority effects in Experiments 1a and 1b, we used the word set from Experiment 1a here, since they were more precisely controlled and matched with digits on an item-by-item basis. In all other respects, the procedure was the same as Experiment 1.

3.2. Results and discussion

An independent samples *t*-test showed no difference in span between the odd-numbered and even-numbered lists (Odd-numbers *M* = 5.09, *CI* = 4.77–5.40; Even-numbers *M* = 5.12, *CI* = 4.69–5.55; $t(66) = .11$, $p = .911$, Cohen's $d = .03$). Subsequent analyses therefore combine the two list types.

Recall of isolated digits, isolated words, digit pairs and word pairs was analysed for the final four lists that participants were presented with (i.e., those lists that were recalled at span). For the sake of simplicity and consistency, we only analysed sequences in terms of pairs of successive items (e.g. *three-seven* and *seven-one* for the triplet *three-seven-one*), rather than investigating longer sequences, the opportunity for occurrence of which would vary as a function of span length. Table 5 shows the proportion of isolated digits, isolated words, digit pairs, and word pairs that were correctly recalled by participants. A 2 (sequence-type: isolated or paired) × 2 (stimulus-type: digits or words) repeated measures

Table 5
Means and 95% confidence intervals representing the proportion of accurate recall of isolated digits, isolated words, digit pairs, and word pairs for the final four span lists of each participant.

Sequence type	Digits	Words
Isolated item	.68 (.63, .73)	.71 (.66, .76)
Item pair	.57 (.49, .64)	.44 (.37, .51)

ANOVA showed a main effect of sequence-type, ($F(1,67) = 32.39$, $p < .001$, $\eta^2_p = .33$), no main effect of stimulus-type ($F(1,67) = 2.49$, $p = .119$, $\eta^2_p = .04$), but an interaction between the two ($F(1,67) = 6.53$, $p = .013$, $\eta^2_p = .09$). Planned comparisons showed that there was no difference in recall of isolated digits and words ($t(67) = 1.28$, $p = .408$, Cohen's $d = .16$) but significantly more digit pairs were recalled than word pairs ($t(67) = 2.33$, $p = .046$, Cohen's $d = .42$).

Undoubtedly, a large range of linguistic variables – typically attributed to characteristics inherent in the to-be-recalled items either individually or as a set – have been shown to impact STM performance. This means that, in principle, to establish the influence of any given variable, all the other candidate variables need to be controlled. The range of such candidates includes lexicality, frequency, concreteness, phonological complexity, phonological similarity, phonological neighbourhood density, phonotactic regularity, and so on (see e.g., Acheson & MacDonald, 2009; Macken et al., 2014, for discussion). The challenge is clearly considerable, even when approaching the question of the influence of aspects of linguistic experience with, as it were, a blank slate onto which sets of materials may be assembled. The challenge here is even greater, in that the critical set of materials under consideration – the digits – is already given. Therefore, at this juncture we should note that while we have shown the survival of the digit superiority effect when a range of variables is matched, it is still the case that digits form a semantic class where items are drawn from a relatively small pool of closely related items, whereas our chosen words (Experiments 1a and 2) and nouns (Experiment 1b) were drawn from relatively large word pools and exhibited considerably less semantic similarity to that inherent in the set of digits. Such semantic similarity has been shown to have robust effects on STM performance (e.g., Poirier & Saint-Aubin, 1995; Saint-Aubin & Poirier, 1999a, 1999b), and an ideal comparison would be with a set of non-digit materials that shared such semantic similarity while not also occurring as common sequences in the linguistic experience of the rememberer. In practical terms, it seems that such a comparison is unavailable, given the possibility of finding another given semantic class that could also be matched on the other key variables. For example, the frequency of occurrence of primary and secondary colour names in natural language significantly tails off outside of black, white, red, green, and blue.

However it is also worth noting in this respect that aspects of our results deviate from those typically found when comparing STM for sets of semantically related items with unrelated ones. The effect of semantic similarity on serial recall has been typically attributed to benefits of item integrity within such lists, such that, for example, the advantage does not interact with serial position and the difference between semantically related and unrelated lists is due primarily to fewer item errors in the former, with little or no effect on order errors (e.g., Saint-Aubin & Poirier, 1999a, 1999b). Furthermore, even in those cases where semantic similarity has an impact on the pattern of order errors, that impact has been argued to be due to increased overall activation of individual items within the lists as a result of reciprocal activation between the associated items (Acheson, MacDonald, & Postle, 2011; Poirier, Saint-Aubin, Mair, Tehan, & Tolan, 2015). Such a mechanism, if it were responsible for the results of Experiments 1 and 2, would be expected to confer enhanced item activation for digits over words within the mixed lists of Experiment 2, although to a lesser extent than in pure lists. Therefore, a digit superiority effect due to semantic similarity might have been expected to lead to better recall of the isolated digits in Experiment 2, as well as digits recalled in sequence. The evidence here does not support such an interpretation.

So, while our results are supportive of the view that the advantage in STM for digits over words is not due to any inherent characteristic possessed by individual digits as verbal items that is not also possessed by individual words, it is worth remaining cautious about this conclusion at this point, given the challenges (if not the impossibility) of fully controlling for all potentially relevant variables. Our results are, nonetheless, concordant with an effect of linguistic experience that is operating on the ability to reproduce sequences of verbal material, rather than on the efficacy of retrieval of individual items within a list which is the more typically invoked explanation for effects of inter-item association (e.g., Gathercole, Frankish, Pickering, & Peaker, 1999; Jefferies et al., 2009; Stuart & Hulme, 2000). So, these findings point to a potential locus of the effect, not within the classical mechanisms of STM per se, but rather in the way in which – regardless of methodological attempts to eliminate them – factors deriving from the linguistic experience of the rememberer manifest themselves in STM performance. Given the caveats expressed above about the possibility of fully precluding an account of the results thus far in terms of classical STM mechanisms, in Experiments 3 and 4, we investigate our alternative account more directly using the British National Corpus as a proxy for the linguistic experience of our participants.

