Review TRENDS in Cognitive Sciences Vol.9 No.11 November 2005

Computational models of working memory: putting long-term memory into context

Neil Burgess[1] and Graham Hitch[2]

[1]Institute of Cognitive Neuroscience and Department of Anatomy and Developmental Biology, University College London, 17 Queen Square, London, WC1N 3AR, UK
[2]Department of Psychology, York University, York, YO10 5DD, UK

Detailed computational modeling of human memory has typically been aimed at either short-term (working) memory or long-term memory in isolation. However, recent research highlights the importance of interactions between these systems for both item and order information. At the same time, computational models of both systems are beginning to converge onto a common framework in which items are associated with an evolving 'context' signal and subsequently compete with one another at recall. We review some of these models, and discuss a common mechanism capable of modelling working memory and its interaction with long-term memory, focussing on memory for verbal sequences.

Introduction

The idea of short-term memory (STM) and its relationship with long-term memory (LTM) has long intrigued psychologists. Much of the short-term organization of behaviour requires some limited memory capacity to support on-line processing, as in the concept of 'working memory' [1]. Equally, there is a need to filter on-going experience and buffer this information so as to organize its efficient entry into long-term memory [2–4]. The distinction between STM and LTM is supported by many sources of evidence. Behaviourally, immediate serial recall (ISR) of a list of words is sensitive to their phonological similarity, but not their semantic similarity, whereas the reverse is true for memory after a delay of more than a few seconds [5]. Neuropsychologically, there is a double dissociation between organic amnesia associated with the hippocampus and related systems, and short-term buffer disorders associated with neocortical damage: amnesic patients present with impaired LTM and preserved STM [6,7], whereas short-term buffer disorders show the reverse pattern [8]. Physiologically, STM is often thought of as maintained neuronal firing [9,10], or short-term potentiation of synaptic connections [11], whereas LTM is thought of as long-term potentiation of synapses [9]. However, the dissociation between STM and LTM in no way denies the link between them: each is clearly important for the other and how they interact is a topical question in memory research [12,13].

One of the clearest examples of the dissociation between STM and LTM comes from the Hebb repetition effect, in which repeated presentation of a digit sequence for ISR results in gradual learning of that sequence, even though presentation and recall of other sequences intervenes between the repetitions [14]. Hebb concluded that there must be some change in LTM from a single trial. Milner and colleagues subsequently found that the Hebb repetition effect for lists of digits or spatial locations depended on the integrity of the medial temporal lobe (more specifically, on the left and right hippocampi, respectively), whereas ISR of the lists did not [15]. We note, however, that the relationship between STM and LTM is stimulus specific. Other stimuli, such as the layout of a spatial scene [16,17], appear to require the hippocampus even over very short delays [13].

In this article, we review a combination of recent computational and empirical investigations of verbal memory that are beginning to address the interrelationship of STM and LTM. We outline recent progress in modeling verbal STM in terms of separate mechanisms for order and item information, where order is controlled by some form of context signal, and argue that independently developed models of LTM are converging on a similar framework. Finally, we suggest that the Hebb repetition effect provides a powerful vehicle for developing and testing models of the relationship between STM and LTM.

Computational modeling of verbal STM

The multi-component model of working memory proposed by Baddeley and Hitch [1] and developed by Baddeley [5] is described in Box 1. The phonological loop component gives a simple account of the sensitivity of ISR of verbal items to the items' spoken characteristics and interference from concurrent articulatory suppression. However, the phonological loop requires translation into a quantitative lower-level model to address the actual mechanism by which serial order is retained.

Serial order: context signals and competitive queuing

The simplest mechanism for storing a sequence of items in order is to associate one directly to the next ('chaining').

Corresponding author: Burgess, N. (n.burgess@ucl.ac.uk).
Available online 5 October 2005

www.sciencedirect.com 1364-6613/$ - see front matter doi:10.1016/j.tics.2005.09.011

Box 1. The Baddeley and Hitch (1974) model of working memory

Three separate but interacting limited capacity components of working memory were assumed: a central executive responsible for control processes and two slave systems providing modality-specific buffer storage (see Figure I). One buffer store is specialized for visuo-spatial information, the other for verbal information. All three subsystems are separate from LTM. The verbal buffer acts as a 'phonological loop' supporting immediate serial recall or rehearsal, such as when remembering a telephone number for long enough to be able to dial it. An early metaphor for this buffer was that of a closed loop tape recorder in which stored speech sounds undergo rapid decay and can be refreshed by a control process of subvocal rehearsal. This account explains the poorer recall of sequences of similar sounding items (the 'phonemic similarity effect'), the relationship between memory span for words and how fast the words can be subvocally rehearsed (the 'word length effect'), and the abolition of this word length effect when subvocal rehearsal is prevented by concurrent 'articulatory suppression'. Subvocalization was also assumed to be required to recode visual stimuli into verbal form, so as to enter the phonological loop, explaining why articulatory suppression removes the phonemic similarity effect for visual but not auditory items (see [5] for details).

