ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Journal of Memory and Language xxx (2006) xxx–xxx

Journal of Memory and Language

www.elsevier.com/locate/jml

A revised model of short-term memory and long-term learning of verbal sequences ☆

Neil Burgess [a,*], Graham J. Hitch [b]

[a] Institute of Cognitive Neuroscience and Department of Anatomy, University College London, 17 Queen Square, London WC1N 3AR, UK

[b] University of York, UK

Received 3 March 2006; revision received 4 August 2006

Abstract

The interaction between short- and long-term memory is studied within a model in which phonemic and (temporal) contextual information have separate influences on immediate verbal serial recall via connections with short- and long-term plasticity [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]. Long-term learning of sequences of familiar items is correctly predicted to interact with temporal grouping but not phonological similarity or articulatory suppression. However the model fails to predict learning of different sequences simultaneously, or of partially repeated lists. In a revised model, sufficiently different sequences recruit different context signals while sufficiently similar sequences recruit the same signal, via a cumulative matching process during encoding. Simulations show this revised model captures the experimental data on Hebb repetition, including the importance of matching at the start of a list, makes novel predictions concerning the effects of partial repetition, and provides a potential mechanism for position specific intrusions and the build up of proactive interference.

Keywords: Long-term memory; Phonological loop; Working memory; Hebb effect; Computational modelling; Serial order

Introduction

The nature of the interface between long-term memory (LTM) and short-term memory (STM) or 'working memory' has long fascinated psychologists (e.g., Atkinson & Shiffrin, 1968; James, 1890), and remains a topic of much theoretical interest (e.g., Baddeley, 2000; Burgess & Hitch, 2005; Ranganath & Blumenfeld, 2005). We consider this problem from the perspective of the working memory model of Baddeley & Hitch (1974; Baddeley 1986). According to this account, STM for a sequence of verbal items reflects the operation of a 'phonological loop', and this system is critical for the

☆ We acknowledge useful discussions with many colleagues including Alan Baddeley, Dennis Norris, Mike Page, Tom Hartley, and Garry Pilmore. We also acknowledge the helpful suggestions of guest editors Stephan Lewandowsky and Evan Heit, and referees Matthew Botvinick and Bennet Murdock. Neil Burgess is supported by a Medical Research Council Senior Fellowship.

* Corresponding author. Fax: +44 020 7679 1145.
E-mail address: n.burgess@ucl.ac.uk (N. Burgess).

0749-596X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2006.08.005

long-term learning of new word-forms (Baddeley, Gathercole, & Papagno, 1998). However, the working memory model is deliberately broad and does not address how such long-term learning takes place. We take the view that an important part of the development of a quantitative understanding of the interaction between STM and LTM will be played by the development of explicit computational models. Starting from our own computational model (Burgess & Hitch, 1999) of the role of the phonological loop, we investigate the potential mechanisms behind the interaction between STM and LTM. Here, we focus on the 'Hebb repetition' paradigm (Hebb, 1961). This paradigm consists of a series of immediate serial recall (ISR) trials over which one of the lists of items is occasionally repeated. ISR reflects the performance of the phonological loop, and the influence of long-term learning can be seen in a gradual improvement of performance for repeated lists compared to novel lists.

One aspect of our computational model is that learning a novel sequence of familiar items involves strengthening connections between representations of the items and states of an internal context or timing signal, see Fig. 1 and Burgess (1995), Burgess and Hitch (1992, 1999). The context signal is entrained by the temporal organisation of the stimuli, and contrasts with the rest of the model, which deals with the dynamics by which items are selected and effects due to their phonological composition. In accordance with this division of labour, associations to the context signal are responsible for effects of both temporal grouping of stimuli and long-term learning of their serial order, but not for phonological/articulatory effects such as concurrent articulatory suppression or manipulations of word-length or phonemic similarity. Thus, and uniquely among contemporaneous models of serial order, the model predicts a specific pattern of interference between effects of Hebb repetition and other experimental manipulations: namely that long-term learning should interact with manipulations of the re-presented list's temporal grouping, but not with manipulations of phonological/articulatory variables. We note that the second prediction is counter-intuitive given that articulatory suppression and phonemic similarity can severely disrupt STM performance and would thus seem more likely to reduce any long-term learning resulting from the immediate recall task.

Recent experimental results using the Hebb repetition paradigm, described more fully elsewhere (Hitch, Flude, & Burgess, 2006), have confirmed these predictions. However, these experiments also draw attention to the fact that different sequences cannot all be coded using the single set of context states assumed in the Burgess and Hitch (1999) model (see also Cumming, Page, & Norris, 2003; Hitch, Fastame, & Flude, 2005). Here, we describe a modification of the model that is capable of long-term learning of multiple sequences, in which different sequences become associated with different context sets. We report simulations of the revised model

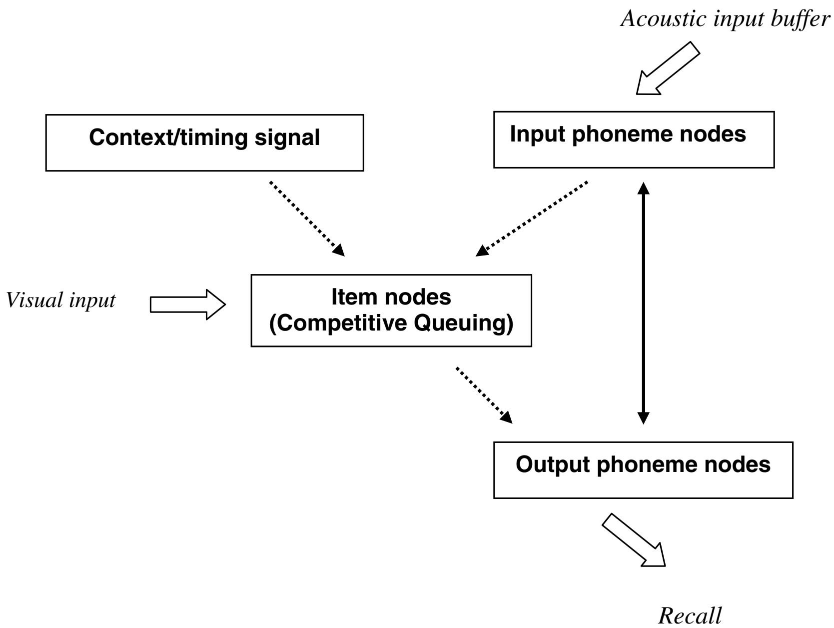

Fig. 1. Outline of the Burgess and Hitch (1999) model. Boxes denote layers of nodes, text in boxes denotes what the nodes represent. Modifiable all-to-all connections are shown as dashed lines. Pre-wired one-to-one connections are shown by a solid line. Block arrows denote where visual and auditory inputs access the system and where recall is effected by motor systems.

Please cite this article as: Neil Burgess, Graham J. Hitch, A revised model of short-term memory and long-term ..., Journal of Memory and Language (2006), doi:10.1016/j.jml.2006.08.005

showing that it captures the pattern of results in these recent experiments and is sufficiently well specified to generate further testable predictions. We begin by giving a brief review of the phonological loop and how the Burgess and Hitch (1999) model extends this account to explain serial ordering and long-term learning.

The phonological loop

According to Baddeley (1986), the phonological loop involves the interplay of a rapidly decaying phonemic store and a control process of subvocal rehearsal that can be used to refresh the contents of the store. Access to the phonemic store depends on presentation modality: the store is fed directly by external speech, but visual items must first be verbally recoded by means of subvocalisation. This simple conceptual model was proposed to account for the effects of a cluster of variables on STM for sequences of verbal items, principally those of word length, phonemic similarity and articulatory suppression. The tendency for lists of long words to be harder to recall than lists of short words (Baddeley, Thomson, & Buchanan, 1975) is explained by the extra time taken in rehearsing longer words, leaving more time for information in the phonological store to decay. The phonemic similarity effect, whereby lists of similar sounding items are less well recalled than lists of dissimilar items (Conrad & Hull, 1964), is explained by the greater difficulty of discriminating partially decayed traces of similar items in the phonological store. Articulatory suppression, the act of repeating an irrelevant utterance, disrupts ISR (Murray, 1967) and is assumed to interfere with subvocal rehearsal. Consistent with this analysis, articulatory suppression removes the dependence of recall on word length and phonemic similarity. However, whereas suppression removes the word length effect for visual and auditory lists, it only removes the phonemic similarity effect for visual lists (Baddeley, Lewis, & Vallar, 1984). This asymmetric effect of presentation modality is explained by the different routes by which spoken and visual stimuli are assumed to gain access to the phonological store.

We note that the idea of the phonological loop has been challenged in various ways (see e.g., Nairne, 2002). Among these are its explanations of the word length effect (see e.g., Lovatt, Avons, & Masterson, 2000; Service, 1998), the phonemic similarity effect (Jones, Hughes, & Macken, 2006; Jones, Macken, & Nicholls, 2004), and its ability to explain other phenomena such as the effect of irrelevant sound on ISR (Macken & Jones, 2003). However, many of these challenges can be met (Baddeley, 2006) and the phonological loop remains influential as a simplistic and highly applicable conceptualisation of verbal STM. Our own interest in modelling (Burgess & Hitch, 1992, 1999) was to explore how the core assumptions might be implemented computationally. Our general aim was to obtain a deeper understanding of the mechanisms that underpin the ISR of verbal material in terms of a simplistic model capable of both explaining existing data and generating novel predictions. Our approach involves assessing the model by testing its predictions, and, if justified by the results, developing it further. We note that a more general goal of our modelling is to map between computational mechanisms and neural levels of explanation (see Burgess & Hitch, 1999). However, we focus here on computational issues. These issues prompted us to note two major omissions from Baddeley's (1986) concept of the phonological loop: a means for storing and retrieving information about the serial order of a sequence and an interface with long-term memory (see also Burgess & Hitch, 1992, 2005)

The absence of a mechanism for serial order in the phonological loop might be considered surprising given the prevalence of order errors in immediate serial recall, where an item is recalled in the wrong place in the sequence (Conrad, 1965). Order errors typically involve items migrating to adjacent positions, and their distribution is a principal determinant of the characteristic bowed shape of the serial position curve (e.g., Henson, 1998). Order errors increase when the list items are phonemically similar (Conrad, 1965) and decrease when the items are presented in rhythmic groups, resulting in a multiply bowed serial position curve and a modified error distribution that reflects the grouping structure (Ryan, 1969). The phonological loop can account for the effect of phonemic similarity on order errors by assuming that order errors are actually item substitutions, as originally suggested by Conrad (1965). For example, incorrect recall of the sequence BTDP as BDTP is attributed to two item substitution errors (D for T and T for D) in retrieving partially decayed information from the phonological store. However, Baddeley's (1986) description of the phonological loop provides no explanation for the characteristic distributions of order errors, or for the effects of temporal grouping on the serial position curve.

Turning to the absence of an interface to LTM, this is reflected in the wholly short-term nature of the phonological loop. Thus the original concept does not embrace evidence that long-term memory contributes to performance in ISR tasks, as suggested by the effects of word frequency and various aspects of word structure (e.g., Gathercole, 1995; Hulme, Maughan, & Brown, 1991). Moreover, evidence that the phonological loop plays a role in vocabulary acquisition (Baddeley et al., 1998) begs the question of how the system contributes to long-term learning.

In contrast to the omission of serial ordering in Baddeley's (1986) account of the phonological loop, a number of computational models of STM have included mechanisms for serial order (see e.g., Brown, Preece, &

Hulme, 2000; Burgess & Hitch, 1992, 1999; Gupta & MacWhinney, 1997; Henson, 1998; Page & Norris, 1998). See also Estes (1972), Dosher (1999), Estes and Maddox (1995), Lewandowsky and Murdock (1989), Murdock (1995), Nairne (1990) for alternative approaches. However, to our knowledge, Burgess and Hitch (1999) is the only model that deals with both serial ordering and the role of STM in long-term learning, but see also (Botvinick & Plaut, 2006; Botvinick & Huffstetler, 2006; Farrell & Lewandowsky, 2002).