4. Experiment 3 – computational modelling of associative influences on short-term memory

One of the most potent learning mechanisms that humans possess is the ability to identify and learn associations across stimuli that appear frequently in our perceptual world. For example, young infants can identify word boundaries on the basis of adjacent sounds that rarely co-occur (e.g., Saffran, 2001; Saffran, Aslin, & Newport, 1996); and Reber (1967) and Braine (1963) amongst many others show how children and adults are sensitive to transitional regularities within the environment. There is ample evidence that this type of learning also exists at the lexical level, not least because many of the young child's utterances are rote learned (e.g., Bannard & Matthews, 2008; Fernald, Pinto, Swingley, Weinbergy, & McRoberts, 1998; see also Conway, Bauernschmidt, Huang, & Pisoni, 2010). Since one cannot know *a priori* those word sequences that will eventually become familiar phrases, it is likely the case that children and adults are sensitive to any sequence that appears with some frequency. Once an association is formed, any repetition of it strengthens the trace (Kilb & Naveh-Benjamin, 2011).

One plausible explanation for the digit span advantage seen in Experiments 1 and 2 therefore relates to the associative learning of the random digit sequences that people are exposed to daily, particularly those sequences to which people are exposed on a regular basis (e.g. dates, times, account numbers). Here we provide a computational model of how this process may influence STM. Importantly, it is not the aim here to provide a computational model of STM or of long-term learning, but rather to demonstrate how basic associative processes operating on the natural encounter with language may determine STM performance.

The computational model of long-term sequence learning is based on that of Jones and colleagues, originally named EPAM-VOC (e.g., Jones, Gobet, & Pine, 2007, 2008) but recently given the more meaningful acronym CLASSIC (Chunking Lexical and Sublexical Sequences in Children; Jones, Gobet, Freudenthal, Watson, & Pine, 2014). CLASSIC is simplified here in order to emphasise the role of long-term instance-based, associative knowledge. From an input utterance or sentence, the model initially learns the individual words; however, when adjacent words are both already known to the model, it learns new information that corresponds to the word sequence. For the purposes of this analysis, we will call any word or sequence of words that is learnt by the model a 'chunk', because the information learned corresponds well with the concept of chunks of information outlined in classical STM research (e.g., Cowan, 2001; Gobet et al., 2001; Miller, 1956) as well as in contemporary, instance-based theories of language (e.g., Bybee, 2010).

4.1. Input to the model

Input to the model is from the British National Corpus (BNC, http://www.natcorp.ox.ac.uk/). The BNC provides a snapshot of spoken (20%) and written (80%) contemporary UK-based English language use drawn from conversations, popular fiction, national newspapers, etc. In total, the BNC contains over 5 million spoken or written utterances/sentences and over 100 million words. A random sample of half a million spoken utterances and written sentences is presented to the model one utterance/sentence at a time.

4.2. Chunk learning

Any given input is first recoded into chunks based on the model's existing knowledge. For example, the input *where's the hat* will be recoded as two chunks *where's the* and *hat* if the model already has knowledge of the chunked phrase *where's the* (but not *where's the hat*) and has already learned *hat* as a chunk.

Chunk learning occurs by a simple process of learning a new chunk for each of the adjacent chunks in the recoded input (or learning as a chunk any new word that has not yet been learnt as a chunk). For example, when first encountering *the big brown cat* no recoding is possible and each word is learnt as a chunk. On next encountering the same input, each word is recoded as a chunk and new chunks are learnt corresponding to the sequences *the big*, *big brown*, and *brown cat*. On third presentation, the model can now use its existing chunked knowledge to recode the input as two chunks (*the big*, *brown cat*) rather than four individual chunks. Learning joins these two chunks together to create a new chunk that contains the whole phrase.

Rather than learning chunked sequences instantaneously, a learning rate of 0.5 is used to reduce the possibility of 'one-off' sequences being learned (i.e., on average, a sequence must exist at least twice in the input to be learnt as a chunk). Although this may seem unrealistically high, the input to the model is miniscule compared to the language that humans are exposed to during childhood and adulthood. For example, even pre-literate infants may be exposed to as many as half a million words in just a 3 week period (Swingley, 2007).

4.3. Limiting the processing of a given input

There is extensive literature on limitations involving the amount of information that can be processed at any one time, with the question of precisely what the nature and dimensions of such capacity limits might be remaining controversial (e.g., Cowan, Ricker, Clark, Hinrichs, & Glass, 2014; Luck & Vogel, 2013; Macken, Taylor, & Jones, 2015). Our aim here is simply to provide some constraint on the amount of information that can be processed at any one time, so for reasons of operational expedience we implemented this limit as 4.5 chunks. Importantly, as we report below, this particular parameter value is not critical for the argument here, and so the choice of this limit is not intended to speak specifically to the question of what the capacity of STM might be, or even how it should be construed (e.g., Cowan, Morey, Chen, Gilchrist, & Saults, 2008; Macken et al., 2015). A sigmoidal function was applied such that accessing a chunk in STM was probabilistic. Fig. 1 shows that these probabilities favour access of chunks that appear at the end of the input. Using Fig. 1 as a guide, when an input is recoded as 9 chunks, an average of 4.5 are accessed; this number decreases marginally for fewer than 9 chunks and increases marginally for more than 9 chunks. This is broadly consistent with serial recall of independent items when list length varies (see Grenfell-Essam & Ward, 2012). Note that learning any chunked sequence relies on adjacent chunks both being accessed in STM via the sigmoidal function. Table 6 gives examples of chunk learning when chunk access is constrained in this way.

4.4. Performing span tests in the model

Span tests are presented to the model in exactly the same way as any language-based input: as a simple word list. Chunks in the model are then used to recode the input as much as possible, with correct span production occurring only when every chunk in the recoded input can be accessed (based on the sigmoidal probabilities shown above). Chunk access is probabilistic, hence each span test is presented to the model 50 times with average accuracy being recorded.

4.5. Results of the model

Experiments 1 and 2 above examined digit span, word span, and mixed span. For the purposes of the model, all three span tests were presented as input. Span lists were administered at list lengths from 2 to 9. The goal of the modelling work is to illustrate that a greater number of digit sequences are learned as long-term associative knowledge than word sequences, causing digit span to be significantly larger than word span.