Growing evidence of additional links between working memory and LTM, for example, the improvement in ISR associated with chunking based on previously acquired knowledge, recently caused Baddeley [12] to suggest the addition of an 'episodic buffer' to the model.

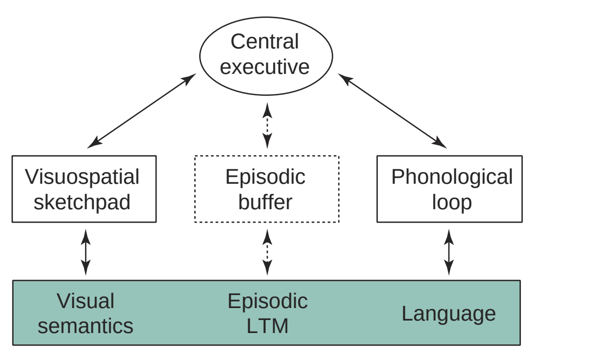

Figure I. The elements of working memory originally proposed by Baddeley and Hitch [1] and developed later by Baddeley [5] are shown (white boxes), together with their interactions with long-term memory (green box), taken from [12]. Dashed lines show the additional 'episodic buffer' more recently suggested by Baddeley [12].

However, although chaining works well for a wide range of memory phenomena [18,19], it is inconsistent with the pattern of errors in ISR [20,21]. For example, in a chaining model, once an item is omitted, it is unlikely to be retrieved until the end of the list, if at all, whereas order errors involving transposition of nearby items are very common in ISR [22–24]. Even more strikingly, in ISR of lists of items of alternating phonemic similarity (e.g. Q-D-R-B-N-P), worse recall of the similar sounding items (D, B, P) does not affect performance on the dissimilar intervening ones [23,25].

Most alternatives to chaining involve associating item information to some kind of 'context' signal that varies with experience or time, coupled with a mechanism for selecting one item at a time during retrieval. One of the earliest, the 'perturbation' model of Estes [21], assumes a direct encoding of the position of each item that is subject to perturbation (changes by 1 position) over subsequent time steps. In more recent models a context representation evolves with the passage of items [11,22,26–28] but is not itself perturbed. Rather, errors occur in the way in which item information is retrieved from context. Various forms of context signal have been proposed, such as a stochastic variation in time [9,22], a moving window of activation [11] or a decaying 'start node' and increasing 'end node' [26,27].

The majority of models [19,22,24,29,30] also converge on a type of selection mechanism known as 'competitive queuing' ([26,31], and reviewed in [32]). In competitive queuing, items are active in parallel and the most active is output; this item is then inhibited to allow the next most active to be selected, and subsequently recovers from inhibition. Such a mechanism generates patterns of transposition errors typical of ISR when noise is introduced into item activations, and naturally incorporates response suppression. It is strikingly consistent with single unit responses in primate prefrontal cortex (where cells representing different forthcoming actions are active in parallel, the most active corresponding to the soonest-to-be-performed; see [32]).

An insight into the nature of the context signal comes from 'intrusion errors' in ISR: items from a previously presented list that are erroneously retrieved in the current list. Intrusion errors tend to occur at a similar position in the current list to their position in the previous list [33]. Experiments using lists of variable length [34] show further that intrusions tend to maintain their position relative to both the start and the end of the list. Thus, if order information is maintained by association of items to a context signal, this signal must reflect position-within-list relative to both ends, as in the start–end models [26,27] and an oscillator implementation of positional context [35] inspired by a model of syllabic parsing [36], see also [28]. Neurophysiological evidence of positional coding has been seen in primate supplementary motor cortices (e.g. cells responding between the second and third action in a sequence, irrespective of the identity of the actions [37]).

One successful model of ISR [24] does not include positional cues. In this model an item representation is activated at presentation, increases in activation with each new item that is presented, and then decays with time during retrieval. It can be regarded as using a context signal consisting of a single node to which items are associated at presentation by connections that increase in strength with each new item presented. This model (as with [11] with context signal removed, see Box 2), demonstrates adequate ISR, including effects of phonemic similarity, articulatory suppression etc. However, non-positional accounts cannot explain typical patterns of intrusion errors in ISR or temporal grouping effects (see below).