A network model of the phonological loop

Burgess and Hitch (1999) modelled the phonological loop as a simplified neural network containing representations of three types of information: lexical items, phonemes and a context/timing signal (see Fig. 1), following Burgess and Hitch (1992), Burgess (1995). Items and phonemes are each represented locally by single nodes within separate layers that can have different levels of activation. The context/timing signal has a distributed representation corresponding to patterns of activation over nodes in the context layer that reflect the similarity of different context states (see Fig. 2). Apart from hard-wired one-to-one connections linking separate layers for input and output phonemes, connections between adjacent layers are modifiable by learning. Thus, item nodes for familiar stimuli such as letters, digits or words are assumed to have strong pre-experimental connections to the nodes representing their constituent phonemes. When each item in a sequence is presented, an item node is selected by competitive queuing (see below) and learning is accomplished by strengthening connections between simultaneously active nodes in adjacent layers. Subsequent decay of these changes and the effects of noise result in forgetting. The strengthening process has two components, one large-amplitude and short-lived, the other small-amplitude but slowly decaying and so more cumulative (perhaps related to the different time-courses observed in synaptic potentiation, e.g.,

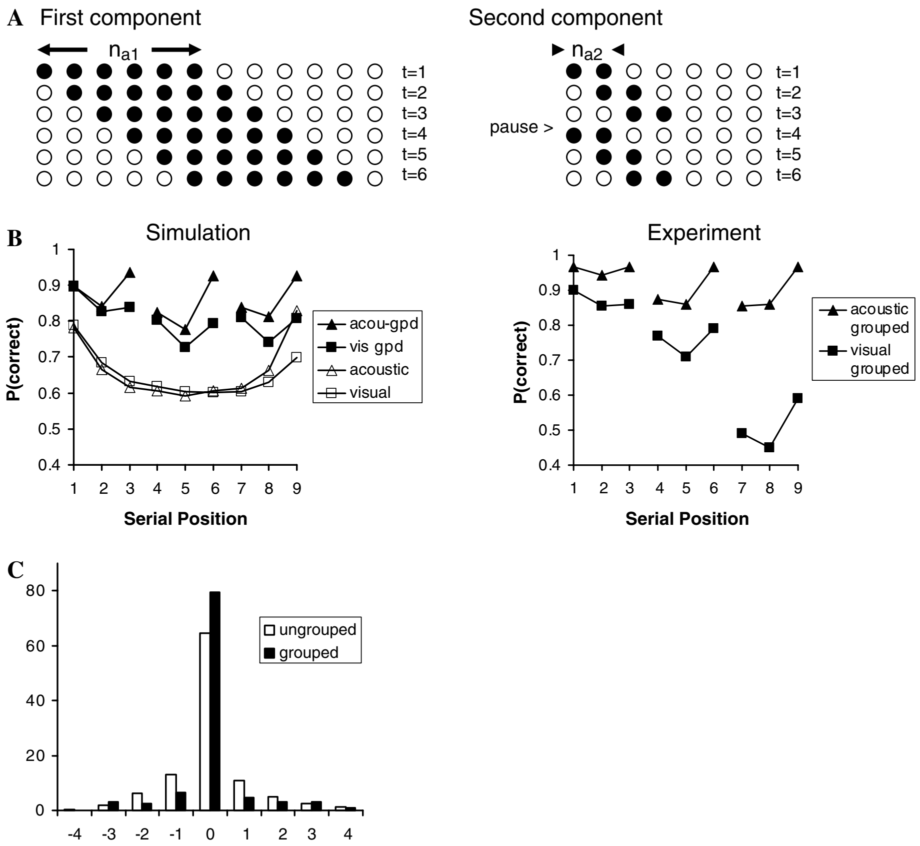

Fig. 2. Modelling temporal grouping effects (adapted from Burgess and Hitch, 1999). (A) The two dimensional context signal for a list presented in regular groups of three items. Filled circles denote active nodes, empty circle denote inactive nodes. One component tracks position within list, the other position within group. (B) Simulations of the effect of temporal grouping of a 9-item sequence of visual or auditory items by inserting a pause after every third item (left), and experimental data adapted from Frankish (1985, right). (C) Simulated distribution of serial positions at which an item is recalled relative to its position at presentation, note the slight increase in errors from serial positions differing by three places when grouping in threes.

Please cite this article as: Neil Burgess, Graham J. Hitch, A revised model of short-term memory and long-term ..., Journal of Memory and Language (2006), doi:10.1016/j.jml.2006.08.005

Bliss & Collingridge, 1993; Zucker & Regehr, 2002, see Burgess, 1995). The quickly decaying process is responsible for short-term memory and rapid forgetting, whereas the slowly decaying process enables long-term knowledge to accumulate when an input is repeated, provided not too much time has elapsed between repetitions.

The context signal is important in remembering serial order and is a feature taken up by some other models (Brown et al., 2000; Gupta & MacWhinney, 1997; Henson, 1998). Nodes in the context layer can be thought of as oscillators entrained by the timing of item presentation, such that patterns of activation for items presented at similar times tend to overlap (see also Henson & Burgess, 1997). During presentation of a list, the learning process strengthens connections between the context and item layers. For items presented at regular intervals, this process encodes order in the form of position–item associations. List presentation also strengthens item–phoneme and phoneme–item connections. However, for items such as digits, letters or words, that are highly familiar as a result of pre-experimental learning, the long-term components of these connections are already at their maximum value.

A critical aspect of the model is that recalling a sequence involves a selection process of 'competitive queuing' (Grossberg, 1978; Houghton, 1990) in which items are active in parallel and the most active item is output (also used during presentation, see above). The selected item is inhibited immediately after output to allow selection of the next item, but then recovers from the inhibition. This short-term queuing mechanism has recently received strikingly direct electrophysiological support (Averbeck, Chafee, Crowe, & Georgopoulos, 2002; Bullock, 2004). The same process occurs during presentation of the list: the item node most strongly activated by each stimulus at presentation is also selected and suppressed by competitive queuing.

Another important feature of the model is that recall involves cycling through two stages for each item: first selecting a candidate item and then retrieving its phonemic composition (see also Burgess, 1995; Page & Norris, 1998). In the first stage, context nodes receive the appropriate pattern of activation and context–item connections activate nodes in the item layer. In the second stage, activation from the item layer spreads to the output phoneme layer and feeds back to the item layer through the loop of connections via the input phoneme layer. The item recalled is taken to be the item node with the strongest activation after the effect of this feedback and the addition of noise to all item activations. After recalling the selected item, its node is subjected to strong but decaying inhibition, context nodes are updated by allowing the internal oscillators to move to their next state, and the cycle is repeated.

The network model correctly reproduces the effects explained by Baddeley's (1986) concept of the phonological loop as follows. Rehearsal is simulated as multiple output cycles, with modification of connection weights at output corresponding to "refreshment of information in the phonological store", and relies on the internal connections from output to input phonology. The word length effect arises because longer words take longer to rehearse and recall, allowing more time for the fast decay of connection weights to occur. The phonemic similarity effect occurs because of increased cross-talk in phonemic feedback in the second stage of retrieval. This cross-talk increases errors in which one similar item is substituted by another, much as in Conrad's (1965) idea of item-substitutions. In this way, phonemic similarity increases order errors in recall, without, according to the model, affecting the context signal encoding order information. More generally, the use of competitive queuing ensures the correct distribution of order errors in ISR, with omitted items likely to be recalled next, having had longer to recover from inhibition during presentation or a previous rehearsal (Burgess, 1995).

The effects of articulatory suppression are approximated by adding noise to all of the output phoneme nodes. The effect of adding noise to the output phoneme nodes feeds back to the input phoneme nodes, disabling the re-strengthening of input phoneme to item connections by output or rehearsal (and their strengthening at presentation in the case of visual items), see Appendix A for details. As a result, suppression not only impairs recall but removes its dependence on visual items' phonemic composition. Simulations of these effects are reported in Burgess and Hitch (1999). However, although the model of articulatory suppression is simple, it exaggerates the size of the effect. In the revision, see below, we only change the parts of the model to do with the context signal, and so retain this flaw of the previous model (see Figs. 3, 5, 6).

As intended, the network model also simulates serial order phenomena not addressed by Baddeley's (1986) concept of the phonological loop. For example, serial position curves and effects of temporal grouping are generated by the way the context-timing signal operates. Previous authors have suggested that serial position effects arise from differences in the distinctiveness of positional cues (e.g., Lee & Estes, 1981; Murdock, 1960). In a similar way, serial position curves arise in the model because the context signals for items at the ends of a list are more distinctive than for items in the middle. Note however, that ISR itself is only mildly dependent on the context signal (e.g., digit span is only mildly reduced by removal of the context signal, Burgess & Hitch, 1999, Table 3) because the recovery of item nodes from inhibition during presentation also provides an ordering mechanism: the earlier items having had longer to recover than the more recent ones.

Temporal grouping effects arise in the model because the context signal reflects the timing of items within groups as well as within the list. Temporally grouped lists recruit a second component in the context signal: a set of units whose pattern of activation tracks position within group rather than position within list (see Fig. 2A, see also Henson & Burgess, 1997). The input from this second component modulates the input from the main component of the context input to provide a more distinctive, two-dimensional coding of position relative to ungrouped presentation. The result is fewer order errors overall and multiply bowed serial position curves for grouped lists, together with a slight increase in errors where an item migrates to a corresponding position in a different group. These are all characteristic features of human performance (Ryan, 1969). Fig. 2B presents experimental data on grouping together with model simulations of these effects. Consistent with separate roles for phonemic and timing information in the model, temporal grouping effects have been found to be relatively insensitive to phonemic similarity, word length and articulatory suppression, (Hitch, Burgess, Towse, & Culpin, 1996).

As should be clear, our model assumes that the passage of time plays a key role in STM: this is evident in the use of oscillators to code contextual information, the decay of connection weights over time, and decay of the inhibitory input driving response suppression. However, it is important to appreciate that these time-dependent processes are driven by events (i.e. the presentation or recall of items) so that performance is not simply dependent on the sheer passage of time. Thus, it is inappropriate to classify our model as 'time-based' in a dichotomy with 'event-based' models (see e.g., Lewandowsky, Brown, Wright, & Nimmo, 2006). Such a dichotomy is overly simplistic given the existence of hybrid models such as ours with both time-based and event-based properties.

One of the reasons the context signal was assumed to be event-driven rather than time-driven in Burgess and Hitch (1999) was that errors in ISR reflect similarity of the ordinal positions of items rather than their times of occurrence (Henson, 1999; see also Ng & Maybery, 2002). Henson and Burgess (1997) provide a detailed description of how a system of temporal oscillators can reproduce these event-driven characteristics, by virtue of the oscillators becoming entrained to the temporal pattern of input items (for auditory input, the oscillators follow the envelope of the speech signal as in Hartley, 2002). Other evidence points to limitations on the role of elapsed time in STM. For example, slowing down the speed of recall can have remarkably little effect on ISR (Lewandowsky, Duncan, & Brown, 2004), as can varying the time intervals between items during sequence presentation (Lewandowsky & Brown, 2005; Lewandowsky et al., 2006), or the time allowed for response inhibition to recover (Duncan & Lewandowsky, 2005). These findings all point to the importance of event-driven processes. However, it is crucial to note that the evidence also implicates time-based processes. For example, the most thorough investigation of the word length effect to date shows an effect of articulatory duration, consistent with temporal decay (Mueller, Seymour, Kieras, & Meyer, 2003). Previous evidence for an item-complexity account of word length effects was based on less reliable methodology. We conclude that any adequate model of ISR should have both time-based and event-based properties. The Burgess and Hitch (1999) model can be viewed as one attempt to meet this requirement.

To return to our main theme, we consider the interaction between STM and LTM in the Burgess and Hitch (1999) model. A simple example of long-term learning in immediate serial recall is given by prior list intrusions, where an error in recall of the current list can be traced to an item in a corresponding position in the immediately preceding list (Conrad, 1960; Henson, 1999). According to our model, prior list intrusions reflect competition between long-term context-item associations learned for the previous list and those for the current list. We return to this issue in the revised model. However, we begin by focusing on another example of long-term learning that occurs when an entire list of familiar items is presented and recalled several times. In this case, learning is assumed to result from cumulative strengthening of context–item connections for the repeated list through their long-term plasticity. In the following section, we use the model to generate more detailed predictions for long-term learning in the Hebb repetition paradigm (Hebb, 1961).

Network model predictions for the Hebb repetition paradigm

In Hebb's (1961) procedure, random sequences of digits were presented for ISR, but unannounced to participants, the same list was presented every third trial. Performance on the repeated list gradually improved, a phenomenon known as the Hebb Effect. Hebb used this observation to argue that encoding a sequence in STM generates a long-term trace that outlives STM. Given that the digits themselves are highly familiar, it follows that what participants learned was primarily sequence information, i.e. a particular ordering of this set of items rather than the items themselves.

As noted above, the Burgess and Hitch (1999) model assumes that learning a sequence of familiar items involves the strengthening of context–item connections through cumulative changes in their slow weights. For familiar items such as digits, sequence repetition does not affect item–phoneme connections as these are already saturated through prior learning. The assumption that slow weights also decay (but over a timescale significantly longer than a single trial) provides a natural account of the observation that the Hebb Effect is contingent on there being only a small number of 'filler' lists separating successive presentations of the repeated list (Melton, 1963).

The Burgess and Hitch (1999) model makes a number of novel predictions about sequence learning. One is that learning will be sensitive to experimental manipulations that affect the context/timing signal such as temporal grouping. In contrast, manipulations that have their effect on the phoneme nodes, such as phonemic similarity and articulatory suppression, should have little or no effect on sequence learning, even though they can strongly affect ISR. Hitch et al. (2006) report a series of experiments that broadly confirmed these predictions. Thus, articulatory suppression and phonemic similarity had their characteristic effects of disrupting ISR, and interacted with each other as expected with visual presentation. However, neither articulatory suppression nor phonemic similarity had any significant effect on sequence learning. Moreover, replication of a study first reported by Bower and Winzenz (1969) confirmed that the Hebb Effect is sensitive to the consistency of the temporal grouping pattern used to present the repeated list. The two studies differed in detailed results but agreed in showing significantly slower long-term learning when the repeated list was presented using different grouping patterns. Overall, we took these results as providing support for the basic assumptions of our model, and in particular the crucial separation between a temporal context signal on the one hand and representations for items and phonemes on the other.

Notwithstanding the above success, at a more detailed level there were clear indications that the model was incorrect. The most telling problem arose from our final experiment in which two lists were repeated in parallel in the same series of trials. The results showed simultaneous Hebb Effects for both sequences. This observation is not itself surprising as nothing in our everyday experience suggests we have to learn sequences such as telephone numbers or new word-forms one at a time. The problem for our network model however was fundamental as it assumed that all sequences were learned by strengthening links from a common set of contextual cues. If so, there would be massive interference when more than one sequence is learned at the same time (see also Cumming et al., 2003). The combination of this specific shortcoming and the overall success of our model led us to see if it could be revised to accommodate the ability to learn multiple sequences.