Mean span size for each type of stimulus was as follows: 5.38 (CIs 5.19, 5.57) for digits (compared to 5.43 for adults), 4.86 (4.65, 5.07) for words (compared to 4.45 for adults), and 5.13 (4.94, 5.32) when lists were a mixture of digits and words (compared to 5.11 for adults). Fig. 2 shows how many chunks were used to recode lists of each stimulus type and length. If sequence knowledge plays no role in list recall, then the number of chunks that are required to recode each list should be equal to the number of items in the list. Fig. 2 clearly shows that this is not the case, since all stimulus types benefit from chunked sequences (for example, at list length 4 an average of 3.5 chunks are required to recode lists of all stimulus types). Fig. 2 also shows that while all stimulus types benefit from chunked sequence knowledge, in general, digit lists require fewer chunks than mixed lists, and mixed lists require fewer chunks than word lists.

The most striking feature of Fig. 2 is at list length 5 where fewer chunks are required to recode digit lists than mixed lists, and fewer chunks are required to recode mixed lists than word lists. Furthermore, these differences are flanked by equality in recoded chunks across stimulus types for list lengths 4 and 6. This is potentially interesting because the majority of participants in Experiments 1 and 2 achieved a span size of either 4, 5, or 6. If linguistic experience plays a role in short-term serial recall performance therefore, differences in span scores should occur at list length 5. This is precisely where participants falter for word lists in particular. 60% of participants achieve a maximum word span of 4 items compared to 32% for mixed lists and 26% for digit lists.

Fig. 2 also shows a clear hierarchy in the number of chunks required to recode lists of 7 or more items, with digit lists requiring

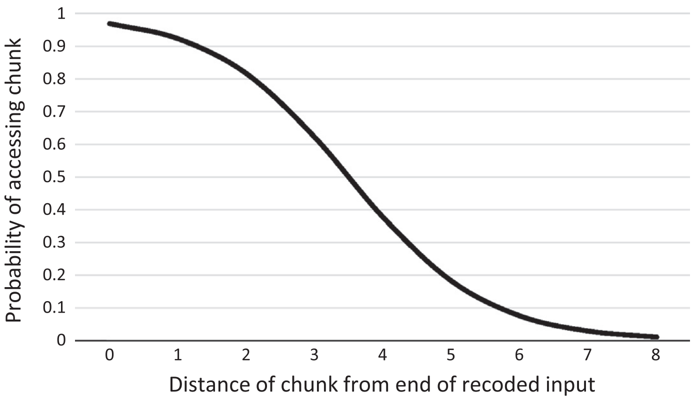

Fig. 1. Sigmoidal probabilities for chunk access. The chunk having a distance of 0 is at the end of the recoded input; a distance of 1 is the next-but-last chunk, and so on.

Table 6
Examples of chunk learning based on how an input is recoded into chunks and whether those chunks can be accessed in STM (chunks in italics are not accessed).

Input	Recoded input	Chunks learned
Where's the doggy gone?	*Where's*, the, doggy, gone	The doggy, doggy gone
How did the bouncy ball get there?	*How*, did the, *bouncy ball*, get, there	Get there
The pink one goes underneath that blue one	The pink one, goes underneath, *that*, blue, one	The pink one goes underneath, blue one

the least number of chunks and word lists requiring the greatest number of chunks. Again, this is supported by the adult data where 15% of participants achieve a digit span of 7 or more compared to 6% for mixed span lists and 1% for word span lists.

The results from the model show that it is possible to broadly simulate participant performance across different span lists on the basis of an associative learning mechanism combined with a constraint on the extent to which new chunks can be formed and accessed for a given sequential input. However, to make clear that greater span size for digits over words is not an artefact of the particular parameter values implemented in the model, we carried out span tests at different chunk capacities. Table 7 shows that regardless of chunk capacity, span size for digit lists is consistently larger than span size for word lists. Performance in the model is consistent with the hypothesis that a greater amount of instance-based, associative learning occurs for random digit sequences than for random word sequences, and this accounts for why span size for random sequences of digits is greater than span size for random sequences of words. We provide a final empirical test of this proposal in Experiment 4.

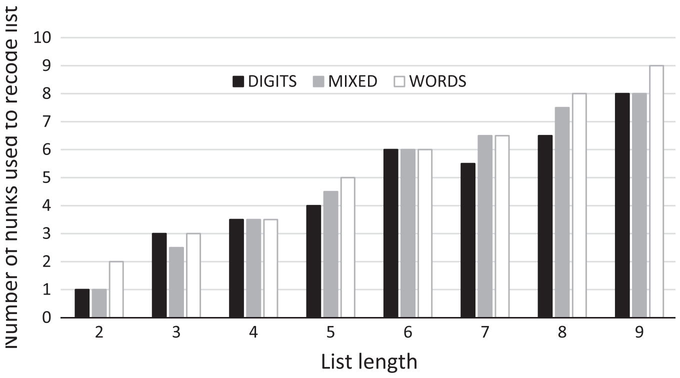

Fig. 2. Number of chunks required to recode span lists, by stimulus type and list length.

Table 7
Digit span and word span across different chunk capacities.

Chunk capacity	Digit span	Word span
3	3.74 (3.52, 3.96)	3.39 (3.17, 3.61)
4	4.77 (4.59, 4.96)	4.39 (4.20, 4.58)
5	5.87 (5.63, 6.10)	5.54 (5.35, 5.73)
6	7.31 (7.07, 7.54)	6.33 (6.12, 6.54)

5. Experiment 4

We now arrive at a point where both the results of empirical data and the computational modelling environment support an associative learning account of the digit span superiority effect based on the number of encounters with random sequences of digits within the experience of the rememberer. If this is the case, then a natural next step is to interrogate the model so that we can construct digit lists containing pseudo-random digit sequences that occur frequently versus those that rarely occur. In this way, any difference in performance cannot be because digits hold some special characteristic that other stimuli such as words do not; rather any difference in performance would arise from the frequency with which the sequences occur in natural language (under the assumption that associative learning is more likely to occur for frequent than infrequent sequences).