Regardless of how they deal with serial order, most models of verbal STM require that output involves accessing items' phonological representations. Effects of noise at this stage generate phonological similarity effects, and decay in the strength of association between items and context with increasing duration of rehearsal or recall

Box 2. Modeling verbal working memory with a context signal and competitive queuing

We describe the operation of one model of ISR (Figure II; see [11] for details, also [22,24,27,28,30]). Item nodes are connected to their corresponding input and output speech representations (modelled as phonemes for simplicity), and to a context/timing signal, via connections capable of large short-term (decaying) modification and smaller (incremental) long-term modification. Corresponding input and output phonemes directly activate each other. Items are selected by 'competitive queuing' at each step of presentation and retrieval; that is, the most strongly activated item is selected, subsequently inhibited, and then slowly recovers from inhibition.

Presentation of an auditory item activates the corresponding input phonemes and thence output phonemes and a single item node (selected by competitive queuing), whilst the context signal changes by a fixed amount, and connection weights are modified. Changes to little-used connections (e.g. context-item connections for a novel list) are dominated by short-term modification, whereas well-used connections (e.g. item–phoneme connections for familiar words) also have a significant long-term component.

At retrieval, the context signal is reset, and reproduces the changes made during presentation: changing by a fixed amount with each recalled item. Item nodes are re-activated via their context-item connections, the most active is selected and activates its output phonemes, whose activation feeds back to item nodes via the input phonemes. Output corresponds to a final selection by competitive queuing, while item nodes receive both phonemic and contextual input. This stage is assumed to be noisy: causing errors in which as-yet-unrecalled items associated to similar context or phoneme representations may replace the target item. See [24] for further discussion of the need for a second (phonological) output stage.

Many aspects of the model have an obvious correspondence to the concept of a phonological loop. Rehearsal (simulated as repeated retrievals) serves to refresh the decaying short-term connection weights. The activation of the input phonemes by the output phonemes during rehearsal or visual input corresponds to hearing one's inner voice. The structure of the model also captures the patterns of impairment due to articulatory suppression (modelled as disrupting the output phoneme representation and hence also the input phoneme representation) and observed in the various types of short-term memory patient.

The inclusion of a context signal, separate from phonological representations, predicts a dissociation between effects of temporal grouping and Hebb repetition on the one hand, and on the other, effects traditionally associated with the phonological loop, such as phonological similarity, word length and articulatory suppression. Indeed this is almost a double dissociation, as ISR itself is only weakly dependent on the context signal - the recovery of items from inhibition during presentation alone allows them to be selected in order at retrieval with only a slight drop in span [11]. In addition, the inclusion of long-term as well as short-term learning allows effects of item and list familiarity to be addressed.

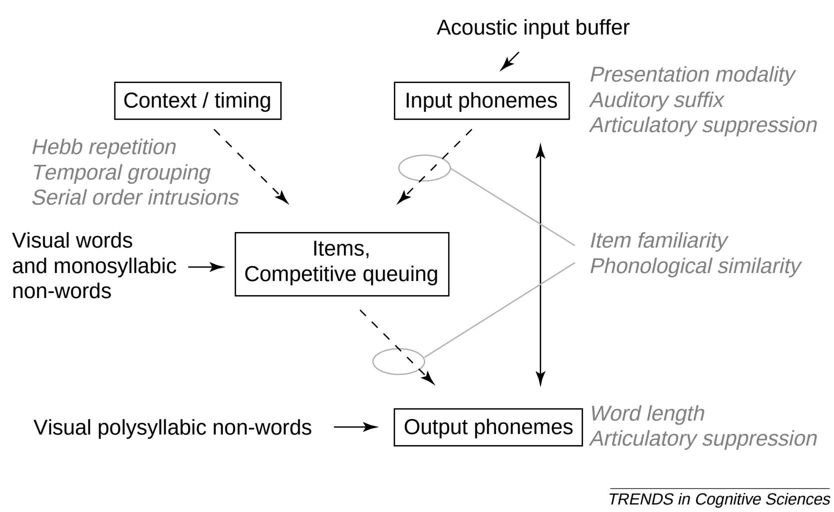

Figure II. Outline model of the phonological loop. Components of the model are shown in black boxes. Dotted arrows show short- and long-term modifiable connections, full arrows show inputs to the model and the fixed one-to-one connections between input and output phonology. Grey text indicates some of the experimental effects captured by the model next to the relevant part of the model. Adapted from [11].

generates word length effects. Box 2 gives further details for one of these models [11].

In summary, models of verbal STM converge on separate mechanisms for serial ordering and retrieving the phonological characteristics of items. Although competitive queuing alone is sufficient to generate serially ordered output, intrusion errors (and temporal grouping; see below) suggest that this process is cued by some form of context signal containing positional information.