A revised network model

There are other pointers to problems in the way the Burgess and Hitch (1999) model handles sequence learning besides its failure to account for our ability to learn different sequences in parallel. For example, the idea that learning a sequence involves strengthening context–item associations predicts that learning one sequence will affect the recall of another sequence with a related structure. Thus, having learned the sequence 59264183, there should be transfer of training to a derived sequence in which every other item occupies the same position, e.g., 53214986. The transfer effect should be reflected in zig-zag serial position curves showing improved recall at positions containing the same items as in the learned sequence. However, despite finding some evidence for short-range position-specific transfer (reminiscent of Conrad's (1960) serial order intrusions), Cumming et al. (2003) found no evidence for long-lasting position-specific transfer. Moreover, Cumming et al. (2003) attempted to simulate the time course of an actual Hebb repetition experiment using the Burgess and Hitch (1999) model and found that a reliable Hebb effect was not possible because the non-repeated lists caused too much interference. This is a further indication that a single set of context–item associations is not a viable mechanism for sequence learning. Other methods of studying transfer effects in sequence learning agree in providing only weak evidence for position–item associations (Hitch et al., 2005), and go further in indicating the importance of the start of the list as the point from which a learned sequence is built up. Hitch et al. (2005) examined transfer of training after ISR of a set of lists that each contained a sub-sequence in common. The training lists varied in length and in separate conditions they either started or ended with the same sub-sequence of items. This sub-sequence was later presented as a separate list for ISR and performance was compared with a previously unseen control list. There was considerable transfer of training when the repeated sub-sequence had appeared at the start of each training sequence and very little transfer when it had appeared at the end. This asymmetry is consistent with Schwartz and Bryden's (1971) observation that the Hebb Effect disappeared if the first two digits of the repeated list were changed at each presentation but remained if the same number of digits were changed at the end of the sequence.

The theoretical challenge, therefore, is to revise the Burgess and Hitch (1999) model in such a way as to cope with its inadequacies while maintaining the features that have received empirical support here and in previous research. As a starting point, we note that the model derives from an attempt to model serial order in word production that also used competitive queuing (Houghton, 1990; see Burgess & Hitch, 1992). In Houghton's model, sequence information is coded by associations between the constituents of a word and successive states of a 'start' and an 'end' node which both evolve over time. This signal is conceptually analogous to the context signal in our own model (see Henson & Burgess, 1997). Houghton was concerned with our ability to produce a range of different words and proposed that as a word becomes familiar it acquires its own dedicated start and end signals. By this means it is possible to avoid catastrophic overload of a single set of cues. We have incorporated this idea in the revision to our own model by the simple expedient of allowing there to be multiple sets of context nodes rather than a single set. At the same time, we altered the learning rule to increase the specificity of the pattern of activation arising from the context signal: allowing connection weights to inactive items to become negative as well as those to active items becoming positive.[1] Thus, a given context state will inhibit items not associated with it as well as activating those that are associated with it. In addition, to allow fair competition between context sets, we normalise the long-term weights of context–item connections after each presentation and retrieval (by dividing all weights by the same amount). This means that the length of the weight vector to each item node remains constant, while the pattern of positive or negative weights can change.[2] Thus learning results in different context sets becoming a better or worse match to a given sequence of items, rather than simply increasing the strength of the input from a single context signal (see Rumelhart & Zipser's, 1985 'competitive learning algorithm' for a similar approach). In addition, more recent weight modifications will be more influential than previous modifications which are washed out by repeated normalisations.

The long-term weights for each set of context nodes at the start of a Hebb experiment reflect the residue of the model's prior sequence learning. On being presented with a further sequence, the model computes, from each context set, the cumulative match between the input from these connections and the currently active item. This matching process enables the model to detect whether a sequence is familiar or not, in a manner reminiscent of Adaptive Resonance Theory (Grossberg, 1976, see also Carpenter & Grossberg, 2003). Any set of context nodes whose cumulative match to all so-far-presented items in the sequence falls below a threshold value is discarded (rendered inactive) at that point.[3] Thus, as successive items are presented, the cohort of candidate context sets gets smaller. As a result, either one set emerges as the winner (the context set with the greatest cumulative match) or there is no winning set (all previously used context sets having been discarded). In the latter case, the sequence recruits a new context set. Learning (connection weight modification) occurs during each step of sequence presentation for all context-sets that are still in the cohort. The winning or new context set is the only one to remain active to control the retrieval phase, during which learning also occurs (this makes sense because weight modification should depend only on activation values, it also makes Hebb repetition dependent on actual retrieval of the items rather than solely on their presentation, see Cohen & Johansson, 1967; Cunningham, Healy, & Williams, 1984).

The re-use of an old context set that was previously used for a similar list will benefit retrieval because the long-term connection weights from the old set will already match the context-item associations corresponding to the current list. The strength of the match will increase with re-use: providing the Hebb repetition effect. The sequential elimination of context sets as soon as they fall below threshold follows the observed importance of the start of the list—if this does not match there is no Hebb effect (Hitch et al., 2005).

In addition to improved performance due to re-use of old context sets, there will also be implications for the occurrence of errors. In the 1999 model, position-specific intrusion errors from prior lists (see Conrad, 1960; Henson, 1999) resulted from the repeated use of a single context set. In the revised model, they resulted from re-use of old context sets. Thus, occasional selection of a previous context set associated with a partially matching list will increase the chance of a mis-matching item from the previous list being retrieved in the current list, possibly creating knock-on errors in the following list positions. Factors that increase the occurrence of partially matching lists will increase the occurrence of position-specific intrusions (e.g., previous presentations of similar lists, or the Hebb repetition paradigm itself, given the high rate of errors in supraspan lists). Accordingly, it is possible that the observed build up of proactive interference over the first few trials of a block (Keppel & Underwood, 1962; Sanders & Willemsen, 1978), or over long-

[1] We increased θ in Eqs. (1) and (2) of the 1999 model from 0 to 0.5 for context-item connections, the same value as for the input phoneme to item connections, see Appendix.

[2] The length l of the vector of connection weights from a context set to each item unit is fixed at $A/\sqrt{n}$, where n is the number of item units (26 for letters of the alphabet), by multiplying each element by $A/(l\sqrt{n})$. This ensures that the length of the entire vector of weights from a context set is A. A value of $A = 3$ was found to produce a rough match to the data, and was used in all simulations. The size of the increase in context-item connections with each trial was increased from 0.1 to 0.3 to counteract the moderating effect of this.

[3] A context set gets a match of 1 if the currently active item receives at least as much activation from the long-term context-item weights as any other item, and a match of 0 otherwise. The cumulative match is the running average of matches to the items in the list so far. If this falls below the threshold value (0.6), that context set is discarded.

er periods (Lustig & Hasher, 2002; Underwood, 1957), might result in part from the increasing likelihood of selecting a partially matching context set. The revised model also makes specific predictions regarding the effects of different types of partial repetition, such as how the Hebb effect is disrupted by novel items and 'capture' effects by previous lists with similar initial subsequences. We come back to these points in Discussion of simulation results.

The threshold value of the goodness of the match required for a context set to be re-used is a free parameter that needs to be chosen. The Cumming et al. (2003) experiment found essentially no transfer of learning to a probe list in which every second item was in the same location as in the repeated list. This indicates that, to recruit a previous context set, the number of matching items must be greater than 0.5. Thus we used a threshold value of 0.6, which was found to produce a rough match to the data, and was used in all simulations. Nonetheless, the behaviour of the model in terms of the pattern of effects of Hebb repetition, temporal grouping, articulatory suppression seems relatively insensitive to the precise choice of this parameter.

In summary, the revised model retains the architecture proposed by Burgess and Hitch (1999), but amends the way the context–item stage operates so as to allow multiple sequences to be learned. This is achieved by introducing multiple sets of context signals together with a cumulative matching process that allows a previously presented sequence to be recognised. It can be seen that by cumulating from the start of the sequence, the matching process is likely to give a good account of the importance of repetition from the start of a sequence in Hebb learning and transfer (Hitch et al., 2005; Schwartz & Bryden, 1971). The new model should also be able to reproduce the effects of articulatory suppression, phonemic similarity and temporal grouping reported in the experiments reported by Hitch et al. (2006). A more detailed description of the revised model is given in the Appendix. To check all of this, we implemented the revised version and carried out a number of simulations.

Simulations

For simplicity and ease of comparison, the model was run as in the 1999 simulations for lists of letters of the alphabet, excepting only the modifications relating to multiple context sets listed above. See Table A1 for the full list of parameters. We began by confirming that, by retaining the same architecture, the revised model replicates the patterns of ISR performance shown by the earlier model, e.g., the 'sawtooth' serial position curve for lists of letters of alternating phonemic similarity (see also Burgess, 1995), the scalloped serial position curve for temporal grouping, effects of presentation modality, articulatory suppression, word length, list length, item familiarity, recognition memory and of various neuropsychological impairments. For reasons of space we do not report these simulations. We focus instead on the effects of the new modifications on performance in the Hebb repetition experiments reported by Hitch et al. (2006), and the interaction of Hebb repetition with other experimental manipulations (e.g., Cumming et al., 2003).

At the start of presentation of a list, all previously used context sets (i.e. those with modified long-term connection weights) and a potential new context set (with uniform, unmodified, weights) are active. So long as the cumulative match of a previously used context set stays above a threshold value, its pattern of activation proceeds with the presentation of each item, and long-term and short-term connection weights are modified, as in 1999. If the cumulative match of a context set falls below threshold, it is made inactive (thus playing no further role in retrieval and receiving no further weight modification). At the end of presentation all surviving context sets except the one with the highest cumulative match are made inactive. The pattern of activation of this context set is then reset to instigate retrieval and proceeds with the retrieval of each item, continuing to receive modifications to its connections, as in 1999. Note that if none of the previously used context sets remains above threshold, then the new context set will be the one used at retrieval. The potential new context set remains active throughout list presentation, being given the minimal cumulative match (i.e. equal to the threshold), and having its connection weights modified. In a final set of simulations, we address the two important questions of the role of the prior history of learning and the total number of context sets that would be required, given that, in principle, a new list can recruit a new context set. All simulations were averaged over 24 runs, each using a different seed for the random number generator, representing 24 subjects.

Simulation 1: Hebb repetition and articulatory suppression

Simulation 1 consisted of the auditory presentation and immediate serial recall of 64 spoken lists of 12 items, arranged in eight blocks of eight trials. Each block of eight ISR trials follows an *aebeced* structure in which *a*, *b*, *c*, *d* trials used non-repeated 'noise' lists and *e* trials used the repeated list (Experiment 1, Hitch et al., 2006). The model was run with and without concurrent articulatory suppression. Fig. 3 shows the results the simulations. It can be seen that they reproduce the main pattern of the experimental data, in that list repetition leads to learning, articulatory suppression disrupts ISR, and crucially, list learning over repetitions occurs despite concurrent articulatory suppression. The two

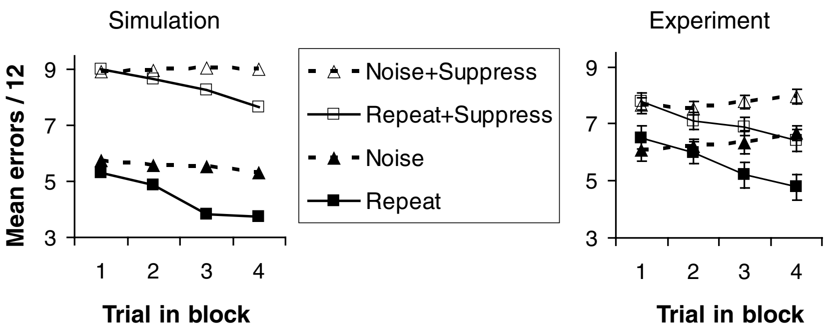

Fig. 3. Hebb repetition and articulatory suppression. (Left) Simulation of the Hebb repetition effect is shown under normal conditions (filled symbols): the mean number of errors per list decreases over the four trials of the repeated list (full lines) within each block of eight trials, but not over the four trials with non-repeated 'noise' lists (dashed lines). The Hebb repetition effect is also shown under articulatory suppression (open symbols). Lists of 12 dissimilar auditory items were used. (Right) Experimental data re-plotted from Hitch et al. (2006) Experiment 1. See text for further details.

effects, of repetition and suppression, appear to act independently, although there is a small influence of overall error rate on Hebb learning, as increasing numbers of errors reduces the probability of context re-use due to the reduced similarity of the presented list to the previous (incorrect) retrieval of the list (see Table 1).