However, for reasons of tractability, the computational model was only trained on one-tenth (half a million utterances/sentences) of the BNC. Given that we have already concluded that even the full set of BNC data is a fraction of that encountered by children and adults, rather than interrogate the model we extracted sequence frequencies from the BNC in full. Experiment 4 therefore compares recall performance for random digit sequences that occur frequently in the BNC with those that occur less frequently, the prediction being that the former should facilitate serial recall more than the latter.[3]

5.1. Method

5.1.1. Design

The within subjects independent variable was sequence type (high frequency or low frequency). The dependent variables were

---

[3] Since repeated random sequences of words other than digits are so rare in the corpus, we are not in a position to make this manipulation with such stimuli, although in principle, the logic would be the same for any type of verbal material.

the number of lists accurately recalled and the proportion of correctly recalled items within each list.

5.1.2. Participants

24 undergraduate and postgraduate psychology students (*M* = 22.21, *SD* = 3.49, 19 female) were paid for their participation.

5.1.3. Materials

Table 8 shows the mean frequency with which digit pairs that begin with the digits *two* through *nine* appear in the BNC. The digit '1' (one) was omitted because it is also used to refer to oneself and so some of its occurrence in language represents a different semantic class to that of the other digits. The table illustrates that *two* occurs most frequently in digit sequences but other digits (in particular *nine*) rarely occur. Eight digit span lists were therefore constructed to each contain 7 digits: 4 digit span lists averaged a relatively high pair frequency (*M* = 25.63, *CI* = 18.46–32.79) by using the digits 2–8; 4 digit span lists averaged a relatively low sequence frequency (*M* = 12.63, *CI* = 9.92–15.33) by using the digits 3–9. Canonical sequences such as *7, 8, 9* and *2, 4, 6* were excluded.

5.1.4. Procedure

The procedure differed slightly from the previous experiments in that participants were presented with 8 digit lists each containing 7 digits, rather than digit lists that gradually increased in size. All other experimental details were identical.

5.2. Results and discussion

Digit lists containing relatively high frequency digit pairs were recalled significantly more accurately (*M* = 1.88, *CI* = 1.30–2.45) than digit lists containing relatively low frequency digit pairs (*M* = 1.33, *CI* = 0.86–1.81), *t*(23) = 2.33, *p* = .029, Cohen's *d* = .44. Furthermore, this effect is not due to the presence of the higher frequency digit *2* in one set of lists relative to the lower frequency digit *9* in the other lists. The proportion of digits that were correctly recalled – excluding the digits *2* and *9* – was greater for the high frequency digit lists than for the low frequency digit lists (*M* = .75, *CI* = .67–.83 for high frequency lists, *M* = .65, *CI* = .56–.74 for low frequency lists, *t*(23) = 2.56, *p* = .018, Cohen's *d* = .50).

Experiments 1 and 2 showed not only that the previously observed digit span superiority occurs under conditions where the effect is unlikely to be due to the typical item- or set-level mechanisms used to account for differences in recall of different types of verbal material (Experiments 1a and 1b), but also that the advantage is only evident when recall of sequences of items is assessed, and does not hold for recall of individual items themselves (Experiment 2). To this pattern, the results of Experiments 3 and 4 add the critical detail that not all random digit sequences tested in STM experiments are created equal in this sense, but that a finer grained analysis shows that those sequences that have likely been encountered more frequently than others by the rememberer support better serial recall than sequences of digits that are more rarely encountered.

Table 8

Means and 95% confidence intervals for the raw BNC frequency of digit pairs beginning with the digits two through nine, excluding same item pairs and pairs involving 'one'. For example, the entry for 'three' is the mean raw frequency of occurrence of the sequences 'three two', 'three four', 'three five'… through to 'three nine'.

Digit	Mean frequency of digit pairs
Two	73.45 (.57, 146.29)
Three	38.43 (4.41, 72.45)
Four	26.29 (9.25, 43.32)
Five	15.86 (4.02, 27.69)
Six	13.71 (7.54, 19.89)
Seven	11.86 (4.83, 18.89)
Eight	13.00 (10.55, 15.45)
Nine	4.29 (1.63, 6.94)

6. General discussion

Experiment 1 showed that digit span was superior to word span even when digit lists did not contain well-known sequences, when digit and word stimuli were matched for phonemic and syllabic length, when the size of the pool from which items were sampled was equal, when words were drawn from either different or the same syntactic class, when frequency of occurrence was matched between digits and words, and when the burden on item memory was reduced by utilizing a serial order reconstruction test. Experiment 2 used lists that were a mixture of digits and words to show that isolated digits and isolated words were recalled equally well but digit sequences were recalled more accurately than word sequences. The British National Corpus was then used in Experiment 3 to show that the superior recall of digit sequences over word sequences arose because random sequences of digits occurred more frequently than random sequences of words in natural language. Furthermore, the model demonstrated that basic associative mechanisms operating on the linguistic experience represented by the corpus could account for the digit span advantage. Experiment 4 then showed that pseudo-random digit lists containing digit sequences that occur relatively frequently in the language environment were recalled more accurately than pseudo-random digit lists containing digit sequences that occur less frequently. These results have specific implications for studies that have used digit span; more generally though, they also have wider implications concerning how long-term influences on STM are conceived.

Our findings suggest that the archetypal measure of verbal STM capacity – that of digit span – is in part a function of the structure of the natural linguistic environment (e.g., the relative frequency of dates, times, telephone numbers, compared to other types of verbal strings). Indeed, if one begins to search for seemingly random digit sequences, their prevalence in the environment becomes startlingly obvious. Using the UK as an example, the primary national sport (soccer) consistently lists results from matches as digit sequences (e.g., *3–1, 2–4, 5–3*); company phone numbers (i.e., random sequences of digits) are often consistently repeated on national television and radio advertisements as an aide memoire; and bank account numbers and sort codes, used by almost every adult, are series of random digits that are frequently used and encountered. This provides an explanation for performance differences that are seen across different stimulus sets. For example, Dempster (1978) shows span size for randomized lists of digits is greater than span size for randomized lists containing consonants and vowels, which in turn is greater than span size for consonant-only sequences, all of which is predicted by the frequency with which people encounter random sequences of the particular stimuli (see also Crannell & Parrish, 1957; Jacobs, 1887).