Serial order and rhythm

One important influence on ISR that appears to be independent of articulatory or phonological factors is provided by the temporal rhythm with which verbal items are presented. Performance is enhanced by presenting items with a specific temporal grouping, as opposed to evenly-spaced presentation [38]. This effect is stronger for auditory than visual presentation, is characterized by mini U-shaped serial position curves for each group, and by a decreased incidence of order errors between items that occupy different within-group positions [38,39]. When the number of items or duration of each group is varied, the patterns of order errors indicate that, as for within-list position, within-group position is encoded relative to both the start and the end of the group ([34,40]; see [35] for a model).

The effects of temporal rhythm on ISR are independent of items' phonemic similarity and word length [41], consistent with the separation between item (phonological) and order (contextual) components. They can be readily accommodated in models that associate items to context by making the context signal multi-dimensional – one dimension tracking position within list and another tracking position within group [11,41]. The second dimension reduces the overall similarity between the context representations for all pairs of items excepting those with the same within-group position. We argue below that grouping is important not only for specifying the nature of the context signal in STM, but in identifying the context signal as a crucial link in STM–LTM interactions.

Context and models of long-term memory

In parallel to the working memory models reviewed above, several models of LTM propose that the memory trace of an item's presentation includes a time-tag for use in subsequent recall [42–44]. Thus these models effectively also use item–context associations, but often with a different context signal for each list [9,45–47]. These models usually focus on free recall of supra-span lists of items. We note that the use of a context representation at the higher level of lists is consistent with the lower level use of context states that vary within lists or sub-groups of a list discussed above. Indeed these ideas have also been used at the syllable level [29], and some models attempt to capture effects of learning at syllable and item levels by combining them hierarchically [30].

Another approach to modelling both LTM and STM is provided by 'temporal distinctiveness' theories. A recency effect corresponding to better recall of the most recently presented items is found in both immediate free recall (where it is usually regarded as reflecting STM) and under 'continuous distractor' conditions (in which each item is followed by a distractor task sufficient to wipe out the contents of STM). These models assume that the probability of correctly remembering an item relates to the distinctiveness of its presentation time from the presentation times of other items. Again, they can be implemented in terms of association to a slowly time-varying context [44,46,48–50]. Distinctiveness models capture a significant continuity between STM and LTM but at the same time tend to ignore the many dissociations between these two systems. Our argument here is that we need an account that captures both sets of features.

One potentially important difference between LTM and STM concerns item-to-item ('chaining') associations. In contrast to ISR, there are signs that item-to-item association does play a role in LTM, such as the forward bias in free recall (one item is more likely to lead to the next item than the preceding one). This observation led to the 'temporal context model' (TCM) [49]. Again, item representations are associated with context representations such that a given item will be retrieved according to the similarity between the context at retrieval and that associated with the item. However, unlike the STM models above, in which context representations evolve independently, the context representation is derived from the items themselves: becoming a recency-weighted sum of the context arising from each item (see also [19]). After presentation of the list the context vector will be most similar to the most recent items, producing the well-known recency effect. In addition, when an item is presented it affects the context vector to which subsequent items are associated. Thus the recall of a given item increases the likelihood of retrieval of immediately subsequent items in the list (by making the context vector more similar to its state when those items were presented), causing the observed forward bias in retrieval.

An alternative to the TCM model [9] simply assumes a linear set of context units on which the location of the single active unit performs a random walk biased in one direction. Again, items are retrieved via associations to the context units, whereas the context signal continues to walk from its final state (producing long-term recency) and then from its initial state (retrieving items from the start with a forward bias). Partial separation of the mechanisms behind STM and LTM phenomena, as in [9], is advizable given the many qualitative differences between them.

In summary, models of LTM typically involve forming associations between items and their context, suggesting a point of correspondence with models of STM. However, distinctiveness models tend to overplay the degree of similarity. In our view the different nature of the context representations in models of LTM compared with models of STM (e.g. the inclusion of item information in the TCM) implies that a transition occurs with repeated experience of lists or with increasing list-length. The nature of this transition is likely to be an important factor in the passage from STM to LTM and may correspond to 'chunking', in which short-term representations become permanently grouped together.

Interactions between working memory and LTM

Although ISR is often discussed in terms of the phonological loop alone, it shows strong influences of LTM. For example, familiar words are better recalled than unfamiliar words [51], and invented nonwords are harder to recall than words [52]. These and other influences of language knowledge on ISR exemplify the link from LTM to the phonological loop in Box 1, and provide strong constraints for modelling. For instance, unlike ISR, immediate serial recognition shows virtually no effect of lexicality [53], indicating that item and order information have separate links to long-term memory.