As noted earlier, the reason for the dissociation between effects of repetition and suppression is that suppression disrupts the phonemic component of the model but has no effect on the context-item associations that mediate sequence learning. Some insight into the way the model operates is given by noting how often it re-uses an old context set rather than recruiting a new one (see Table 1). As can be seen, list repetition increases the proportion of re-used context sets from the baseline for non-repeated lists, and this pattern also persists under articulatory suppression. Nonetheless, the reduction of context re-use as performance decreases (as under suppression) is also clearly seen in this Table. The re-use of context sets illustrates how the matching process enables the model to build on prior learning. It is intriguing to note that for learning to occur the match need not be with the particular context set used to encode the previous presentation of the Hebb list (see Table 1, re-use of context set for list $n-2$). The context sets for Hebb list $n-2$ can be different to those for lists $n-4$ and $n-6$, depending on the different errors made on each recall. However, re-use of these previous context sets, or any sufficiently similar previous set, will also carry an advantage. Thus, the model does not have to 'recognise' the specific previous presentation of the Hebb list for learning to take place, suggesting a possible point of correspondence with data indicating that human participants need not be aware of the list repetition for learning to occur (Hebb, 1961; McKelvie, 1987, but see also Sechler & Watkins, 1991).

We also simulated a situation comparable to the Cumming et al. (2003) experiment, using lists of 10 visual items (as they did). As above, the simulation involved blocks of eight ISR trials following an *aebeced* structure in which *a*, *b*, *c*, *d* trials used non-repeated 'noise' lists and *e* used the repeated list. However, as in Cumming et al. (2003), the final 'repeated' list (or 'probe' list) only contains every second item in the same location as the preceding repeated lists, with the intervening items in a new order. See Fig. 4. The choice of 0.6 for the strength

Table 1
Percent of re-used context sets for repeated and noise lists in Simulation 1 of Hebb repetition, as a function of position in trial-block and articulatory suppression

	Position in block				
	1	2	3	4	Mean
No suppression					
Repeat	0 (0)	85 (85)	96 (71)	98 (72)	70
Noise	1 (0)	0 (0)	0 (0)	0 (0)	0
Suppression					
Repeat	0 (0)	34 (33)	62 (22)	82 (26)	45
Noise	1 (0)	0 (0)	0 (0)	0 (0)	0

See text and Fig. 3 for details. Figures show the mean proportion over trials of each type of list (repeated or 'noise'), figures in brackets denote the proportion of re-used contexts that were also used for list $n-2$ (i.e. the previous presentation of a repeated list).

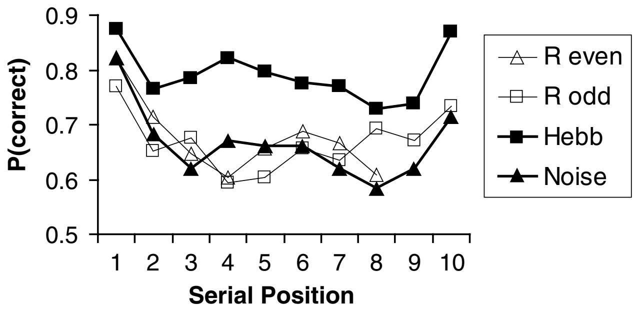

Fig. 4. Simulation of partial repetition: alternate items in a final probe list are repeated in the same locations as in previous repetitions, as in Cumming et al. (2003). A Hebb repetition simulation was run as in Fig. 3, with lists of 10 visual items, but with the fourth repetition of the Hebb list replaced by a 'probe' list repeating only the odd or even items from previous repetitions in the correct locations. The Figure shows serial position curves for the final probe lists (R odd, R even) as well as the final noise list and the final presentation of a fully repeated Hebb list. No Hebb effect is found for the partially repeated list.

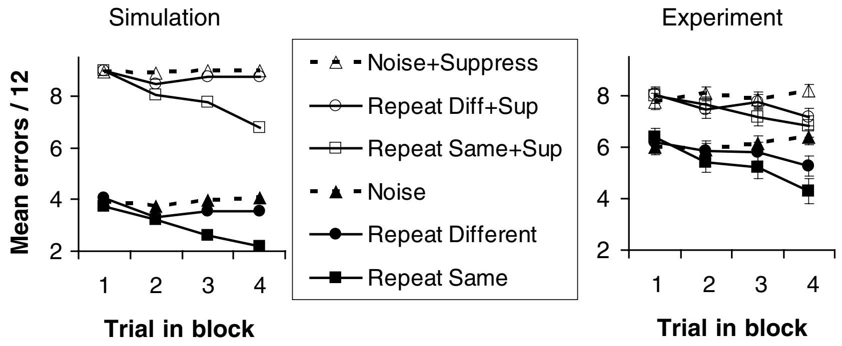

Fig. 5. Hebb repetition and temporal grouping. (Left) Simulation of the Hebb repetition effect is shown for lists repeated over four trials with the same temporal grouping (squares), while significantly less effect is shown for lists repeated with different temporal grouping (circles), compared to non-repeated 'noise' lists (in which neither sequence nor grouping is repeated, triangles and dashed line). This pattern remains under normal conditions (filled symbols) and under articulatory suppression (open symbols). Lists of 12 dissimilar auditory items were used, with grouping patterns taken from permutations of (4, 4, 2, 1, 1). (Right) Experimental data re-plotted from Hitch et al. (2006) Experiment 2. See text for further details.

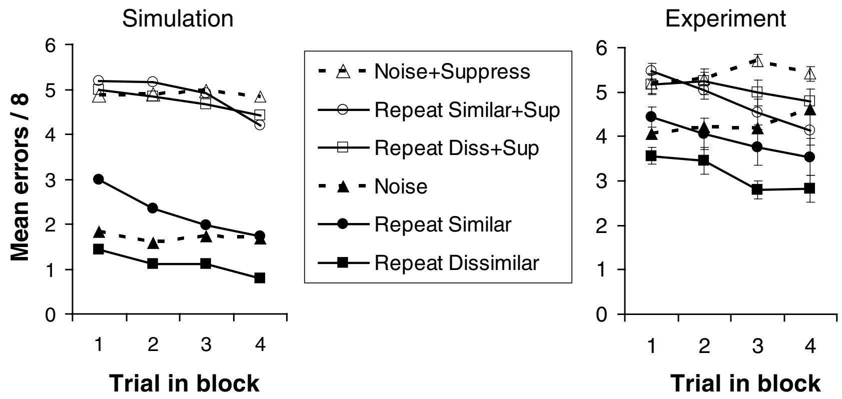

Fig. 6. Hebb repetition and phonemic similarity. (Left) Simulation of the Hebb repetition effect is shown for lists of phonemically similar (circles) or dissimilar (squares) items, compared to non-repeated 'noise' lists (triangles, dashed line). This pattern remains under normal conditions (filled symbols) and under articulatory suppression (open symbols). Lists of eight visual items were used. (Right) Experimental data re-plotted from Hitch et al. (2006) Experiment 3. See text for further details.

Please cite this article as: Neil Burgess, Graham J. Hitch, A revised model of short-term memory and long-term ..., Journal of Memory and Language (2006), doi:10.1016/j.jml.2006.08.005

of the context match means that these probe lists do not recruit previous context sets, and do not show the increased performance of normal Hebb repetition, as expected.

Simulation 2: Hebb repetition and temporal grouping

Before reporting the simulations it is necessary to describe the way the modified model handles Hebb repetition of temporally grouped lists. First, we review the unmodified model of temporal grouping, details of which are described more fully in Burgess and Hitch (1999) or Hitch et al. (1996). As discussed above, presentation of a temporally grouped list recruits a second component of the context signal, in which the pattern of activation proceeds with each item to reflect position-within-group (cf. the first component in which it reflects position-within-list, see Fig. 2A). The input from the second component modulates the input from the first component to the item units. Connections from currently active units in the second context component to each presented item (i.e. to the currently active item node) are set to value $1/n_{a2}$, where n_{a2} is the number of active units in each state ($n_{a2} = 2$ in all simulations). The effect of this learning is that the input to an item i from the second component at time t during recall equals the fraction of units in the second component that were active during the presentation of item i that are again active at time t during recall, referred to as $F_i^2(t)$. This leaves the total context input to an item unchanged at positions equal to the item's within-group position during recall (when $F_i^2(t) = 1$), but reduces it at different within-group positions (when $F_i^2(t) < 1$): reducing the likelihood of order errors in which the item ends up at a new within-group position.

In the revised model, the re-use of a context set depends on the match provided by the combined input from both components and includes re-use of both components. To enable this more complex matching process, the connection weights from the second component of context sets are normalised after presentation and retrieval, as are those of the first component. For both components, the normalisation of weights ensures that recent modifications are more influential than previous modifications, which are washed out by repeated normalisations. Given its *modulatory* function it is important that the input from the second component to an item is at most 1.0 (this maximum input occurs at recall positions with the same within-group position as the item). This is ensured by normalising the connection weights onto a given item by division so that the maximum weight is $1/n_{a2}$. The input from the second component of context modulates the long-term input during presentation (affecting a given context set's cumulative match) and the short-term input during retrieval (causing the effects of temporal grouping, as before). Thus, if a previous list is re-presented with the same temporal grouping, the previously used context set will likely be re-used as the cumulative match for it will be strong (the modulatory input will most often have value 1). However, if it is re-presented with a different temporal grouping, the previously used context set will less likely be selected to control retrieval. This is because some items will have a much worse match due to being presented at a different within-group position of the new grouping structure (and thus having a low modulatory input). See Appendix for details.

Fig. 5 shows the results of using the revised model to simulate the manipulation of grouping in Hitch et al. (2006) Experiment 2. The simulation consisted of 64 trials arranged in the same way as in Simulation 1. Repeated lists were either consistently grouped across their 4 presentations within each block of trials or given a different grouping pattern each time. In simulations, permutations of groups of size 4,4,2,1,1, were used; in the experimental data grouping included permutations of 4,4,2,1,1 and also of other sets of group sizes (4,3,2,2,1; 4,3,3,1,1; 4,4,3,1; 4,3,3,2; 4,4,2,2). Auditory presentation was used to maximise the influence of grouping. Once again, there is a reasonable overall correspondence with the experimental data. Most importantly, the Hebb effect is reduced in that the model learns more slowly when a repeated list is presented using different temporal grouping patterns rather than the same grouping. Detailed inspection of the operation of the model shows that this difference reflects a much lower frequency of re-using old context sets when the grouping pattern changes, see Table 2. The Hebb effect is also similar for lists repeated with the same grouping as for lists repeated with no (i.e. even) grouping, as in the experimental data, see Fig. 3. We note also that Fig. 5 confirms that, to a first approximation, articulatory suppression does not interact with these long-term learning effects in the model, as observed experimentally.

Simulation 3: Hebb repetition and phonemic similarity

Simulation 3 considers the effect of phonological similarity on Hebb repetition and consists of presentation of 72 lists of eight visual items each, arranged in six blocks of 12 ISR trials. The 12 ISR trials in a block follow a *sdnsdnsdnsdn* structure in which *s* trials used a repeated list of phonemically similar items, *d* trials used a repeated list of phonemically dissimilar items, and *n* trials used non-repeated 'noise' lists of items of mixed phonemic similarity, following Hitch et al. (2006) Experiment 3. Visual presentation was used to allow conformation that articulatory suppression removed the phonological similarity effect under these conditions.

Fig. 6 shows the results of the simulations. As can be seen, they correspond reasonably closely to the experimental data in that phonemic similarity and articulatory suppression disrupt ISR and interact with each other in

Table 2
Percent of re-used context sets for repeated and noise lists in Simulation 2, as a function of temporal grouping, position in trial-block and articulatory suppression

	Position in block				
	1	2	3	4	Mean
No suppression					
Repeat same	0 (0)	99 (99)	100 (99)	100 (97)	75
Repeat different	0 (0)	13 (13)	23 (11)	34 (15)	18
Noise	1 (0)	0 (0)	0 (0)	1 (0)	1
Suppression					
Repeat same	0 (0)	97 (97)	100 (72)	100 (70)	74
Repeat different	0 (0)	10 (10)	20 (10)	27 (11)	14
Noise	1 (0)	0 (0)	0 (0)	0 (0)	0

Figures show the mean proportion over trials of each type of list (repeated with same temporal grouping, repeated with different temporal grouping, or 'noise' lists in which neither sequence nor grouping are repeated), figures in brackets denote the proportion of contexts that were also used for list $n-2$ (i.e. the previous presentation of a repeated list).

Table 3
Percent of re-used context sets for repeated and noise lists in Simulation 3, as a function of phonemic similarity, position in trial-block and articulatory suppression

	Position in block				
	1	2	3	4	Mean
No suppression					
Phonologically similar	1 (0)	58 (57)	86 (71)	95 (78)	60
Phonologically dissimilar	1 (0)	79 (78)	97 (91)	99 (93)	69
Noise	4 (0)	4 (0)	1 (0)	3 (0)	3
Suppression					
Phonologically similar	1 (0)	24 (20)	51 (31)	71 (37)	37
Phonologically dissimilar	1 (0)	22 (20)	53 (34)	76 (43)	38
Noise	2 (0)	2 (1)	3 (1)	3 (0)	3

Figures show the mean proportion over trials of each type of list (repeated similar, repeated dissimilar, or 'noise'), figures in brackets denote the proportion of contexts that were also used for list $n-3$ (i.e. the previous presentation of a repeated list).

the correct way. Articulatory suppression largely removes the phonemic similarity effect, while long-term learning is still present under suppression. The re-use of a context set associated with one list by a subsequent list depends on the number of items in common to both lists and their orders. Their phonemic composition does not affect this. However, as noted above, there is a reduction of context re-use as the overall error rate increases (Table 3). Thus the exaggerated effect of articulatory suppression in the simulations means that the Hebb effect under articulatory suppression is harder to see than in the experimental data. Nevertheless it is significant across the 24 simulated subjects.[4] We return to this in the discussion below. The simulations also provide an important demonstration that the revised model meets its goal of being able to learn more than one sequence at a time.

