A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms

M. P. A. Page[1],* and D. Norris[2]

[1]*School of Psychology, University of Hertfordshire, UK*
[2]*MRC Applied Psychology Unit, Cambridge, UK*

We briefly review the considerable evidence for a common ordering mechanism underlying both immediate serial recall (ISR) tasks (e.g. digit span, non-word repetition) and the learning of phonological word forms. In addition, we discuss how recent work on the Hebb repetition effect is consistent with the idea that learning in this task is itself a laboratory analogue of the sequence-learning component of phonological word-form learning. In this light, we present a unifying modelling framework that seeks to account for ISR and Hebb repetition effects, while being extensible to word-form learning. Because word-form learning is performed in the service of later word recognition, our modelling framework also subsumes a mechanism for word recognition from continuous speech. Simulations of a computational implementation of the modelling framework are presented and are shown to be in accordance with data from the Hebb repetition paradigm.

Keywords: serial recall; repetition learning; word-learning; neural network model

1. INTRODUCTION

In this paper, we will be reviewing the link between the immediate serial recall (ISR) task, the Hebb repetition effect (Hebb 1961) and the learning of phonological word forms. We will then present some simulations of a model that is able to capture some key aspects of the combined data and which, we believe, offers some promise for a future integration of these hitherto rather separate domains.

ISR is probably one of the best known of all the tasks in cognitive psychology. In the last two decades or so, there has been a renewed interest in the task, specifically as a target for computational modelling efforts. ISR is, in some respects, an ideal task for such computational modelling, given the wide variety of experimental manipulations that have been performed, most prominently within the framework of the working memory model of Baddeley & Hitch (1974) and Baddeley (1986). The data provide solid evidence that a number of factors affect ISR performance, including the phonological similarity of list items (Conrad & Hull 1964; Baddeley 1968; Henson *et al.* 1996; Farrell & Lewandowsky 2003; Page *et al.* 2007 etc.) and irrelevant sound during list presentation (Colle & Welsh 1976; Salamé & Baddeley 1982, 1986; Jones & Macken 1995; Tremblay *et al.* 2000 etc.). The detailed pattern of data underlying these and other effects has constituted a rich dataset on which to test competing models. A number of such models have been proposed (e.g. Burgess & Hitch 1992, 1999, 2006; Henson 1998; Page & Norris 1998*a*,*b*; Brown *et al.* 2000; Neath 2000; Farrell &

Lewandowsky 2002; Brown *et al.* 2007), and there is still healthy debate on the merits of each.

Our own model (Page & Norris 1998) is perhaps the most closely aligned with the working memory framework, effectively constituting a connectionist, computational model of its phonological loop component. In recent years, we have evinced evidence in support of the idea that the phonological loop is essentially a high-level (i.e. lexical level in the case of ISR for lists of familiar words) utterance plan designed to accomplish the reproduction of a recently encountered sequence of verbal items (Page *et al.* 2007). Other recent data have been supportive of this characterization (Acheson & MacDonald 2009) and of the close links that this suggests between ISR and speech production (cf. Ellis 1980; Page & Norris 1998*b*).

2. LINKS BETWEEN IMMEDIATE SERIAL RECALL, NON-WORD REPETITION AND WORD-FORM LEARNING

In a parallel body of experimental and theoretical work, there has been a good deal of support expressed for links between the ISR task and another speech/ language system, namely that involved in the learning of phonological word forms. In a number of papers (e.g. Gathercole & Baddeley 1989; Brown & Hulme 1996; Gathercole *et al.* 1997, 1999; Gupta & MacWhinney 1997; Baddeley *et al.* 1998; Gupta 2003; Majerus *et al.* 2006), the case was made that there was a tight relationship between the ability to perform ISR and the ability to learn the phonological forms of newly presented words. We shall refer to this as the phonological word-form learning (PWFL) hypothesis. As part of the explanation for this link, it was further noted that performance in both of these

* Author for correspondence (m.2.page@herts.ac.uk).

tasks correlated strongly with performance in the non-word repetition (NWR) task. In the NWR task, a non-word is presented either once, or several times, for immediate repetition on each occasion. The natural interpretation, and one that has found plenty of support (Gupta 1996, 2002, 2005; Hartley & Houghton 1996; Cumming *et al.* 2003; Gupta *et al.* 2005; Page & Norris 2009), is that the novel non-word is perceived as a sequence of smaller units, say phonemes or syllables. What links these tasks, therefore, is the primary requirement to maintain, in short-term memory, the sequence of smaller units that the list/non-word comprises, and to use that representation to generate a reproduction of the recently presented stimulus. Gupta (2002, 2005) has been to the fore in gathering data in support of this hypothesized relation and in incorporating them into computational models of the systems presumed to be involved (e.g. Gupta 2009).