The reverse link, from working memory to LTM, is most obvious in the role of the phonological loop in long-term learning of novel word forms, reviewed in [54]. Thus, individual differences in verbal ISR in children predict their scores on vocabulary tests, and the phonological short-term memory patient PV was unable to learn non-words despite showing normal learning in a corresponding task involving words. Convergently, normal adults' learning of novel word forms is much more sensitive to word length, phonemic similarity and articulatory suppression than their performance on corresponding tasks involving words. A second role for the phonological loop in long-term learning concerns the order of familiar items, for example, in learning a nursery rhyme or telephone number through repeated exposure, studied experimentally as the Hebb repetition effect described earlier. An important result here is that the rate of learning in Hebb repetition is sensitive to manipulations of temporal grouping, but not articulatory suppression or phonological similarity (which we noted

Box 3. Modeling Hebb repetition and temporal grouping with context/timing signals

The Hebb repetition effect tells us about the nature of the interaction between STM and LTM in ISR. Because the effect is observed with familiar items (e.g. digits), learning must concern the serial order of the repeated list. Further, the effect is weaker when each presentation of the repeated list has a different rhythm [59], and, for familiar items, is unaffected by their phonemic similarity or by articulatory suppression (Hitch, Flude and Burgess, unpublished data), even though both factors disrupt ISR. This pattern is consistent with Hebb repetition occurring primarily through the long-term context-item connection weights in a model like that in Box 2 (in which the context signal is sensitive to rhythmic grouping). However, the long-term learning of different lists would interfere with each other (see [66]) unless a specific context signal is developed for each one as it becomes familiar. To do this we modified the model (Box 2) to include multiple context sets. For all potential context sets, the match between the long-term context-item connections and the presented item is calculated for successive items in the list. After each item, context sets whose running average match falls below a threshold value are discarded. The best-matching surviving context set remains to control recall, and if none survive a new set is recruited. Modification of the long-term connection weights of surviving context sets during presentation and recall causes them to become even better matches. This model successfully captures the pattern of data for Hebb repetition (Hitch, Flude and Burgess, unpublished data), and predicts some new effects, such as the capture of a context set by repeated sequences whose ending slowly changes.

above do affect the learning of novel word forms) (see Box 3). Again, this is consistent with the idea that item and order information are not only distinct within STM but also have separate links to LTM (but see [55] for a model of ISR which does not explicitly make this separation, but nonetheless captures effects of long-term learning such as the bigram frequency effect).

In summary, any account of the interaction between STM and LTM should reflect the apparent dissociations between learning mechanisms for familiar and unfamiliar verbal materials, and between item and order information. The acknowledgement of multiple independent links between STM and LTM for different kinds of information is reflected in Baddeley's addition of a multi-modal 'episodic buffer' to the working memory model ([12], and Box 1). We tentatively interpret the link between language knowledge and the phonological loop as involving item information, and that between the episodic buffer and LTM as involving order information in the form of a context signal. We note that serial order effects in STM for non-verbal sequences resemble those in corresponding verbal tasks [56], consistent with a multi-modal capacity for serial ordering. However, we shall see that attempting to model the interaction between STM and LTM in more detail raises questions for any simplistic distinction between order and item information.

Modelling the interaction between WM and LTM

The different timescales of synaptic potentiation, including 'long-term potentiation', which seems to last for as long as can be measured [57], and 'post-tetanic potentiation', which lasts for a few seconds [58], suggests modelling STM and LTM via connections with both short- and long-term plasticity [11]. Thus, whereas short-term (decaying) connection weights mediate the association between item nodes and context in ISR of a novel sequence, the Hebb repetition effect results from a longer-lasting cumulative association, and residual associations from previous lists produce position-specific intrusions in ISR. This kind of architecture correctly predicts patterns of interaction between experimental manipulations in the Hebb repetition task (Box 2). Similarly, the connection strengths between item nodes and input/output phonology would have a larger long-term component for words than non-words, explaining effects of item familiarity in ISR [11].

However, to properly model the learning of phonology, one should include its timing, effectively using competitive queuing at both list [22] and item [29] timescales (see also [30]). Similarly, to model the learning of lists properly requires models of working memory to have multiple context signals. If so, an existing context signal with long-term connections that match a presented list would be able to control retrieval of the list and, over repetitions, become better matched to it – effectively storing the order information it represents (see Box 3). The mechanism for finding the best-matching context signal during sequential presentation of a list highlights the importance of the familiarity of the start of the list for Hebb learning to occur [59,60]: an interesting parallel with the 'cohort' model of spoken word recognition [61] and the learning of new categories in Adaptive Resonance Theory (reviewed in [62]). As noted earlier, a more realistic computational model may be possible in which a hierarchy of context signals deals with ordering at different levels of representation, from phonemes through words to groups or lists. In such models, the matching process should find the highest level in the hierarchy at which a good correspondence can be found, as in the Pandemonium account of perceptual recognition [63], thus providing a potential account of chunking.