[4] A 3 (list type) $\times$ 4 (position in block) ANOVA gives a significant effect of position in block ($F(3,69) = 9.15$, $p < 0.001$) modified by an interaction with list type ($F(6,138) = 2.74$, $p < 0.05$), with simple effects of position in block for similar ($F(3,69) = 11.54$, $p < 0.01$) and dissimilar repeated lists ($F(3,69) = 3.25$, $p < 0.05$) but not for noise lists ($F(3,69) < 1$).

Simulation 4: Disruption of the Hebb Effect by partial repetition

To a first approximation, the revised model correctly reproduces the complex pattern of experimental findings regarding the Hebb effect and articulatory suppression, phonemic similarity and temporal grouping (Hitch et al., 2006). We now test the revised model in novel situations, partly to check that it works as we think, and partly to demonstrate that it can make clear predictions for new experiments. Following Hitch et al. (2005), Cumming et al. (2003) and Schwartz and Bryden's (1971), we investigate the effects of partial repetitions of a list.

First we investigated insertion of a single new item into a repeated list (replacing one of the repeated items).

Please cite this article as: Neil Burgess, Graham J. Hitch, A revised model of short-term memory and long-term ..., Journal of Memory and Language (2006), doi:10.1016/j.jml.2006.08.005

We replicated Simulation 1, and looked also at the effect of introducing a new item on the fourth presentation of a repeated list, varying the serial position of the new item. Fig. 7 shows the mean simulated serial position curves for the fourth 'noise' list and for the fourth presentation of the repeated list within each block of eight trials in Simulation 1. It also shows the serial position curves for a 'probe' list used instead of the fourth repetition of the repeated list, in which the nth item of the repeated list has been replaced by a new item (curves shown for $n = 1, 2, 3, 6, 9, 12$ in the 12 item auditory lists).

When a new item is introduced at positions $n = 1$ or 2, performance on the list is equivalent to that on a non-repeated 'noise' list, and the context sets used for the previous (third) repetition of the list is never recruited (the cumulative match after one or two item will be zero and the context set rejected). When the new item is introduced at positions $n = 3$, 6, 9 or 12, the previously used context set is recruited as often as in the normal Hebb Effect (98%, see Table 1), but performance drops for that position and subsequent positions. Performance at the position of the new item drops below that for a 'noise' list (i.e., more than negating any Hebb Effect), and slowly recovers for subsequent positions. Inspection of the simulated data shows that a majority of the errors at the position of the new item are premature retrievals of the next item in the list (item $n + 1$). The next item will normally be the second most active item at position n, after item n. However when a new item is presented at position n its connections from the context layer have not received additional long-term learning, unlike item $n + 1$, making item $n + 1$ more likely than usual to be selected erroneously over item n by the competitive queuing mechanism. Subsequent errors arise from premature retrieval of the following item and retarded retrieval of the new item, with these disruptive effects slowly dying out over the following positions. Interestingly, relatively little disruption is caused by replacement of the final item in these auditory lists, showing the benefit of the increased short-term connections from the input phoneme nodes for this item (see Burgess & Hitch, 1999).

We next investigated the effect of introducing a paired transposition of adjacent items into the repeated list. Fig. 8 shows the mean simulated serial position curves for the fourth 'noise' list and the fourth repeat of the repeated list within each block of eight lists, as above. It also shows the serial position curves for simulations when a 'probe' list replaces the fourth repeat, in which the nth and $n + 1$th item of the repeated list have been transposed. As with the insertion of a new item at the start of the list, transposing items 1 and 2, or 2 and 3, causes a massive reduction in context re-use (down to 1% and 4%, respectively), and consequent loss of any Hebb repetition effect. A possible exception to this is a slight benefit for the first item in lists in which items 2 and 3 are swapped. Again similarly to insertion of a new item, transposing later pairs has very little effect on re-use rates (all at least 97%). Now, however, the initial dip in performance at the location of the transposed pair (caused by a range of order errors, including restitution of their previous order) has relatively little knock-on effect for the remainder of the list.

The final simulation investigated the effect of partial repetition. As with simulations 1 and 2, these simulations consist of presentation of 64 lists of 12 items each, arranged in eight blocks of eight immediate serial recall (ISR) trials. Now each block follows the *aebeced* structure in which *a*, *b*, *c*, *d* trials use 'noise' lists and trial *e* uses a partially repeated list in which the initial n items always have the same order, and the remaining items have new orders in each list. Fig. 9 shows the serial position curve for the fourth Hebb list and fourth 'noise' list

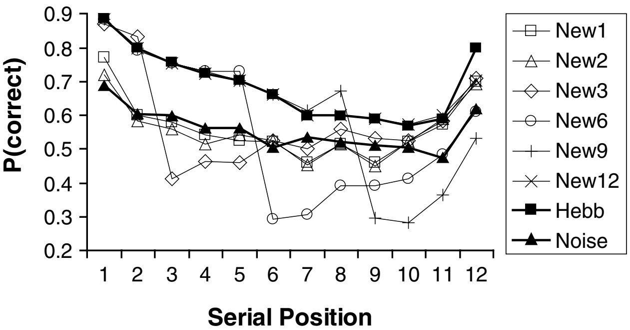

Fig. 7. Disrupting the Hebb Effect with a new item. Serial position curves are shown for Simulation 1 (see Fig. 3), showing the fourth repeat of the repeated list in each block ('Hebb') and the fourth non-repeated list in each block ('Noise'). Serial position curves are also shown for simulations in which the nth item of the fourth repeat is replaced by a new item, for $n = 1, 2, 3, 6, 9, 12$. The curves for $n = 1$ and 2 resemble that of a non-repeated list, the curve for $n = 12$ resembles that of a regular repeated list, while a new item at positions $n = 3,6,9$ causes disruption at position n and several following positions.

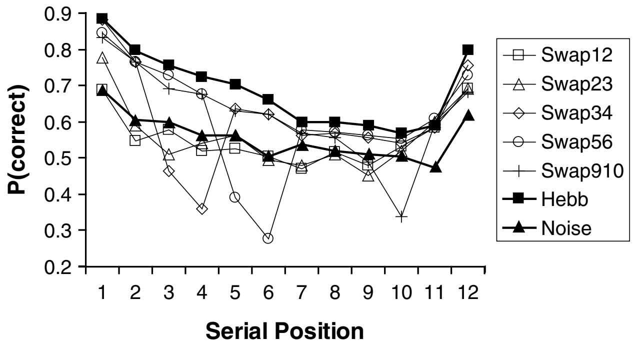

Fig. 8. Disrupting the Hebb Effect by transposing adjacent items. Serial position curves are shown for Simulation 1 (see Fig. 3), showing the fourth repeat of the repeated list in each block ('Hebb') and the fourth non-repeated list in each block ('Noise'). Serial position curves are also shown for simulations in which the nth and n+1th item of the fourth repeat are transposed, for n = 1, 2, 3, 5, 9. The curves for n = 1 and 2 resemble those of a non-repeated list, the curves for n = 3,5,9 show the Hebb Effect at positions other than n and n+1.

as in the usual Hebb paradigm using lists of 12 dissimilar items (see Simulation 1), and for lists in which only the initial n items have been repeated for n = 2, 5, 8. When only the first two items are repeated, there is no context re-use, and performance resembles that for a noise list. When items 1 to 5 are repeated, there is some context re-use (46% for the fourth repeat), and some performance advantage at the start of the list. When items 1 to 8 are repeated, context re-use is as high as for the Hebb list (99% for the fourth repeat), and a performance advantage is clearly present for the first eight items. Further, the advantage for the first 5 or 8 repeated items in partially repeated lists falls short of that for fully-repeated Hebb lists, consistent with the 'retrograde compatibility effect' found experimentally (Botvinick & Huffstetler, 2006).

In summary, the fourth set of simulations indicate that the model is consistent with previous data regarding partial repetition of sequences. They show the dependence of the Hebb Effect on the overall similarity of the repeated list (see also Fig. 4 and Cumming et al., 2003), and stress the importance of similarity at the start of the list (Hitch et al., 2005; Schwartz & Bryden, 1971). In addition, they make predictions regarding the pattern of performance in specific novel experimental situations.

Simulation 5: Effect of prior experience

In this simulation we examined the dependence of the matching process on the model's prior learning history. This is clearly a factor of considerable theoretical interest, and crucial to the question of how many context sets might be required in total, see Discussion of simulation results, below. In addition to the idea that long-term connections also slowly decay (not simulated here), returning a 'used' context set to the pool of 'new' context sets if the associated list is not rehearsed, it might be that a large enough fixed pool of context sets will suffice. Thus, as prior experience grows, eventually every list will have a context set with a close-enough match (i.e., one

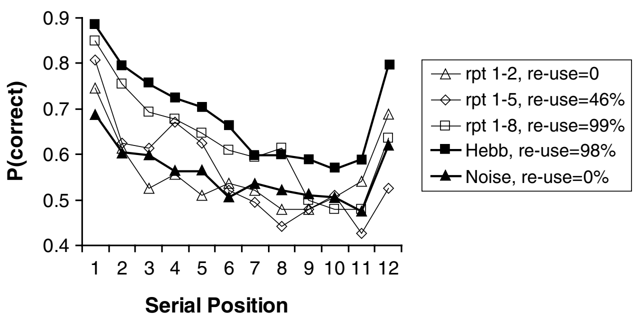

Fig. 9. Partial repetition of the start of a list. Serial position curves are shown for simulation 1 (see Fig. 3), showing the fourth repeat of the repeated list in each block ('Hebb') and the fourth non-repeated list in each block ('Noise'). Serial position curves are also shown for the fourth 'repeated' list in simulations in which only the first n items of the 'repeated' list are repeated in the correct positions on each trial. When n = 2, there is no context re-use and performance follows that of the Noise list. When n = 5, there is some context re-use (46%) and performance on the early items is somewhat improved. When n = 8, there is full context re-use (99%) and performance on the first eight items is clearly improved.

previously associated with a list with a similar start, see above).

The simulation runs the model after different amounts of prior exposure to previous lists (following the same *aebecede* pattern as in Simulations 1 and 2, with lists of 8 dissimilar items as in Simulation 3). Fig. 10 shows how performance and the frequency of context re-use alter with the extent of prior experience. It can be seen that the rate of re-use rapidly approaches ceiling for repeated lists, as would be expected. Interestingly, there is also a steady increase in the re-use of old context sets for novel lists. As outlined above, high levels of re-use indicate that a finite number of context sets may suffice: eventually all lists will re-use a context set. The Hebb effect remains because the context sets used by recently repeated lists will tend have better matched long-term weights. A Hebb effect caused in this way would also be accentuated by slow decay of long term connection weights (not simulated here) which would lend further advantage to recently repeated lists.

The overall level of performance reduces slowly with prior experience, reflecting errors caused by re-use of context sets associated with slightly different lists. Interestingly, the extra errors tend to occur more frequently in the later serial positions (because re-used context sets have to provide a good match at the beginning of the list), steepening the otherwise rather shallow primacy gradient shown by the model (see Fig. 11). Compared to this baseline, the relative effect of Hebb repetition increases slightly, as the repeated lists show less reduction in performance with prior experience (Fig. 10), reflecting the greater likelihood that the context sets they re-use are associated with the same list (specifically, with the previous presentation of that list, Fig. 10 dashed lines). These data indicate that the model might actually require only a relatively small set of context sets (e.g., hundreds), despite the large total number of possible lists in the world. At the same time, our data draw attention to the importance of exploring prior experience systematically in future modelling and experimental investigations.

Discussion of simulation results

We have demonstrated that our 1999 model can be extended to capture the pattern of performance in ISR under manipulations of Hebb repetition, temporal grouping, articulatory suppression and phonemic similarity, see Hitch et al. (2006). The model makes specific predictions regarding the effects of various ways in which a list can be partially repeated, and performs consistently with previous such experiments, in particular reproducing the importance of similarity at the start of the repeated list (Cumming et al., 2003; Hitch et al., 2005; Schwartz & Bryden, 1971; Botvinick & Huffstetler, 2006).

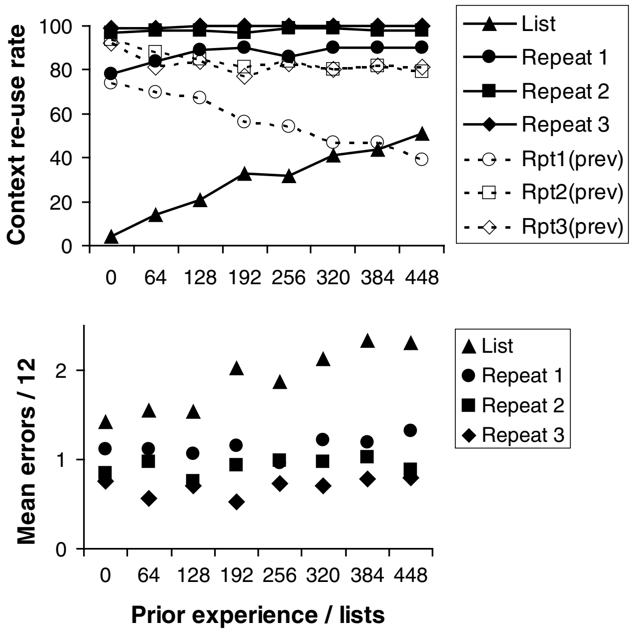

Fig. 10. Effect of the extent of prior experience on Simulation 1 (Fig. 3), i.e., the number of lists presented before the Hebb experiment begins. (Above) The rate of context re-use as a function of prior experience for the first presentation of a list ('List') and for the first, second and third repetition of it ('Repeat 1–3'). Dashed lines show the proportion of lists for which the context set associated with their immediately previous presentation is re-used ('Rpt1-3(prev)'). (Below) Performance in recalling lists presented once, twice, three and four times as a function of prior experience.