The evidence that PWFL is functionally related to ISR and hence, in that framework at least, to the phonological loop component of working memory goes beyond simple correlations between tasks. Baddeley *et al.* (1998) provided a comprehensive review, integrating data from a variety of sources. They brought together evidence from neuropsychological case studies of so-called 'short-term memory' patients (Warrington & Shallice 1969; de Renzi & Nichelli 1975; Basso *et al.* 1982; Trojano & Grossi 1995 etc.). These patients had experienced, either developmentally or as a result of later insult, a form of brain damage that had drastically limited their auditory verbal span, in some cases to such an extent that they appeared to have no effective memory for the sequential aspects of unfamiliar auditorily presented lists. The key finding was that these patients had a correspondingly reduced ability to learn the phonological word forms corresponding to novel, to-be-acquired words. To our knowledge, this association between ISR and word-learning is without exception in the neuropsychological literature.

Other evidence reviewed by Baddeley *et al.* (1998) came in the form of cross-lagged correlations from children learning their native language, showing a pattern suggesting that NWR ability was causally effective in promoting the development of a larger vocabulary (Gathercole *et al.* 1992). Note that vocabulary size was the measure here. The acquisition of an item of vocabulary can be conceived not only as the acquisition of a word form but also the forming of a link between that acquired form and its meaning (Gupta & Tisdale 2009). Other evidence (e.g. Gathercole & Baddeley 1989) suggested that performance in NWR correlated rather specifically with the word-form acquisition component of this larger process and not with, say, the learning of familiar word pairs. Finally, Baddeley *et al.* also drew attention to supportive data from other groups, including adults with a 'gift' for languages, and people with learning disabilities.

3. THE HEBB REPETITION EFFECT
As a logical extension of construing the NWR task as a form of ISR, our own group focused on the learning

aspects of the PWFL hypothesis. If NWR is akin to ISR, is there a laboratory analogue of the process by which a non-word form gradually becomes a familiar word form over a number of presentations? Our attention focused on the Hebb repetition effect (Hebb 1961). In his seminal experiment, Hebb found that a nine-digit list that was repeatedly presented every third trial among a number of such lists for ISR became progressively better recalled even though the repetition was unannounced to participants and even among participants who claimed not to recognize that any repetition had occurred (cf. McKelvie 1987; Stadler 1993).

Looking at the Hebb (1961) repetition effect, it seemed plausible (to our group, at least) that there was a relationship between that effect and the learning of phonological word forms by repeated presentation. To use an example from Page & Norris (2009), suppose that a participant in a Hebb-type experiment is repeatedly presented with a list of letters 'B J F M L', with the requirement to recall the list on each occasion. Is it likely that performance on this task is entirely unrelated to that on a task in which someone is presented with a novel object, repeatedly told that it is called a 'bejayeffemmelle' and asked, on each occasion, to make an attempt to recall the object name (i.e. a Hebb-like manipulation of the NWR task)? Armed with this intuition, we set about a series of experiments in an attempt to corroborate the working hypothesis that the Hebb repetition effect is, in some sense, a laboratory analogue of the word-form learning process. In what follows, we will briefly review some evidence relating to the Hebb repetition effect that bears on the plausibility of this idea, in each case drawing attention to a particular property of the effect itself.

(a) *The Hebb repetition effect is not always dependent on spacing*
Following up on Hebb's (1961) experiment, Melton (1963) showed that the learning of a repeatedly presented list was weakened to the extent that repetitions were spaced further apart; he was unable to observe learning when the repeated list was presented on every sixth, rather than on every third, list. If true, this would be discouraging for our working hypothesis because it would imply that successful word-learning would require repeated presentations that were themselves closely spaced, perhaps implausibly so. In Cumming *et al.* (submitted), we showed that it was, in fact, possible to show strong Hebb repetition effects for repetitions at spacings up to every 12th list (the largest spacing tested). The key factor that differentiated our experiment from Melton's was that in our experiment we used a pool of stimulus words that permitted us to manipulate the item sets from which repeating lists, and the non-repeating filler lists in between, were drawn. When all lists (repeating and fillers) were drawn from the same item set, Hebb repetition learning was weak or non-existent at all spacings. By contrast, when fillers were drawn from a different item set than were repeating lists, strong repetition learning was restored. From the point of

view of our working hypothesis, this was more encouraging, given the low likelihood that repeated presentations of a to-be-learned word form would be separated by other (also unfamiliar) word forms that were its phonemic 'anagrams'.