Conclusion

The interaction between working memory and LTM is a topic of much current interest, and computational models will be required for a quantitative understanding to emerge. Here, we have discussed the mechanisms relating STM and LTM for verbal items (see [64,65] for examples of similar approaches to spatial memory). We have shown how models of working memory that involve short- and long-term plasticity can explain some of the effects of STM on long-term learning and of LTM on immediate recall. We have argued that models of STM and LTM might fit within a common framework whereby associations are formed between states of a context signal and representations of items. This suggestion raises many questions (see Box 4). Within these models, hierarchical use of context signals over different timescales potentially enables modelling of both immediate memory and long-term learning at the levels of syllables, items, chunks and lists, including the learned familiarity of words and of oft-repeated sequences of words.

Box 4. Questions for future research

- How should temporal or positional accounts of context in STM be combined with accounts of LTM in which context depends on the items themselves (as in the TCM [49])? For example, is the rate of change or distinctiveness of the various sources of context important, such that, for very long lists, context based on the rhythm of presentation alone is not distinctive enough – requiring item information to be included.
- Does chaining begin to occur as lists are repeated in STM experiments, indicating a shift to a TCM-like [49] use of context (there are hints of this in [60])?
- Which aspects of the context signal come from the medial temporal lobe? The Hebb repetition effect (but not ISR) seems to depend on the hippocampus [15], provision of a context signal is one way to interpret the hippocampus as providing a dynamic index into LTM [4,67,68] and the physiology of the hippocampal formation may support slowly-varying patterns of activity reflecting temporal [65,69,70] or spatial context [65,71].
- Does the additional implication of the context signal in short-term serial order and temporal grouping effects relate to recent findings linking the medial temporal lobe to some aspects of sequential processing [72]?
- How do the oscillatory context signal/competitive queuing models [11,28,29,35,36] relate to the temporal mechanisms proposed for working memory [73] and long-term encoding of sequential information in medial temporal lobe involving the theta and gamma rhythms of the EEG [74]?
- Do the separate mechanisms for long-term learning of item and order information correspond respectively to the separate links with LTM of the phonological loop and the episodic buffer (see Box 1)?

Acknowledgements
We thank Alan Baddeley, Tim Shallice and Tom Hartley for useful discussions. This work was supported by the MRC, UK.