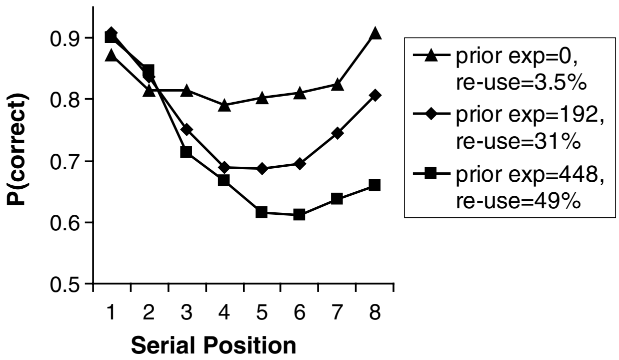

Fig. 11. Effect of the extent of prior experience on serial position curves. Serial position curves are shown for items in the first presentation of a noise or Hebb list in each block in the simulations shown in Fig. 10. The amount of prior experience is varied: i.e. the number of previously learned lists (0, 192, 448, producing the following average percentages of context re-use, respectively: 3.5, 31, 49).

An interesting aspect of the model concerns the occurrence of position-specific intrusions and the build up of proactive interference in general. In addition, the model implies that position-specific intrusions result from re-use of partially matching context sets. For example, simulations of retrieving a repeated list show increased position-specific intrusion of items recalled out of order in the previous repetition of the list (and decreased intrusions from the immediately preceding noise list), compared to simulations of retrieving a noise list. More generally, re-use of partially matching context sets will cause additional errors that are not scored as position-specific intrusions, suggesting a mechanism for proactive interference. In experiments in which prior lists contain re-orderings of the same items, re-use of a partially matching context set will cause order errors, as seen in Fig. 8. We note that these errors will tend to be towards the end of the list (since there must be a good match between the starts of the previous and current lists for re-use to occur). It follows that the marked primacy gradient seen in longer lists may be due in part to the build up of proactive interference within an experiment (Keppel & Underwood, 1962; Sanders & Willemsen, 1978), and over longer timescales (Lustig & Hasher, 2002; Underwood, 1957). In addition, the frequency of errors due to prior learning will depend on the characteristics of the vocabulary of items from which lists are formed (e.g., digits, letters, phonemically similar letters), and might explain some aspects of performance in experiments using lists of mixed composition (see e.g., Farrell & Lewandowsky, 2002)

An important question for the model concerns the total number of context sets that might be required. We noted in Introduction that long term learning might exist in equilibrium with very slow decay of the long-term connection weights (which we have not simulated here). Thus, over longer timescales (e.g., months) without rehearsal of a Hebb list, the strength of the association of a particular context set to it, and its improved retrieval, will wane and the context set will return to the pool of 'new' context set to be used. In principal, the rate of decay of these long-term associations might be estimated from the relative rates of proactive interference from recent and more distant exposure to lists. In addition, the final simulation indicated that, as prior experience increases, so does the extent of re-use of context sets even for novel lists. Thus, depending on the nature of the stimuli, a finite pool of context sets might suffice, with the Hebb effect resulting from better matching context sets being available for recently repeated lists than for novel lists.

For continuity with the previous model, and the wide range of data that it explains, we have changed only those aspects of the 1999 model necessary for this extension. Thus, as can be seen, the performance of the current model makes only approximate fits to the data in each experiment. For example, as in 1999, the model's overall performance on long lists is too good (because the serial position curve does not fall steeply enough, see Figs. 2, 4, 7–9), and the effect of articulatory suppression is too strong, see Figs. 3–5. The over-strong effect of articulatory suppression combines with an influence of performance on Hebb repetition (such that increasing errors decreases the Hebb effect by making the retrieved Hebb lists less similar). This produces an apparent interaction between articulatory suppression and the Hebb effect (most clearly in Fig. 6) that is not obvious in the experimental data. A dependence of the Hebb effect on actual retrieval is consistent with previous experiments (Cohen & Johansson, 1967; Cunningham et al., 1984), and a weak interaction between performance level and Hebb effect is present as a non-significant trend across the experiments in Hitch et al. (2006). However, a more sensitive experimental design will be required to check this aspect of the model.

Some of the shortcomings of the 1999 model impact on the revised model presented here. Firstly, the role of list length in determining the number of units used in a context set seems a little unnatural (the moving window of activation requires $n + n_a - 1$ units, where n is the number of items and n_a is the number of units active in each context state, see Fig. 2). This hinders the model's ability to make predictions regarding the effects of Hebb repetition over partially repeated lists of different length, and to correctly model the pattern of errors in temporal grouping experiments using variable group lengths. A related problem is that the rate of context re-use depends to some degree on list length, since the chance of getting a cumulative match greater than 0.6 after repetition of a list (whose retrieval will be imperfect) will be lower for a long list than for a short one. A second shortcoming concerns the inelegant role of the second component of context in modulating the first component's long-term input to the matching process during presentation while modulating its short-term input during recall. In addition, the lower overlap between states of the second component of context than those of the first component reflects the greater precision of positional coding within the shorter group than within the entire list: as indicated by results showing relative rather than absolute coding of position (Henson, 1998). All of these shortcomings might well be reduced by using a start-end context representation rather than a moving window of activation (see e.g., Henson, 1998; Henson & Burgess, 1997; Houghton, 1990), which will be the subject of future work. We return to some of these issues in General discussion.

Notwithstanding these deficiencies, we hope to be able to use the model to make predictions concerning the effects of experimental manipulations of the type investigated here, and to improve the model, make more predictions, and so on in future work. This paper, and Hitch et al. (2006) represents one cycle of this process. In this way we would hope to advance understanding of the functional organisation of short and long-term memory.

General discussion

We begin by briefly summarising the position so far. Our starting point was the phonological loop model of verbal short-memory (Baddeley, 1986), arguably the most successful current account despite its many flaws and critics. We noted that in computational terms, the phonological loop fails to incorporate any ideas about how order information is maintained or how this subsystem links to long-term memory. These omissions motivated our network model (Burgess & Hitch, 1992, 1999) which assumes separate mechanisms for order and item information. According to the model, order is coded via connections between lexical representations of items and a context/timing signal item, whereas item information is coded in terms of associations between lexical and phonological representations. Both types of connections are modifiable, having short-term and long-term plasticity (Burgess, 1995). The limit on immediate recall is explained by the rapid decay of short-term changes to connections in the presence of noise, whereas learning comes about through cumulative changes in the same connections as a result of their long-term plasticity. We showed how, according to this account, STM and long-term learning of sequences of familiar items should be sensitive to different factors.

In recent experimental work (Hitch et al., 2006), we tested these qualitative predictions of the 1999 model using a modification of Hebb's (1961) procedure for studying incidental long-term learning of sequences held in STM. The predictions were confirmed in the form of a broad experimental dissociation. Thus, phonemic similarity and articulatory suppression (affecting the lexical–phonological component of the model) disrupted STM but had little effect on long-term learning, whereas altering the timing of a repeated sequence (affecting the temporal context signal) impaired long-term learning without affecting STM. These experiments also produced the incidental observation that participants were capable of learning more than one sequence simultaneously. This ability is inconsistent with the assumption made by Burgess & Hitch (1999) that a single context/timing signal is used to code all sequences. This observation, together with weak evidence for solely positional transfer effects in the Hebb procedure (Cumming et al., 2003; Hitch et al., 2005), stimulated the attempt to revise our model.

The revision of the 1999 model, presented here, introduces the idea of multiple context-sets, each reflecting sequences that the model has learned previously. In all other respects, the architecture and core features of the original model are retained. However, now when a sequence is presented, the revised model tries to match the input against the pattern of associations to all context sets. As successive items are presented, the cumulative match to each context set is maintained, and sets with cumulative matches that fall below a fixed threshold are discarded. In this way, the cohort of active context-sets reduces until the sequence is either 'recognised' as similar to one that has been encountered before or perceived as 'new' and assigned a new context-set. Simulations showed that the revised model is capable of reproducing the pattern of findings reported in Hitch et al. (2006). We begin General discussion by considering some of the implications of our research for other accounts of verbal short-term and long-term memory.

First, we note that despite its simplicity, our model invokes a more subtle relationship between STM and LTM than is evident in current debate, much of which tends to revolve around the question of whether STM is a separate system or merely the currently activated region of LTM (see e.g., Cowan, 2001). Our model has dual features in that the presence of long-term and short-term weights in the same connections could be seen as corresponding to the common system view, whereas on the other hand competitive queuing is a purely short-term process. What emerges as of more importance than the terms of current debate is the crucial role played by the balance between the systems ability to respond to and remember novel inputs and its ability to benefit from prior learning. Thus, in the terms of our model, a matching process that was too stringent would impair the ability to learn from past experience while being maximally responsive to new patterns whereas one that was too lax would leave the model trapped in the past.

In terms of implications for other computational models of serial order, a number of models besides ours make a distinction between the process of selecting an item and retrieving its phonological features during recall (e.g., Brown et al., 2000; Henson, 1998; Jones & Polk, 2002; Page & Norris, 1998). As has been noted, these models are more limited in scope in that none of them addresses long-term learning. However, they could in principle be elaborated by specifying suitable learning mechanisms to accommodate the dissociation between phonological and timing variables in the Hebb Effect. Such changes could be readily incorporated into models in which item selection involves a context signal that varies with experience or time (Brown et al., 2000; Henson, 1998). However, they would appear to be more difficult for the Primacy Model (Page & Norris, 1998) as this model does not include a mechanism for temporal grouping effects.

We regard it as important that our revised model generates further predictions that are open to experimental test. The model generates predictions for the rate of learning when a list is only partially repeated on each presentation, according to where in the list the variation occurs. An example of this is how the Hebb Effect will be disrupted by introduction of a novel item at different positions (Fig. 7), swapping two items (Fig. 8) and repeating only the first section of a list (Fig. 9). In addition, 'capture effects' are predicted, where the context set for a previously repeated list is captured by a new list if it is similar enough, particularly at the start of the list, such that repetitions of the new list should impair retrieval of the non-matching portion of the previous list. More generally, we note that although the revised model is inevitably more complex than its predecessor, it retains the desirable feature of being sufficiently well specified to generate clear qualitative predictions.

There are some interesting parallels between our present model and models of long-term memory that also assume a time-varying context signal, see Burgess & Hitch (2005). Such models typically associate items to a single context state for each list (e.g., Anderson, Bothell, Lebiere, & Matessa, 1998; Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Mensink & Raaijmakers, 1988; Raaijmakers & Shiffrin, 1981). It is particularly interesting to note that the use of a context signal at the higher level of lists is consistent with its use at the lower level of position within lists or subgroups of lists, as discussed here. For example, different levels of context might be combined hierarchically. Another possibility is that retrieval involves temporal discriminations and can operate on multiple time-scales (e.g., Neath & Brown, 2006; Neath, 1993).

Despite incorporating extra assumptions, we wish to emphasise that our revised model remains deliberately simplistic, like its predecessor. Thus we acknowledge many aspects of serial order and the interaction between verbal short-term and long-term memory that the revised model does not address. For example, patterns of errors in ISR suggest that context varies relative to the ends of sequences as well as the beginnings (Henson, 1998; Ng & Maybery, 2002). However, these observations, and the processing of lists or temporal groups of variable lengths can likely be addressed by changing the detail of the context representation without having to alter the structure of the model (see Henson & Burgess, 1997), see also Discussion of simulation results. A more substantial limitation is that the revised model still deals with serial order only at the lexical level, ignoring order at lower levels of representation such as phonology, and higher levels such as syntax for natural language inputs. For example, the model cannot simulate spoken recall or subvocal rehearsal as a sequence of phonemes. We note, however, that serial ordering can be successfully modelled at the phonological level by an evolving context signal combined with competitive queuing (Hartley & Houghton, 1996). Moreover, Gupta & MacWhinney (1997) have presented a model that combines serial ordering at the lexical and phonological levels. Modelling serial order at the phonological level is particularly important if one wishes to simulate the learning of new word forms, a crucial function of the phonological loop according to Baddeley et al. (1998). We conclude that our present model requires further elaboration to deal with nonword learning, but that it provides a sound basis for such development. Hartley & Houghton's (1996) model provides a useful starting point for this enterprise, but requires an analogous modification to the present revision of our own model to deal successfully with the learning of multiple nonwords.