Digit span is most often administered as part of an assessment battery. The experiments presented here used the digit span test from the Wechsler intelligence scales (e.g., Wechsler, 2008, 2009), though other standardized scales include similar digit span tests (e.g., Elliot & Smith, 2011; Kaufman & Kaufman, 2004). These tests generally assume that (forward) digit span measures short-term verbal memory capacity as a basic cognitive mechanism. This is somewhat troubling because studies that include digit span – of which there are thousands – are in effect exaggerating the role of STM capacity on the basis of digit span results, and playing down the role of long-term instance-based associative learning (or lack thereof). For example, links between dyslexia and poor digit span performance, often taken to implicate verbal STM processes in reading deficits (e.g., Helland & Asbjørnsen, 2004; Paulesu et al., 2001), could reflect a deficit in instance-based associative memory, rather than STM, per se. Support for this possibility comes from computational modelling work (Jones et al., 2008, 2014) that shows how long-term associative learning of phonological knowledge can explain performance differences for another verbal memory task that children with dyslexia show deficits for – that of nonword repetition. Moreover, recent research involving children and adults with dyslexia is starting to discover difficulties in associative learning (Du & Kelly, 2013; Hedenius et al., 2013).

Similarly, developmental increases have consistently been shown for digit span (e.g., Chi, 1977; Karakaş, Yalın, Irak, & Erzengin, 2002) and have often been assumed to be due to increases in verbal STM capacity. The current experiments illustrate how increases in digit span could plausibly arise from the gradual instance-based learning of associations across digit sequences rather than from changes in verbal STM capacity per se. This is also supported by empirical work showing how concentrated learning of digit sequences significantly improves span for digits only (Ericsson, Chase, & Faloon, 1980). Under this explanation, developmental increases in tasks such as digit span are dictated more by the extent of one's experience with sequences of verbal material, rather than increases in the capacity of basic STM processes (see also French & O'Brien, 2008; Ottem, Lian, & Karlsen, 2007; though also see Cowan et al., 2014). That such linguistic experience, rather than an increase in the capacity of some distinct STM system, may account for the developmental increase in STM performance is also supported by age-related increases in digit span being much greater than for either letter or word span (Dempster, 1981). As evidenced in the British National Corpus, increasing linguistic experience leads to more rapid accumulation of encounters with random digit sequences than random sequences of other types of verbal material. An instance-based, associative account of performance in the STM setting therefore predicts that span for random sequences of digits will show greater age-related increases than span for random sequences of other stimuli. It seems likely that such experience operates both on perceptual and productive aspects of episodic experience with language. So, perceptual exposure to sequences of verbal material leads to the implicit learning of the sequential probabilities within that material (e.g., Saffran et al., 1996). Episodic experience involving the production of sequences of verbal material also leads to increased STM for that material by specific enhancements in the fluency with which extended sequences of that material may be articulated (Woodward et al., 2008). At the other end of the developmental spectrum, this richer episodic perceptual and productive repertoire for digit sequences may also explain the much slower decline in digit span with aging compared to that found with other measures of STM (Bopp & Verhaeghen, 2005; Hester, Kinsella, & Ong, 2004; Salthouse & Babcock, 1991).

Our argument here also begins to question the relationship between STM and intelligence. Strong links between digit span performance and intelligence have often been used to argue for a key role for STM capacity in constraining intelligence (e.g., Bachelder & Denny, 1977; Hornung et al., 2011; Kyllonen, 1996). However, the current findings point to a role for ability to learn sequential information and apply it in novel settings, as opposed to STM capacity per se, in underpinning performance on intelligence tests. This suggestion is supported by other studies where people who score highly on intelligence tests more readily learn sequences or associations than people who score poorly on intelligence tests (Feldman, Kerr, & Streissguth, 1995; Kaufman, DeYoung, Gray, Brown, & Mackintosh, 2009). Furthermore, one of the most fundamental markers of human intelligence – language learning – depends critically on the ability to learn sequential information and individual differences in both language comprehension and verbal short-term memory are predicted by individual differences in learning sequential transitional probabilities (Misyak & Christiansen, 2012). Unsurprisingly, from this perspective, studies have shown that digit span performance strongly relates to language learning (Gathercole, Hitch, Service, & Martin, 1997; Payne & Holzman, 1983).

There are also more general theoretical and methodological issues to which our results speak. Critically, as discussed in the introduction, the methodology that is used to examine STM explicitly seeks to prevent a systematic influence of long-term experience with sequences of verbal material within the STM setting by excluding (what the experimenter judges to be) familiar sequences of stimuli such as letters or digits. However, the current results cast doubt on the efficacy of such precautions: the digit span advantage derives from the increased frequency of occurrence of digit sequences (as opposed to digits themselves) in the linguistic experience of the rememberer; those sequences of digits that are more frequent than others (even excluding canonical runs, etc.) sustain superior short-term serial recall than less common digit sequences; and, as indicated by the modelling of the BNC in Experiment 3, instance-driven sequential associations are also present for other stimuli such as the random lists of words that were used in the current experiments.

A view that performance in a STM setting is influenced by aspects of the participant's experience outside that setting is neither novel nor, in itself, controversial. Certainly current theoretical approaches account for such influence, broadly speaking, by either suggesting that STM is influenced by long-term knowledge via supplementary mechanisms that impact on the individual item (e.g., redintegration, Hulme et al., 1997); or for those views that construe STM as that portion of long-term memory which is currently activated, long-term knowledge influences the activation of individual items by virtue of their location within lexico-phonological semantic networks (e.g., Jefferies et al., 2009). By showing that STM is influenced at the level of the sequence as well as the individual item, our results have significant implications for how the relationship between STM and long-term memory is best conceived, as well as the more general question regarding what aspects of cognitive functioning are actually being measured in a typical STM setting.