References

1 Baddeley, A.D. and Hitch, G.J. (1974) Working memory. In *The Psychology of Learning and Motivation* (Bower, G.H., ed.), pp. 47–89, Academic Press
2 Atkinson, R.C. and Shiffrin, R.M. (1971) The control of short-term memory. *Sci. Am.* 225, 82–90
3 Waugh, N. and Norman, D.A. (1965) Primary memory. *Psychol. Rev.* 72, 89–104
4 Marr, D. (1971) Simple memory: a theory for archicortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 262, 23–81
5 Baddeley, A.D. (1986) *Working Memory*, Oxford University Press
6 Baddeley, A.D. and Warrington, E.K. (1970) Amnesia and the distinction between long-and short-term memory. *Journal of Verbal. Learning and Verbal. Behavior.* 9, 176–189
7 Drachman, D.A. and Arbit, T. (1966) Memory and the hippocampal complex: II. Is memory a multiple process? *Arch. Neurol.* 15, 52–61
8 Shallice, T. (1988) *From Neuropsychology to Mental Structure*, Cambridge University Press
9 Davelaar, E.J. *et al*. (2005) A context activation model of list memory: dissociating short-term from long-term recency effects. *Psychol. Rev.* 112, 34–53
10 O'Reilly, R.C. *et al.* (1998) A hippocampal model of recognition memory. In *Neural Information Processing Systems*, Vol. 10 (Jordan, M.I. *et al.*, eds), pp. 73–79, MIT Press
11 Burgess, N. and Hitch, G.J. (1999) Memory for serial order: a network model of the phonological loop and its timing. *Psychol. Rev.* 106, 551–581
12 Baddeley, A. (2000) The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4, 417–423
13 Ranganath, C. and Blumenfeld, R.S. (2005) Doubts about double dissociations between short- and long-term memory. *Trends Cogn. Sci.* 9, 374–380
14 Hebb, D.O. (1961) Distinctive features of learning in the higher animal. In *Brain Mechanisms and Learning* (Delafresnaye, J.E., ed.), pp. 37–46, Oxford University Press
15 Milner, B. (1971) Interhemispheric differences in the localization of psychological processes in man. *Br. Med. Bull.* 27, 272–277
16 Hartley, T. *et al.* (2005) The 4 mountains test: hippocampal damage selectively impairs topographical processing of spatial scenes over very short delays. *Soc. Neurosci. Abstr.* 191.14
17 Lee, A.C. *et al.* Specialization in the medial temporal lobe for processing of objects and scenes. *Hippocampus* (in press)
18 Kleinfeld, D. and Sompolinsky, H. (1988) Associative neural network model for the generation of temporal patterns. Theory and application to central pattern generators. *Biophys J.* 54, 1039–1051
19 Lewandowski, S. and Murdock, B.B. (1989) Memory for serial order. *Psychol. Rev.* 96, 25–57
20 Lashley, K.S. (1951) The problem of serial order in behavior. In *Cerebral Mechanisms in Behavior* (Jeffress, L.A., ed.), pp. 112–136, Wiley
21 Estes, W.K. (1972) An associative basis for coding and organization in memory. In *Coding processes in human memory* (Melton, A.W. and Martin, E., eds), pp. 161–190, V.H.Winston & Sons
22 Burgess, N. and Hitch, G.J. (1992) Toward a network model of the articulatory loop. *J. Mem. Lang.* 31, 429–460
23 Henson, R.N. (1996) Unchained memory: error patterns rule out chaining models of immediate serial recall. *Q. J. Exp. Psychol.* 49, 80–115
24 Page, M.P. and Norris, D. (1998) The primacy model: a new model of immediate serial recall. *Psychol. Rev.* 105, 761–781
25 Baddeley, A.D. (1968) How does acoustic similarity influence short-term memory? *Q. J. Exp. Psychol.* 20, 249–263
26 Houghton, G. (1990) The problem of serial order: a neural network model of sequence learning and recall. In *Current Research in Natural Language Generation* (Dale, R. *et al*., eds), pp. 287–319, Academic Press
27 Henson, R.N. (1998) Short-term memory for serial order: the Start–End Model. *Cogn. Psychol.* 36, 73–137
28 Brown, G.D. *et al*. (2000) Oscillator-based memory for serial order. *Psychol. Rev.* 107, 127–181
29 Hartley, T. and Houghton, G. (1996) A linguistically constrained model of short-term memory. *J. Mem. Lang.* 35, 1–31
30 Gupta, P. and MacWhinney, B. (1997) Vocabulary acquisition and verbal short-term memory: computational and neural bases. *Brain Lang.* 59, 267–333
31 Grossberg, S. (1978) Behavioral contrast in short-term memory: serial binary memory models or parallel continuous memory models? *J. Math. Psychol.* 17, 199–219
32 Bullock, D. (2004) Adaptive neural models of queuing and timing in fluent action. *Trends Cogn. Sci.* 8, 426–433
33 Conrad, R. (1960) Serial order intrusions in immediate memory. *Br. J. Psychol.* 51, 45–48
34 Henson, R.N. (1999) Positional information in short-term memory: relative or absolute? *Mem. Cogn.* 27, 915–927
35 Henson, R.N. and Burgess, N. (1997) Representations of serial order. In *4th Neural Computation and Psychology Workshop* (Bullinaria, J.A. *et al*., eds), pp. 283–300, Springer
36 Hartley, T. (2002) Syllabic phase: a bottom-up representation of the structure of speech. In *Proceedings of the 7th Neural Computation and Psychology Workshop: Connectionist Models of Cognition and Perception* (Bullinaria, J.A. and Lowe, W., eds), pp. 277–288, World Scientific
37 Tanji, J. and Shima, K. (1994) Role for supplementary motor area cells in planning several movements ahead. *Nature* 371, 413–416
38 Ryan, J. (1969) Grouping and short-term memory: different means and patterns of grouping. *Q. J. Exp. Psychol.* 21, 137–147
39 Frankish, C. (1985) Modality-specific grouping effects in short-term memory. *J. Mem. Lang.* 24, 209
40 Ng, H.L. and Mayberry, M.T. (2002) Grouping in short-term verbal memory: is position coded temporally? *Q. J. Exp. Psychol. A* 55, 391–424
41 Hitch, G. *et al*. (1996) Temporal grouping effects in immediate recall: a working memory analysis. *Q. J. Exp. Psychol. A. (Hum. Exp. Psychol.)* 49, 116–140
42 Yntema, D.B. and Trask, F.P. (1963) Recall as a search process. *J. Verbal Learn. Verbal Behav.* 2, 65–74
43 Hasher, L. and Zacks, R.T. (1979) Automatic and effortful processes in memory. *J. Exp. Psychol.* 108, 356–388