Another reflection of the simplistic nature of our revised model concerns the way in which it deals with past experience. We have focused here on experience of sequences rather than items. As in the Burgess & Hitch (1999) & Burgess (1995) models, previous learning of the phonemic composition of individual items is represented in the long-term weights of phoneme–item connections. However, these connections were simply adjusted by hand. A fuller account would investigate how they are modified by experience, possibly by taking note of attempts to model age of acquisition effects in lexical processing (Ellis & Lambon Ralph, 2000). The revised model allows previous learning of sequences to be simulated—showing increasing proactive interference, particularly towards the ends of lists (see Figs. 10 and 11). However, it is still not clear how much prior experience should be simulated, or how this might be made to correspond more closely to a participant's previous linguistic experience, which would include item learning too. The amount of prior experience available to influence performance also relates to the timescale over which the 'long-term' connection weights for a given list might decay in the absence of rehearsal. Both of these issues affect the total number of context sets required by the model. The simulations of varying prior experience indicate that the model could reach a steady state in which a fixed number of several hundred context sets might suffice. That is, 'noise' lists could re-use context sets associated with partially matching lists from prior experience, and repeated lists re-use context sets associated with recent well-matching lists. The advantage from repeated lists would come from the better match (and more recently modified long-term connection weights) of the re-used context set. Empirically, the balance between the contributions of recent and more distant experience might be reflected in the balance between proactive interference within an experiment and over longer timescales. One final implication of proactive interference resulting from re-use of partially matching context sets is that the strength of the primacy effect in ISR will depend in part on proactive interference, see Discussion of simulation results.

The context signals are ambiguous with respect to the traditional distinction between STM and LTM. They are responsible for aspects of ISR traditionally associated with LTM, such as learning over repetitions and position-specific intrusions (as with the context signal in Burgess, 1995; Burgess & Hitch, 1999), as well as effects of temporal grouping within STM (as with Burgess & Hitch, 1999; Hitch et al., 1996). In the revised model, we have sought to make clear their role in the transition of order information from STM to LTM. Thus, in the absence of repetition of lists, context signals play a role in maintaining order in STM (but not a crucial one, given that recovery from inhibition during presentation provides an alternative ordering mechanism, see Burgess & Hitch, 1999) and mediate effects of temporal grouping, and long-term connection strengths from context sets to item nodes are not reliably strengthened. When order information is reliably repeated, however, a context set becomes associated with the repeated pattern and effectively provides a form of long-term memory for that sequence.

A role for the context signals in the transition of information from STM to LTM bears some parallels to Baddeley's (2000) concept of a multi-modal episodic buffer. Baddeley suggested this buffer in addition to the modality-specific short-term stores (phonological loop and visuospatial sketchpad) which each link to specific regions of semantic LTM (language in the case of the phonological loop). The episodic buffer is suggested to mediate the transition of multimodal information into episodic memory. In the case of serial order information, this would identify the temporal context component of our model with the episodic buffer. The association of hippocampal damage with deficits in Hebb repetition, but not immediate serial recall, both associates the hippocampus with the episodic buffer, and associates Hebb repetition with episodic memory (which is thought to depend on the hippocampus). See Burgess & Hitch (2005) for further discussion.

The multi-modal character of the episodic buffer would, in turn, lead us to predict that the context/timing signal governing serial order is not be specific to verbal materials. Interestingly, recent studies have identified a number of functional similarities between ISR for nonverbal and verbal stimuli (Avons, 1998; Jones, Farrand, Stuart, & Morris, 1995; Smyth, Hay, Hitch, & Horton, 2005), suggesting there may be a common mechanism. The Hebb repetition effect has also been shown for visuo-spatial sequences (see Gagnon, Foster, Turcotte, & Jongenelis, 2004), and shares some characteristics with the verbal Hebb effect, for example, the importance of the start of the list in partial repetition (Fastame et al., 2006). The question of whether there is a supra-modal serial ordering system and the nature of the mechanism is clearly an important topic for future work.

The more general question of what starting knowledge to give a model and what to expect it to acquire is fundamental, being no less than the computational equivalent of the nature/nurture debate concerning human linguistic and mnemonic abilities. We note that attempting to model the relationship between short-term and long-term memory has forced us to confront these issues, and that they will have to be addressed in any successful theoretical account (see also Botvinick & Plaut, 2006). We note also that the plasticity inherent in connectionist models makes them well-suited for this purpose (Elman et al., 1996).

A recent alternative approach starts from long-term learning of the capacity to perform ISR (training the connection strengths of a recurrent error back-propagation network over hundreds of thousands of trials), and showed that the resulting network can perform ISR on novel lists, without any further change in connection strengths (Botvinick & Plaut, 2006). In this model, ISR results from the dynamic patterns of activation supported by the recurrent connections rather than by short-term weight modification. This approach is particularly interesting because effects of LTM such as the bigram frequency effect occur in the model as a result of its explicit long-term training. Botvinick and Huffstetler (2006) extend this model to show that, by continuing the modification of connection strengths during an

ISR experiment, an effect of Hebb repetition can also be seen. The model makes no explicit separation between order information and phonological information, so it would be interesting to see if it shows the pattern of results in Hitch et al. (2006).

To conclude, we have shown how a simplistic network model of the phonological loop that assumes separate mechanisms for item and order information (Burgess & Hitch, 1999) can be used to derive qualitative empirical predictions about long-term sequence learning. The basic architecture of the model is such that variables affecting STM for sequences of familiar items (e.g., phonemic similarity, articulatory suppression) dissociate from variables affecting the learning of repeated sequences (e.g., timing). These predictions were broadly confirmed in experiments on the Hebb Effect (Hitch et al., 2006). However, in the course of that research a limitation of the model became evident, namely its inability to learn more than one sequence at a time. A modest but important revision addressed this shortcoming by introducing multiple context/timing signals and a matching process whereby an incoming sequence is compared with long-term memory for previously-learned sequences, re-using the context signal of any prior list that was sufficiently similar. In this way, the behavior of the revised model is affected by prior learning. Simulations showed that the revised model was capable of reproducing the pattern of existing experimental findings (Hitch et al., 2006; Cumming et al., 2003; Botvinick & Huffstetler, 2006; Bower & Winzenz, 1969; Schwartz & Bryden, 1971) and of generating further predictions. The model remains simplistic but draws attention to three further areas for exploration. One is the way in which the learning of several similar sequences will mutually interact depending on the serial positions of any differences and the rate at which they are introduced over trials. Another concerns extension to the learning of novel word forms, where the important sequencing process goes on at the phonological level rather than the lexical level studied here. A final area concerns the interaction between short-term and long-term memory where presumably more complex matching processes are required to deal with effects of prior learning for more realistic verbal materials, such as chunking and syntactic and semantic constraints.

Appendix A

We outline the equations and parameter values governing our simulations. Except where explicitly stated otherwise, these are the same as used in Burgess & Hitch (1999). See Table A1. All simulations were performed using letters of the alphabet as stimuli, as in Hitch et al. (2006) Experiment 3 (and avoiding the re-use of items within a list in the 12 digit strings used in their Experiments 1 and 2).

Activation values of all nodes are simply the sum of the activations of the nodes connected to them, weighted by the strength or 'weight' of the connections. Context nodes, input phoneme nodes activated by auditory presentation and item nodes activated by visual presentation are set to value 1.0. Competitive queuing corresponds to selection of the most active item node: setting its activation value to 1.0 and those of the other item nodes to zero. In the final selection of an item for output, Gaussian noise (mean zero, standard deviation $\sigma = 0.3$) is first added to activation values, causing errors in selection. The winning item node at presentation and output additionally receives a decaying inhibitory input (initially $\Omega = -2.0$).

Below, we consider the context signal: the part of the model to which we have made changes. We first consider the learning and normalisation of connection weights, focussing on the two components of the context signal separately, due to their rather different nature: the first component provides part of the input to item nodes and signals position within list; the second component is only recruited by temporally modulated presentation and modulates the input to item nodes from the first component (according to position within group), see main text and Fig. 2. We then consider how the selection of an old context set for re-use, or the allocation of a new one, depends on the match between old context sets and the presented list.

A. Connections from the first component of a context set

As with the other modifiable connections in the model (input phoneme–item and item–output phoneme connections, see Burgess & Hitch, 1999), the context–item connections have short and long term components that sum to provide the net weight of the connection. That is, the connection weight from unit j (activation a_j) to unit i (activation a_i) at time t is:

$$W_{ij}(t) = W^s_{ij}(t) + W^l_{ij}(t). \tag{1}$$

The short-term component decays according to:

$$W^s_{ij}(t+\delta) = \Delta^{\delta} W^s_{ij}(t), \tag{2}$$

where the time-step δ for processing an item is its spoken "word-length" in seconds (0.4 s for letters of the alphabet), and the decay rate $\Delta = 0.75\ \mathrm{s}^{-1}$. The inhibition given to item nodes following selection by competitive queuing also decays at rate Δ. Both types of connection weight are modified with experience according to:

$$W^s_{ij}(t+\delta) = \begin{cases} \alpha\kappa[a_j(t) - \theta]a_i(t) & \text{if } a_i(t) > 0; \\ W^s_{ij}(t) & \text{otherwise.} \end{cases} \tag{3}$$

$$W^l_{ij}(t+\delta) = \begin{cases} W^l_{ij}(t) + \varepsilon\kappa[a_j(t) - \theta]a_i(t) & \text{if } a_i(t) > 0; \\ W^l_{ij}(t) & \text{otherwise.} \end{cases} \tag{4}$$

where: κ is the inverse of the number of units that are active in the layer the connection comes from (i.e. κ is $1/n_{a1}$, 1, and $1/n_{ph}$ for connections from context, item and input phoneme layers, respectively: n_{ph} being the number of phonemes in the current item learned); ε determines the rate of learning of long-term connections, and reduces as connections approach their maximum value ($\varepsilon = 0.1$ for item–phoneme and phoneme–item con-

Table A1
Summary of parameter values

	Value
Parameters taken from 1999 model	
Δ, decay rate for ST connection weights and for inhibitory input to item following output	0.75
σ, standard deviation of Gaussian noise during selection of item for output	0.3
Ω, inhibitory input to an item following its selection at output or presentation	−2.0
Size of LT phoneme–item & item–phoneme connection weights for familiar items	0.2
α, max size of a ST connection weight, with the exceptions below:	1
phoneme–item weights for auditory item followed by a pause	2
phoneme–item weights during retrieval under articulatory suppression, and during presentation of visual items	0
θ, presynaptic threshold for ST connection weight increment: phoneme–item	0.5
item–phoneme (zero for other connections, but see below)	0
n_{a1}, number of active units in the first component of a context set	6
n_{a2}, number of active units in the first component of a context set	2
New parameter values	
Cmatch, value below which a context set is rejected as a potential match	0.6
ε, size of increment to LT context-item connection weights	0.3
A, fixed total length of the vector of LT context-item connection weights	3
θ, presynaptic threshold for ST context-item connection weight increment	0.5
α, max size of a connection weight from the second component of context	2

nections for nonwords, but $\varepsilon = 0.0$ for item–phoneme and phoneme–item connections for familiar items such as letters of the alphabet—for which these long term connections have saturated at maximum values corresponding to a single increment with $\varepsilon = 0.2$); $\alpha = 1$ for all connections excepting for input phoneme to item connections for auditory items that are followed by a pause (giving an auditory recency effect at the end of lists or groups, articulatory suppression is modelled by setting $\alpha = 0$ for these weights). In the 1999 model, θ was zero for all connections excepting for input phoneme to item connections where $\theta = 0.5$ to increase the specificity of activation of item nodes (activating those that should be on and deactivating those that should be off). Here, we make $\theta = 0.5$ also for the connections from the context layer, for the same reason. In the 1999 model, ε was 0.1 for context–item connections, here we make $\varepsilon = 0.3$ to counteract the moderating effect of the normalisation of the context–item weights, described below.

In the present model the long-term connection weights from all context sets are normalised after the presentation and the recall of a list, such that:

$$\hat{W}^l_{ij} = \frac{AW^l_{ij}}{\sqrt{n\sum_j (W^l_{ij})^2}} \quad \text{for all } i, \tag{5}$$

where $n = 26$ is the number of item units. This ensures that the length of the vector of long-term connection weights from a context set to an item unit i (i.e. $\sqrt{\sum_j (W^l_{ij})^2}$) is fixed at $A/\sqrt{n}$, and the length of the entire vector of weights from a context set to the item layer is fixed at A. Thus the activation of a given item by a given context set only depends on the specificity of its pattern of connection weights, and not on their overall size (as in, e.g., 'competitive learning', Rumelhart & Zipser, 1985). These connection weights also differ from 1999 in having an initial overall length A, rather than zero. A value of $A = 3$ was found to produce a rough match to the data, and was used in all simulations.

B. Connections from the second component of a context set

Connections from the second component of a context set have unitary weights which do not decay with time. In the 1999 model, the input from these connections was assumed to modulate the entire input from the first set (carried by both long-term and short-term connection weights), although long-term connections were in fact zero as repeated temporally grouped lists were never simulated. In the current model, the input from the second component modulates the short-term input of the first component, but not its long-term input.[5] That is, the net input to item node i from the context set, at time t, is:

$$I_i = \sum_j W_{ij}(t)a_j(t) \times \sum_k V^s_{ik}(t)b_k(t) + \sum_k V^l_{ik}(t)b_k(t), \tag{6}$$

where W_{ij} and a_j denote weights and activations from the second component, and V^s_{ik}, V^l_{ik} and b_k denote the short-term weights, long-term weights and activations of the first component, respectively.