It might be argued that these implications are relatively inconsequential: although digit span is often measured using an established test, STM recall for other stimuli is typically measured over many trials involving many different random permutations of list items, and so any advantage for particular sequences due to prior encounters may be 'washed out' over trials. However, when performance is being compared across different classes of verbal material (e.g., abstract versus concrete nouns, low versus high frequency words, words with dense versus sparse phonological neighbourhoods, different grammatical categories), one cannot be certain that the different classes appear sequentially with equal frequency within the language of the rememberer. Without such an analysis, explanations of the effects of different types of verbal item on STM performance that invoke inherent aspects of the items themselves – as is the habit of theories of STM – can be called into question.

In broad terms, then, our proposal is that the ostensibly novel sequences presented in the STM setting can be thought of as varying on a continuum of similarity to the remembers' linguistic experience, and the closer they correspond to that experience, the more readily that experience may be brought to bear in the novel setting and the better performance will be (see e.g., Ericsson & Kintsch, 1995 for a similar view of STM). At one end of this continuum we could locate the general skill of perceptual–motor mapping whereby the perception of verbal material may be utilized to control the articulatory apparatus, such that any sequence of verbal material may be converted into an articulatory control programme that may subsume subvocal rehearsal and subsequent reproduction of the sequence (e.g., Macken et al., 2014). Language-specific aspects of experience will mean that particular verbal forms are more readily mapped in such a way, such that, for example, particular phonological forms will sustain better STM than others (see e.g., Vihman, 2014 for discussion of how such perceptual–motor interactions are involved in the infant's acquisition of phonological forms). The continuum extends into the type of setting focused upon here, where experience with specific instances of extended sequences of verbal material may lead to integrated perceptual and/or articulatory representations ('chunks') that correspond directly to the to-be-remembered sequence, or part thereof. In this way, the limits to STM can be thought of as limits in the processes that accumulate task-relevant experience (perceptual–motor mapping, chunking, statistical learning, etc.) and limits in the correspondence between that experience and the material presented in any given STM setting (Macken et al., 2015). Indeed, modelling work has shown that performance in a verbal STM task (nonword repetition) can match that of children of different ages purely from greater experience of phonological sequences over time based on linguistic exposure, with STM merely being a mechanism by which associative learning is constrained (Jones et al., 2007).

Such a perspective differs from the typical view within cognitive science in which STM processes are construed as basic, primitive cognitive mechanisms whose functioning underpins and constrains higher functions, including their development. This notion has been challenged elsewhere by showing that many of the canonical empirical hallmarks of the operation of such putative short-term verbal memory processes can in fact be fully accounted for by reference to domain-general perceptual and motor control processes that are opportunistically co-opted within the STM setting in order to accomplish the particular task (e.g., Jones, Hughes, & Macken, 2006; Jones, Macken, & Nicholls, 2004; Macken et al., 2014; Maidment & Macken, 2012; Maidment, Macken, & Jones, 2013). From this perspective, the STM setting is one in which participants are presented with sequences of verbal material from which typical transitional probabilities (e.g., those derived from meaning or syntax) have been removed, and the participants' task is to reproduce that sequence. Performance in this setting, rather than being a reflection of the operation of bespoke mechanisms whose function is the temporary maintenance and manipulation of such material, is a function of the extent to which the participants' perceptual and motor abilities readily afford accomplishment of the task with those materials. For verbal material, the key aspect of the participants' skill and knowledge that can be brought to bear is that which derives from their linguistic experience. Indeed, it is already well-established that the closer the material corresponds to the linguistic knowledge of the participant (e.g., with respect to lexical, semantic, syntactic and phonotactic characteristics; see for example Macken et al., 2014 for a discussion), then the better short-term sequence memory is for that material. The possibility highlighted here is that, rather than this advantage occurring due to enhanced processing of the items that constitute the tested material, it is due instead to the correspondence between sequence-level properties of the to-be-remembered material and the sequential characteristics of the participant's linguistic skill and experience.

The type of account we are proposing here accords with this perspective, and provides a parsimonious broad account of the range of linguistic influences – be they at the level of the item or the sequence – on STM performance: The setting is one in which participants must deal with a novel verbal event, and language users do so by drawing on their linguistic repertoire such that the readiness with which the novel event may be processed is a function of the extent to which appropriate analogies are available within that repertoire that can support processing of a given sequence. Such instance-based accounts of language use generally have received increasing wide empirical support in recent decades (see e.g., Bybee, 2010; Pierrehumbert, 2003; Tomasello, 2005; Vihman, 2014) and what we see here is further evidence that performance in verbal STM tasks is better thought of as reflecting the application of linguistic knowledge, rather than the operation of STM systems, per se (see also Acheson & MacDonald, 2009; MacDonald & Christiansen, 2002; Martin & Saffran, 1997). From this perspective, the relationship between STM and long-term memory is merely that STM is a particular task setting in which participants apply their long-term associative knowledge of language in an opportunistic and task-specific way in order to accomplish the particular requirements in that setting (Macken & Jones, 2003).

Minimally, what we have shown here is that the usual methodological expedient of attempting to exclude familiar sequences from the material presented to participants – the raison d'être being to enable more or less pure examination of ostensible STM processes – cannot be assumed to do so successfully without a detailed analysis of the frequency not only of particular types of 'item' but also of particular types of sequence within the linguistic repertoire of the participant. This is particularly the case for digit sequences which appear more readily in the environment than first thought, questioning the implications of studies that draw conclusions based on digit span performance. More broadly, however, we also argue that evidence such as this lends weight to a radical reconceptualization of what is actually being assessed in STM research, away from the idea that we are assessing distinct, primitive cognitive mechanisms for the temporary maintenance and manipulation of information towards a view that sees it rather as being parasitic on broader, domain-general processes involved with long-term learning of sequential information.

References

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin, 135*, 50–68.

Acheson, D. J., MacDonald, M. C., & Postle, B. R. (2011). The effect of concurrent semantic activation on delayed serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 44–59.

Archibald, L. M. D., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders, 41*, 675–693.

Bachelder, B. L., & Denny, M. R. (1977). A theory of intelligence: I. Span and the complexity of stimulus control. *Intelligence, 1*, 127–150.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science, 19*, 241–248.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., et al. (2009). Language is a complex adaptive system: Position paper. *Language Learning, 59*, 1–26.

Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 60*, 223–233.

Botvinick, M. M., & Bylsma, L. M. (2005). Regularization in short-term memory for serial order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 351–358.

Braine, M. D. (1963). On learning the grammatical order of words. *Psychological Review, 70*, 323–348.

Bunting, M. (2006). Proactive interference and item similarity in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 183–196.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*, 551–581.

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55*, 627–652.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.

Cherry, B. J., Buckwalter, J. G., & Henderson, V. W. (1996). Memory span procedures in Alzheimer's disease. *Neuropsychology, 10*, 286–293.

Chi, M. T. (1977). Age differences in memory span. *Journal of Experimental Child Psychology, 23*, 266–281.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition, 114*, 356–371.

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163–183.

Cowan, N. (1995). *Attention and memory: An integrated framework*. Oxford: Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 97–185.

Cowan, N., Morey, C. C., Chen, Z., Gilchrist, A. L., & Saults, J. S. (2008). Theory and measurement of working memory capacity limits. *Psychology of Learning and Motivation, 49*, 49–104.

Cowan, N., Ricker, T. J., Clark, K. M., Hinrichs, G. A., & Glass, B. A. (2014). Knowledge cannot explain the developmental growth of working memory capacity. *Developmental Science*.

Crannell, C. W., & Parrish, J. M. (1957). A comparison of immediate memory span for digits, letters, and words. *The Journal of Psychology, 44*, 319–327.

Dempster, F. N. (1978). Memory span and short-term memory capacity: A developmental study. *Journal of Experimental Child Psychology, 26*, 419–431.

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89*, 63–100.

Du, W., & Kelly, S. W. (2013). Implicit sequence learning in dyslexia: A within-sequence comparison of first-and higher-order information. *Annals of Dyslexia, 63*, 154–170.

Elliot, C. D., & Smith, P. (2011). *British abilities scales* (3rd ed.). London, UK: GL Assessment.

Ericsson, K. A., Chase, W. G., & Faloon, S. (1980). Acquisition of a memory skill. *Science, 208*, 1181–1182.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211–245.

Feldman, J., Kerr, B., & Streissguth, A. P. (1995). Correlational analyses of procedural and declarative learning performance. *Intelligence, 20*, 87–114.

Fernald, A., Pinto, J. P., Swingley, D., Weinbergy, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science, 9*, 228–231.

French, L. M., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics, 29*, 463–487.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 84–95.

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33*, 966–979.

Gathercole, S. E., & Pickering, S. J. (2000). Assessment of working memory in six-and seven-year-old children. *Journal of Educational Psychology, 92*, 377–390.

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., et al. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5*, 236–243.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences, 7*, 219–224.

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

Hale, S., Bronik, M. D., & Fry, A. F. (1997). Verbal and spatial working memory in school-age children: Developmental differences in susceptibility to interference. *Developmental Psychology, 33*, 364–371.

Halford, G. (1993). *Children's understanding: The development of mental models*. London, UK: Routledge.

Hansson, P., Juslin, P., & Winman, A. (2008). The role of short-term memory capacity and task experience for overconfidence in judgment under uncertainty. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1027–1042.

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning* (pp. 37–46). London, UK: Oxford University Press.

Hedenius, M., Persson, J., Alm, P. A., Ullman, M. T., Howard, J. H., Jr., Howard, D. V., et al. (2013). Impaired implicit sequence learning in children with developmental dyslexia. *Research in Developmental Disabilities, 34*, 3924–3935.

Helland, T., & Asbjørnsen, A. (2004). Digit span in dyslexia: Variations according to language comprehension and mathematics skills. *Journal of Clinical and Experimental Neuropsychology, 26*, 31–42.

Henson, R. N. A. (1998). Short-term memory for serial order: The Start-End Model. *Cognitive Psychology, 36*, 73–137.

Hester, R. L., Kinsella, G. J., & Ong, B. (2004). Effect of age on forward and backward span tasks. *Journal of the International Neuropsychological Society, 10*, 475–481.

Hinton, V. J., De Vivo, D. C., Nereo, N. E., Goldstein, E., & Stern, Y. (2001). Selective deficits in verbal working memory associated with a known genetic etiology: The neuropsychological profile of Duchenne muscular dystrophy. *Journal of the International Neuropsychological Society, 7*, 45–54.

Hornung, C., Brunner, M., Reuter, R. A., & Martin, R. (2011). Children's working memory: Its structure and relationship to fluid intelligence. *Intelligence, 39*, 210–221.

Hulme, C., Maughan, S., & Brown, G. D. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*, 685–701.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, M., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1217–1232.

Jacobs, J. (1887). Experiments on "prehension". *Mind, 12*, 75–79.

Jalbert, A., Neath, I., Bireta, T. J., & Surprenant, A. M. (2011). When does length cause the word length effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 338–353.

Jefferies, E., Frankish, C., & Noble, K. (2009). Lexical coherence in short-term memory: Strategic reconstruction or "semantic" glue"? *Quarterly Journal of Experimental Psychology, 62*, 1967–1982.

Jones, G., Gobet, F., Freudenthal, D., Watson, S. E., & Pine, J. M. (2014). Why computational models are better than verbal theories: The case of nonword repetition. *Developmental Science, 17*, 298–310.

Jones, G., Gobet, F., & Pine, J. M. (2007). Linking working memory and long-term memory: A computational model of the learning of new words. *Developmental Science, 10*, 853–873.

Jones, G., Gobet, F., & Pine, J. M. (2008). Computer simulations of developmental change: The contributions of working memory capacity and long-term knowledge. *Cognitive Science, 32*, 1148–1176.

Jones, D. M., Hughes, R. W., & Macken, W. J. (2006). Perceptual organization masquerading as phonological storage: Further support for a perceptual-gestural view of short-term memory. *Journal of Memory and Language, 54*, 265–281.

Jones, D. M., Macken, W. J., & Nicholls, A. P. (2004). The phonological store of working memory: Is it phonological and is it a store? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 656–674.