(b) Multiple lists can be learned simultaneously

As part of the experiment by Cumming *et al.* (submitted), described in the previous paragraph, we were able to show that Hebb repetition learning could be seen for multiple lists simultaneously. To be specific, in that experiment, participants might be exposed to one list repeating every third list, interleaved with another repeating every sixth list, another repeating every ninth list and another repeating every twelfth list. All lists were well learned provided that they and the intervening fillers were not drawn from the same item set. This was encouraging for several reasons. First, it suggested that if the mechanism for Hebb repetition learning is shared at some level with that of the learning of phonological word forms, then there would be no necessary problem if people were required to be in the process of learning several different words over the same time period (as is surely the usual case). Second, it supported the conclusion of various other studies (e.g. Cumming *et al.* 2003; Hitch *et al.* 2005) that Hebb repetition learning does not proceed via the strengthening of generic position–item associations (though see below for a discussion of Burgess & Hitch's (2006) modification to their model).

In place of position–item association models, we have preferred an alternative class of models based on the idea of chunking. This idea, reminiscent of that famously explored by Miller (1956), involves the proposition that sequence learners tend increasingly to perceive items that are part of a frequently presented list as being part of a larger, and to some extent indivisible, chunk. For example, assuming one is familiar with the chunk YMCA, it is unlikely to be activated to any extent by the sequence DMPA solely on account of the fact that the two lists contain two letters (M, A) in the same list positions. There is some intuition that once a chunk has been learned, it acts as what Johnson (1970) called an 'opaque container', activating significantly only when stimulated by the correct items occurring in the correct order. Where words are concerned, this intuition has been supported within a research tradition centred on the Cohort model of word recognition (Marslen-Wilson & Tyler 1980): the familiar list YMCA would not be in the cohort of words activated by the stimulus DMPA, just as the word 'mouse' is not in the cohort of words activated by the word 'house'. The model that is discussed later is a chunking model, and one that is qualitatively and quantitatively consistent with both the findings relating to the non-positional nature of the Hebb repetition effect and the general precepts of a word-recognition model in the Cohort tradition.

(c) Hebb repetition learning is fast and long lasting

Most experiments investigating the Hebb repetition effect involve around eight repetitions of the repeating list, evenly distributed among 20 or so filler lists. In a typical such experiment, performance, measured as the percentage of items recalled in the correct list position, will rise from a baseline of around 50 per cent at a rate of approximately 3–4% per repetition, with performance on filler lists staying fairly stable at 50 per cent throughout the experiment (i.e. showing little propensity to improve with general practice). In order for this reasonably rapid rate of improvement to occur, learning itself must be rapid. In particular, no learning could be observed unless significant learning took place on the first presentation of the list: if such learning did not take place, then the first repetition would be indistinguishable from the first presentation, and the process would never 'get going'. A necessary corollary of this relative fast, one-trial, learning is that all lists in a given experiment, including filler (non-repeating) lists, are learned to some extent. At the time at which each is first presented, it is unknown whether they will repeat. Provisional learning is thus necessary for every list.

In Page & Norris (2009, p.142) we gave details of an informal experiment, in which participants showed enhanced familiarity in an unexpected test for a list that they had heard, presented among a group of 10 such lists on one occasion at least 20 min earlier. A more formal test of the longevity of Hebb repetition learning was conducted by Cumming *et al.* (submitted), in which participants in the spaced-learning experiment described above were invited back unexpectedly some three to four months after the initial experiment. At this later date, participants demonstrated enhanced recognition and enhanced recall of the previously repeated lists that had been presented only eight times over three months before. Hebb repetition learning is, we concluded, relatively fast and relatively long lived, facts that are at the very least consistent with our working hypothesis of a relationship with word-form learning. Among others, Dollaghan (1985, 1987) has demonstrated so-called fast-mapping for children learning words, whereby both word forms and meanings of novel word stimuli are apprised after two or three presentations.