In the 1999 model, the weights from the second component of a context set (initially zero) were modified, like short-term weights, according to:

$$W_{ij}(t+\delta) = \begin{cases} \alpha\kappa[a_j(t) - \theta]a_i(t) & \text{if } a_i(t) > 0; \\ W_{ij}(t) & \text{otherwise.} \end{cases} \tag{7}$$

where: $\kappa = 1/n_{a2}$; $\alpha = 1$ and $\theta = 0$. That is, weights between active item and context nodes are set to $1/n_{a2}$, and did not decay. In the current model, these weights are modified using the same

[5] Otherwise the effect of temporal grouping would interact with Hebb repetition: getting stronger with each repeat, and this is not seen experimentally.

equation, but with $\alpha = 2$, i.e. weights between active item and context nodes are set to $2/n_{a2}$. Normalisation after the presentation and the recall of a list is used to reduce maximum weight to values $1/n_{a2}$, as in the previous model. Normalisation occurs such that:

$$\hat{W}_{ij} = \frac{W_{ij}}{n_{a2} \max(W_{ij})}, \tag{8}$$

which also reduces any connections strengthened during presentation of a previous (differently grouped) list: these will be halved in value each time the weights are normalised. This normalisation replaces the temporal decay in the previous model.

C. Calculating the match of a context set to a presented list

In the current model, multiple context sets are simulated: all of the context sets associated with previous lists, as well as potential 'new' context sets (which have un-modified long-term connection weights). As each item is presented, all context sets go through the same sequence of activation reflecting position in list (in the first component) and, in the case of temporal grouping, position within group (in the second, modulatory, component). A previous context set is recruited to control the recall of a new list on the basis of how well its long-term and modulatory connection weights match the sequence of items in the list. If several context sets show a good match, the best matching is recruited, if none does, then a new set is recruited.

The match score for each context set is calculated as follows. As an item is presented (item p say), the long-term input from the first component of context, modulated by the input from the second component, is calculated:

$$M_i = \sum_j W_{ij}(t)a_j(t) \times \sum_k V^l_{ik}(t)b_k(t), \tag{9}$$

where W_{ij} and a_j denote weights and activations from the second component, and V^l_{ik} and b_k denote the long-term weights and activations of the first component, respectively.[6] The context set gets a match score of 1 if this input is maximal for the item that is being presented, and a score of zero otherwise i.e., for presentation of item p:

$$\text{match}(p) = \begin{cases} 1 & \text{if } \max_i\{M_i\} = M_p; \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

As successive items $p_1, p_2, \ldots, p_n$ are presented, the cumulative or running average match score is kept:

$$\text{Cmatch}(n) = \frac{1}{n}\sum_{i=1}^{n} \text{match}(p_i). \tag{11}$$

If Cmatch falls below a threshold value after any number of items (n), the context set is discarded (i.e. rendered inactive). If more than one context set survives to the end of presentation, all but the one with the largest Cmatch value is discarded. If none survive then a new context set is recruited. Either way, only one context set will be active to control recall. A threshold value of 0.6 was found to produce a rough match to the data, and was used in all simulations.

[6] We note that it is not clear how M_i would be calculated by the network, since the actual input to item nodes is the product of the input from the second component and the *short-term* input from the first component, while M_i is the product of the input from the second component and the *long-term* input from the first component.

References

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language, 38*, 341–380.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.). *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89–195). New York: Academic Press.

Averbeck, B. B., Chafee, M. V., Crowe, D. A., & Georgopoulos, A. P. (2002). Parallel processing of serial movements in prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 99*, 13172–13177.

Avons, S. E. (1998). Serial report and item recognition of novel visual patterns. *British Journal of Psychology, 89*, 285–308.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences, 4*, 417–423.

Baddeley, A. D. (2006). *Working memory, thought and action*. Oxford: Oxford University Press.

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

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–89). New York, London: Academic Press.

Baddeley, A. D., Lewis, V. J., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology, 36*, 233–252.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575–589.

Bliss, T. V., & Collingridge, G. L. (1993). A synaptic model of memory: long-term potentiation in the hippocampus. *Nature, 361*, 31–39.

Botvinick, M., & Plaut, D. C. (2006). Short-term memory for serial order: a recurrent neural network model. *Psychological Review, 113*, 201–233.

Botvinick, M., & Huffstetler, S. (2006). Sequence learning in short-term memory: computational and empirical investigations of the Hebb effect. *Journal of Memory and Language*, Submitted for publication.

Bower, G. H., & Winzenz, D. (1969). Group structure, coding, and memory for digit series. *Journal of Experimental Psychology Monograph, 80*, 1–17.

Neath, I., & Brown, G. D. A. (2006). SIMPLE: Further applications of a local distinctiveness model of memory. *Psychology of Learning and Motivation: Advances in Research and Theory, 46*, 201–243.

Brown, G. D., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107*, 127–181.

Please cite this article as: Neil Burgess, Graham J. Hitch, A revised model of short-term memory and long-term ..., Journal of Memory and Language (2006), doi:10.1016/j.jml.2006.08.005

Bullock, D. (2004). Adaptive neural models of queuing and timing in fluent action. *Trends in Cognitive Sciences, 8*, 426–433.

Burgess, N. (1995). A solvable connectionist model of immediate recall of ordered lists. In G. Tesauro, D. Touretzky, & T. Leen (Eds.). *Neural information processing systems* (7, pp. 51–58). Cambridge, MA: MIT Press.

Burgess, N., & Hitch, G. J. (1992). Toward a network model of the articulatory loop. *Journal of Memory and Language, 31*, 429–460.

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. (2005). Models of working memory: putting long term memory into context. *Trends in Cognitive Science, 9*, 535–541.

Carpenter, G. A., & Grossberg, S. (2003). Adaptive resonance theory. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (2nd ed., pp. 87–90). Cambridge, MA: MIT Press.

Cohen, R. L., & Johansson, B. S. (1967). Some relevant factors in the transfer of material from short-term to long-term memory. *The Quarterly Journal of Experimental Psychology, 19*, 300–308.

Conrad, R. (1960). Serial order intrusions in immediate memory. *British Journal of Psychology, 51*, 45–48.

Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior, 4*, 161–169.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology, 55*, 429–432.

Cowan, N. (2001). An embedded-process model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge: Cambridge University Press.

Cumming, N., Page, M. P. A., & Norris, D. (2003). Testing a positional model of the Hebb effect. *Memory, 11*, 43–63.

Cunningham, T. F., Healy, A. F., & Williams, D. M. (1984). Effects of repetition on short term retention of order information. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10*, 575–597.

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: empirical and computational investigations of recency effects. *Psychological Review, 112*, 3–42.

Dosher, B. A. (1999). Item interference and time delays in working memory: immediate serial recall. *International Journal of Psychology, 34*, 276–284.

Duncan, M., & Lewandowsky, S. (2005). The time course of response suppression: no evidence for gradual release from inhibition. *Memory, 13*, 236–246.

Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: insights from connectionist networks. *Journal of Experimental Psychology: Learning, memory, and cognition, 26*, 1103–1123.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press/Bradford Books.

Estes, W. K. (1972). An associative basis for coding and organisation in memory. In A. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 161–190). Washington, DC: Winston and Sons.

Estes, W. K., & Maddox, W. T. (1995). Interactions of similarity, base rate, and feedback in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1075–1095.

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9*, 59–79.

Fastame, M. C., Hitch G. J., & Vecchi T., (2006). Is Immediate serial order recall driven by modality-specific mechanisms? *Poster presented at 3rd European Working Memory Symposium*, Working memory, long-term memory and attention, Genoa.

Frankish, C. (1985). Modality-specific grouping effects in short-term memory. *Journal of memory and language, 24*, 200–209.

Gagnon, S., Foster, J. K., Turcotte, J., & Jongenelis, S. (2004). Involvement of the hippocampus in implicit learning of supra-span sequences: the case of SJ. *Cognitive Neuropsychology, 21*, 867–882.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition, 23*, 83–94.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding, I: parallel development and coding of neural feature detectors and II: feedback, expectation, olfaction, and illusions. *Biological Cybernetics, 23*, 121–134, 187–202.

Grossberg, S. (1978). Behavioral contrast in short-term memory: serial binary memory models or parallel continuous models? *Journal of Mathematical Psychology, 3*, 199–219.

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: computational and neural bases. *Brain and Language, 59*, 267–333.

Hartley, T. (2002). Syllabic phase: a bottom-up representation of the temporal structure of speech. In J. A. Bullinaria & W. Lowe (Eds.), *Connectionist models of cognition and perception: Proceedings of the seventh neural computation and psychology workshop*. Singapore: World Scientific.

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language, 35*, 1–31.

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

Henson, R. N. A. (1998). Short-term memory for serial order: the start-end model. *Cognitive Psychology, 36*, 73–137.

Henson, R. N. A. (1999). Positional information in short-term memory: relative or absolute? *Memory & Cognition, 27*, 915–927.

Henson, R. N. A., & Burgess, N. (1997). Representations of serial order. In Bullinaria, J. A., Glasspool, D. W., & Houghton, G. (Eds.), 4th Neural Computation and Psychology Workshop, London 9–11 April 1997: Connectionist representations (pp. 283–300). New York: Springer.

Hitch, G. J., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal grouping effects in immediate recall; a working memory analysis. *Quarterly Journal of Experimental Psychology, 49A*, 116–139.

Hitch, G. J., Fastame, C., & Flude, B. (2005). How is the serial order of a verbal sequence coded? Some comparisons between models. *Memory, 3–4*, 247–258.

Hitch, G. J., Flude, B. & Burgess, N. (2006) Slave to the rhythm: dissociation of temporal versus phonological factors in long-term learning of verbal sequences. *Journal of Memory and Language*. Under review.

Houghton, G. (1990). The problem of serial order: a neural network model of sequence learning and recall. In R. Dale, C. Mellish, & M. Zock (Eds.), *Current research in natural language generation* (pp. 287–319). London: Academic Press.

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

James, W. (1890). *Principles of psychology*. New York: Holt.

Jones, D. M., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 1008–1018.

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.

Jones, T., & Polk, T. A. (2002). An attractor network model of serial recall. *Cognitive Systems Research, 3*, 45–55.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 2*, 153–161.

Lee, C. L., & Estes, W. K. (1981). Item and order information in short-term memory: evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 149–169.

Lewandowsky, S., & Brown, G. D. A. (2005). Serial recall and presentation schedule: a microanalysis of local distinctiveness. *Memory, 13*, 283–292.

Lewandowsky, S., Brown, G. D. A., Wright, T., & Nimmo, L. M. (2006). Timeless memory: evidence against temporal distinctiveness models of short-term memory for serial order. *Journal of Memory and Language, 54*, 20–38.

Lewandowsky, S., Duncan, M., & Brown, G. D. A. (2004). Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin & Review, 11*, 771–790.

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review, 96*, 25–57.

Lovatt, P. J., Avons, S. E., & Masterson, J. (2000). The word-length effect and disyllabic words. *Quarterly Journal of Experimental Psychology, 53A*, 1–22.

Lustig, C., & Hasher, L. (2002). Working memory span: the effect of prior learning. *The American Journal of Psychology, 115*, 89–101.

McKelvie, J. S. (1987). Learning and awareness in the Hebb digits task. *Journal of General Psychology, 114*, 75–88.

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

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior, 2*, 1–21.

Mensink, G. J., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review, 95*, 434–455.

Mueller, S. T., Seymour, T. L., Kieras, D. E., & Meyer, D. E. (2003). Theoretical implications of articulatory duration, phonological similarity, and phonological complexity in verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1353–1380.

Murdock, B. B. (1960). The distinctiveness of stimuli. *Psychological Review, 67*, 16–31.

Murdock, B. B. (1995). Developing TODAM: three models for serial-order information. *Memory & Cognition, 23*, 631–645.

Murray, D. J. (1967). The role of speech responses in short-term memory. *Canadian Journal of Psychology, 21*, 263–276.

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

Nairne, J. S. (2002). Remembering over the short-term: the case against the standard model. *Annual Review of Psychology, 53*, 53–81.

Neath, I. (1993). Contextual and distinctive processes and the serial position function. *Journal of Memory and Language, 32*, 820–840.

Ng, H. L. H., & Maybery, M. T. (2002). Grouping in short-term verbal memory: is position coded temporally? *Quarterly Journal of Experimental Psychology, 55A*, 391–424.

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

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134.

Ranganath, C., & Blumenfeld, R. S. (2005). Doubts about double dissociations between short- and long-term memory. *Trends in Cognitive Sciences, 9*, 374–380.

Rumelhart, D. E., & Zipser, D. (1985). Feature discovery by competitive learning. *Cognitive Science, 9*, 75–112.

Ryan, J. (1969). Grouping and short-term memory: different means and pattern of grouping. *Quarterly Journal of Experimental Psychology, 21*, 137–147.

Sanders, A. F., & Willemsen, E. M. (1978). Proactive interference in immediate serial recall. *Acta Psychologica, 42*, 29–38.

Schwartz, M., & Bryden, M. P. (1971). Coding factors in the learning of repeated digit sequences. *Journal of Experimental Psychology, 87*, 331–334.

Sechler, E. S., & Watkins, M. J. (1991). Learning to reproduce a list and memory for the learning. *American Journal of Psychology, 104*, 367–394.

Service, E. (1998). The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *Quarterly Journal of Experimental Psychology, 51A*, 283–304.

Smyth, M. M., Hay, D. C., Hitch, G. J., & Horton, N. J. (2005). Serial position memory in the visual–spatial domain: reconstructing sequences of unfamiliar faces. *Quarterly Journal of Experimental Psychology, 58A*, 909–930.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64*, 49–60.

Zucker, R. S., & Regehr, W. G. (2002). Short-term synaptic plasticity. *Annual Review of Physiology, 64*, 355–405.