Karakaş, S., Yalın, A., Irak, M., & Erzengin, Ö. U. (2002). Digit span changes from puberty to old age under different levels of education. *Developmental Neuropsychology, 22*, 423–453.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Brown, J., & Mackintosh, N. (2009). Associative learning predicts intelligence above and beyond working memory and processing speed. *Intelligence, 37*, 374–382.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman assessment battery for children* (2nd ed.). Circle Pines, MN: American Guidance Service.

Kilb, A., & Naveh-Benjamin, M. (2011). The effects of pure pair repetition on younger and older adults' associative memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 706–719.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Dartmouth Publishing Group.

Kyllonen, P. C. (1996). Is working memory capacity Spearman's g? In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 49–75). Mahwah, NJ: Lawrence Erlbaum Associates.

Lovatt, P., Avons, S. E., & Masterson, J. (2002). Output decay in immediate serial recall: Speech time revisited. *Journal of Memory and Language, 46*, 227–243.

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17*, 391–400.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*, 35–54.

Macken, W. J., & Jones, D. M. (2003). The reification of phonological storage. *Quarterly Journal of Experimental Psychology, 56A*, 1279–1288.

Macken, B., Taylor, J., & Jones, D. M. (2014). Language and short-term memory: The role of perceptual-motor affordance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 1257–1270.

Macken, B., Taylor, J., & Jones, D. (2015). Limitless capacity: A dynamic object-oriented approach to short-term memory. *Frontiers in Psychology, 6*, 293.

Maidment, D. W., & Macken, W. J. (2012). The ineluctable modality of the audible: Perceptual determinants of auditory-verbal short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 989–997.

Maidment, D. W., Macken, B., & Jones, D. M. (2013). Modalities of memory: Is reading lips like hearing voices? *Cognition, 129*, 471–493.

Majerus, S., Martinez Perez, T., & Oberauer, K. (2012). Two distinct origins of long-term learning effects in verbal short-term memory. *Journal of Memory and Language, 66*, 38–51.

Majerus, S., van der Linden, M., Mulder, L., Meulemans, T., & Peters, F. (2004). Verbal short-term memory reflects the sublexical organization of the phonological language network: Evidence from an incidental phonotactic learning paradigm. *Journal of Memory and Language, 51*, 297–306.

Martin, N., & Saffran, E. M. (1997). Language and auditory–verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology, 14*, 641–682.

Mathias, J. L., & Wheaton, P. (2007). Changes in attention and information-processing speed following severe traumatic brain injury: A meta-analytic review. *Neuropsychology, 21*, 212–223.

Messer, M. H., Leseman, P. P., Boom, J., & Mayo, A. Y. (2010). Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology, 105*, 306–323.

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

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning, 62*, 302–331.

Murray, A., & Jones, D. M. (2002). Articulatory complexity at item boundaries in serial recall: The case of Welsh and English digit span. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28*, 594–598.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*, 251–269.

Neath, I. (1997). Modality, concreteness, and set-size effects in a free reconstruction of order task. *Memory & Cognition, 25*, 256–263.

Ottem, E. J., Lian, A., & Karlsen, P. J. (2007). Reasons for the growth of traditional memory span across age. *European Journal of Cognitive Psychology, 19*, 233–270.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*, 761–781.

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*, 3737–3753.

Pascual-Leone, J. (1970). A mathematical model for the transition rule in Piaget's developmental stages. *Acta Psychologica, 63*, 301–345.

Paulesu, E., Démonet, J. F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., et al. (2001). Dyslexia: Cultural diversity and biological unity. *Science, 291*, 2165–2167.

Payne, M. C., & Holzman, T. G. (1983). Auditory short-term memory and digit span: Normal versus poor readers. *Journal of Educational Psychology, 75*, 424–430.

Perham, N., Marsh, J. E., & Jones, D. M. (2009). Syntax and serial recall: How language supports short-term memory for order. *The Quarterly Journal of Experimental Psychology, 62*, 1285–1293.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46*, 115–154.

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words: Further evidence for the influence of semantic factors in immediate serial recall. *Quarterly Journal of Experimental Psychology, 48A*, 384–404.

Poirier, M., Saint-Aubin, J., Mair, A., Tehan, G., & Tolan, A. (2015). Order recall in verbal short-term memory: The role of semantic networks. *Memory & Cognition, 43*, 489–499.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*, 855–863.

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition, 81*, 149–169.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*, 1926–1928.

Saint-Aubin, J., & Poirier, M. (1999a). The influence of long-term memory factors on immediate serial recall: An item and order analysis. *International Journal of Psychology, 34*, 347–352.

Saint-Aubin, J., & Poirier, M. (1999b). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *Quarterly Journal of Experimental Psychology, 52A*, 367–394.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27*, 763–776.

Salthouse, T. A., Mitchell, D. R., Skovronek, E., & Babcock, R. L. (1989). Effects of adult age and working memory on reasoning and spatial abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 507–516.

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

Stone, M., Gabrieli, J. D., Stebbins, G. T., & Sullivan, E. V. (1998). Working and strategic memory deficits in schizophrenia. *Neuropsychology, 12*, 278–288.

Stuart, G., & Hulme, C. (2000). The effects of word co-occurence on short-term memory: Associative links in long-term memory affect short-term memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 796–802.

Swingley, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Psychology, 43*, 454–464.

Taylor, J., Macken, B., & Jones, D. M. (2014). A matter of emphasis: Linguistic stress habits modulate serial recall. *Memory & Cognition*. Published online October 2014.

Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*. Harvard, MA: Harvard University Press.

Vihman, M. (2014). *Phonological development: The first two years*. Chichester: John Wiley & Sons.

Wechsler, D. (2008). *Wechsler adult intelligence scale* (4th ed.). San Antonio, TX: Pearson.

Wechsler, D. (2009). *Wechsler memory scale* (4th ed.). San Antonio, TX: Pearson.

Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., et al. (2011). Improving digit span assessment of short-term verbal memory. *Journal of Clinical and Experimental Neuropsychology, 33*, 101–111.

Woodward, A., Macken, W. J., & Jones, D. M. (2008). Linguistic familiarity in short-term memory: A role for (co-)articulatory fluency? *Journal of Memory and Language, 58*, 48–65.