(d) Young children exhibit a Hebb repetition effect

It is fortunate for our working hypothesis that young children do exhibit a Hebb repetition effect. In our own unpublished experiments, we were able to show that such effects could be observed in children as young as 5 years old. Having said that, the effects that were observed were surprisingly weak, and it was necessary to employ strong Hebb-type manipulations (e.g. closely spaced repetitions) for them to emerge clearly. In Page & Norris (2009), we speculated that this might have had something to do with the levels at which learning might be expected. Although it has been mostly implicit up until this point, it should be clearly acknowledged that the Hebb effect in ISR operates at a different level from the learning of lists of phonemes (or other sublexical items, e.g. syllables). Nevertheless, the data described in the introductory sections strongly suggest the existence of a single short-term sequencing mechanism that

operates across words in a list, and across sublexical items in a word/non-word. For children, however, it may be that attention is focused on the learning of lists of sublexical items that make up a to-be-learned word, rather than on the across-word sequence information that is present in the (rather unusual) ISR task. Such a focus would go some way to accounting for the weakness of the observed Hebb effects in children. As far as our working hypothesis was concerned, we were happy to observe such Hebb effects in children at all.

Finally, while on the subject of children's sequence learning, we should note the growing literature based on Saffran and colleagues' original finding of so-called statistical sequence learning in infants (Saffran *et al.* 1996, etc.). Although we will not emphasize the point in the description that follows, to the extent that some of these 'statistical learning' effects depend on the acquisition of frequently repeated subsequences (chunks), we are optimistic that they will fall within the domain of application of the model set out below.

(e) *Other effects and recent data*

In Page & Norris (2009), we discuss the consequences for the Hebb repetition effect of a number of other manipulations. For example, data have shown that, in an experiment in which participants are only asked to recall a proportion of the lists with which they are presented (with the recall requirement's being indicated at the end of list presentation), there is considerably better learning for repeating lists that are recalled than for repeating lists that are heard but not recalled. Our own work on this subject (Page *et al.* submitted) has suggested that list recall is important but not entirely necessary for learning to accrue, particularly where list recognition rather than list recall is the measure of concern. Nevertheless, it is tempting to predict that word-learning should be better for words that the participant attempts to recall, compared with those that are simply heard.

Mosse & Jarrold (2008) examined the correlations between Hebb repetition effects in both verbal and spatial serial recall, with both word and non-word paired-associate learning. Their participants were 5- and 6-year old children, and the results confirmed that Hebb effects are (rather weakly) present in such populations. More detailed analysis revealed an interesting and, from our perspective, supportive pattern: Mosse and Jarrold found that there was Hebb repetition learning in both verbal and spatial ISR and that the magnitude of this learning, in each of the modalities, correlated significantly with the non-word paired-associate learning but not the word paired-associate learning. This result is entirely consistent with the data reviewed above: it suggests that the sequence-learning element of the Hebb repetition effect is related to the sequence-learning elements of the non-word paired-associate learning. Note that the significant correlations are between across-word (for the verbal task) or across-location (for the spatial task) sequence learning in the ISR task, and within-non-word (sublexical) learning in the non-word paired-associate learning task. This once again suggests some shared mechanism across levels, even

for this participant group. Interestingly, there is an element of the Hebb repetition learning that seems to be domain general, given that even spatial Hebb repetition learning correlated with non-word paired-associate learning. As we have noted in more detail elsewhere (Page *et al.* 2006), the existence of a spatial Hebb effect does not in any way weaken the hypothesized relation between a phonological Hebb effect and PWFL. Importantly, in Mosse and Jarrold's data, correlations with performance on the non-word paired-associate task were separably reliable for both domain-general Hebb repetition learning and domain-specific performance on the ISR task itself. This distinction suggests that learning is affected both by the 'quality' of the list representation and by an independent learning factor—this is reflected in aspects of the model that is presented below. Mosse and Jarrold's study is the first that has explicitly tested the relationship between Hebb repetition learning and word-learning. It is supportive of the framework that we outlined in Cumming *et al.* (2003) and in subsequent papers (e.g. Page *et al.* 2006).