44 Murdock, B.B. (1997) Context and mediators in a theory of distributed associative memory (TODAM2). *Psychol. Rev.* 104, 839–862
45 Raaijmakers, J.G. and Shiffrin, R.M. (1981) Search of associative memory. *Psychol. Rev.* 88, 93–134
46 Mensink, G.J. and Raaijmakers, J.G. (1988) A model for interference and forgetting. *Psychol. Rev.* 95, 434–455
47 Anderson, J.R. *et al*. (1998) An integrated theory of list memory. *J. Mem. Lang.* 38, 341–380
48 Dennis, S. and Humphreys, M.S. (2001) A context noise model of episodic word recognition. *Psychol. Rev.* 108, 452–478
49 Howard, M. and Kahana, M.J. (2002) A distributed representation of temporal context. *J. Math. Psychol.* 46, 269–299
50 Neath, I. (1993) Distinctiveness and serial position effects in recognition. *Mem. Cogn.* 19, 332–340
51 Hulme, C. *et al*. (1991) Word-frequency effects on short-term memory tasks: evidence for a redintegration process in immediate serial recall. *J Exp. Psychol. Learn. Mem. Cog.* 23, 1217–1232
52 Hulme, C. *et al*. (1991) Memory for familiar and unfamiliar words: evidence for a long-term memory contribution to memory span. *J. Mem. Lang.* 30, 685–701
53 Gathercole, S.E. *et al*. (2001) Dissociable lexical and phonological influences on serial recognition and serial recall. *Q. J. Exp. Psychol. A* 54, 1–30
54 Baddeley, A.D. *et al*. (1998) The phonological loop as a language learning device. *Psychol. Rev.* 105, 158–173
55 Botvinick, M. and Plaut, D.C. Short-term memory for serial order: a recurrent neural network model. *Psychol. Rev.* (in press)
56 Smyth, M.M. *et al*. (2005) Serial position memory in the visual-spatial domain: reconstructing sequences of unfamiliar faces. *Q. J. Exp Psychol.* 58A, 909–930
57 Bliss, T.V. and Collingridge, G.L. (1993) A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* 361, 31–39
58 Zucker, R.S. and Regehr, W.G. (2002) Short-term synaptic plasticity. *Annu. Rev. Physiol.* 64, 355–405
59 Bower, G.H. and Vinzenz, G. (1969) Group structure, coding and memory for digit series. *J Exp Psychol Monographs*, 80
60 Hitch, G.J. *et al*. (2005) How is the serial order of a verbal sequence coded? Some comparisons between models. *Memory* 3–4, 247–258
61 Marslen-Wilson, W.D. and Tyler, L.K. (1980) The temporal structure of spoken language understanding. *Cognition* 6, 194–198
62 Carpenter, G.A. (2001) Neural-network models of learning and memory: leading questions and an emerging framework. *Trends Cogn. Sci.* 5, 114–118
63 Tarr, M.J. (1999) News on views: pandemonium revisited. *Nat. Neurosci.* 2, 932–935
64 Burgess, N. *et al*. (2001) Memory for events and their spatial context: models and experiments. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 1493–1503
65 Howard, M.W. *et al.* The Temporal Context Model in spatial navigation and relational learning: toward a common explanation of medial temporal lobe function across domains. *Psychol. Rev.* (in press)
66 Cumming, N. *et al*. (2003) Testing a positional model of the Hebb effect. *Memory* 11, 43–63
67 Teyler, T.J. and DiScenna, P. (1986) The hippocampal memory indexing theory. *Behav. Neurosci.* 100, 147–154
68 Kali, S. and Dayan, P. (2004) Off-line replay maintains declarative memories in a model of hippocampal-neocortical interactions. *Nat. Neurosci.* 7, 286–294
69 Levy, W.B. (1996) A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. *Hippocampus* 6, 579–590
70 Wallenstein, G.V. *et al*. (1998) The hippocampus as an associator of discontiguous events. *Trends Neurosci.* 21, 317–323
71 Wills, T. *et al*. (2005) Attractor dynamics in the hippocampal representation of the local environment. *Science* 308, 873–876
72 Eichenbaum, H. and Fortin, N.J. (2003) Episodic memory and the hippocampus: it's about time. *Curr. Dir. Psychol. Sci.* 12, 53–57
73 Raffone, A. and Wolters, G. (2001) A cortical mechanism for binding in visual working memory. *J. Cogn. Neurosci.* 13, 766–785
74 Jensen, O. and Lisman, J.E. (1996) Novel lists of 7+/− 2 known items can be reliably stored in an oscillatory short-term memory network: interaction with long-term memory. *Learn. Mem.* 3, 257–263