Finally, another recent experimental study has directly investigated this hypothesized relationship between Hebb repetition learning and word-learning. Szmalec *et al.* (2009) showed in one experiment that Hebb repetition effects were observable in the recall of visually presented nine-item lists of single-syllable non-words, where those nine non-words were grouped by pauses into three groups of three non-words. The data showed that individual groups were subject to repetition learning: learning was evident even when the 'repetitions' were not of the whole list of nine items but, rather, of the individual groups of three items, with the groups themselves being presented in a different order on each occasion. From the perspective of a chunking model, therefore, the pauses introduced during presentation had apparently, as intended, led to the partial learning of individual three-syllable chunks. In a second experiment, employing the same participants and performed some 5 min after the first, participants were required to perform an auditory lexical decision task. Some of the non-word foils in this task comprised the same three-syllable groupings as had been learned (or partly learned) in the preceding Hebb repetition experiment, but this time pronounced as a single three-syllable non-word. The key finding was that participants in the auditory lexical decision task found it harder (in the sense of an increased RT) to reject foils derived from groups learned via repetition in the first experiment, than they did to reject appropriate control non-words. Note that this result held even though in the first experiment the ISR stimuli had been presented visually. The authors' interpretation was that participants in the first experiment had recoded the visual stimuli into a speech-based (phonological) form that was thus more easily rehearsed and recalled. In doing so, the repeating three-syllable groups had started to become lexicalized. This process of incipient lexicalization in the verbal system had led to their becoming more difficult to reject as non-word foils in the auditory lexical decision task. Recent experiments have gone even further by showing that

three-syllable groups learned in the context of a visual Hebb repetition experiment show lexical competition with familiar words in a subsequent auditory lexical decision task.

These last two experiments have, in their different ways, directly addressed the key prediction of the framework that we have been setting out. It appears that the repeated presentation of sequential stimuli that either are directly perceivable as phonological sequences or can be recoded as such, results in the establishment of chunk-like representations in memory. The degree to which these chunks are learned is dependent both on the quality of the short-term representation of the stimulus list itself and on the effectiveness of a domain-general sequence learning mechanism. The chunk-learning is flexible, in as much as it can be rapidly deployed to initiate learning of sequences that have been presented only once, yet it results in relative stable representations that can survive multiple lists being learned in parallel and that can have a duration extending to months. Put together with the variety of data that indicate a key role for phonological short-term memory in the learning of phonological word forms, there is a clear demand for a model that integrates verbal ISR, Hebb repetition effects, the learning of phonological word forms and, almost inevitably, the recognition of those word forms from continuous speech. The model that we describe in the remainder of this paper is an initial attempt to meet that demand.

4. A UNIFIED MODEL OF IMMEDIATE SERIAL RECALL, HEBB REPETITION LEARNING, WORD-FORM LEARNING AND RECOGNITION

The model presented here has a heritage in the models of Grossberg (1978), the masking-field model of Cohen & Grossberg (1987) and the SONNET model of Nigrin (1993), adapted and extended by Page (1994). Having said that, it departs from those models in some important ways, which we will try to identify during its description. Our model also inherits some of its structure from the Norris (1994) Shortlist model, particularly in its treatment of the competitive parsing of lexical items from continuous speech. Because of the close relationship our model bears with this, and other, competitive models of word recognition, we will not be dwelling on the technical aspects of that faculty here. We will, instead, focus on the modelling of the Hebb repetition effect, attempting to capture some of the key characteristics of this effect described above.

The model is formulated as a localist, connectionist model. It comprises a set of units connected together in various structured ways by connections, some of which can vary in strength. Because the model is localist in flavour, the activation of a given unit indicates in a fairly direct way the presence (or degree of presence, see below) of a specific, identifiable stimulus in the world (Page 2000). The description that follows necessarily neglects some of the implementational details—these can be found in the appendix.

The model itself comprises units arranged in four distinct layers. These are shown in figure 1, and they are called the occurrence layer, the recognition layer, the order layer and the production layer. (Some of the functionality assumed in the operation of the model presumes the existence of some other subsidiary processing units, omitted for the sake of clarity.) The layers are connected together in the manner shown in the figure. Note that the thick black arrows represent one-to-one connections between units in connected layers: a given unit in one layer is connected only to the corresponding unit in the connected layer. For that reason, each of the four layers should be thought of as containing the same number of units. It is a corollary of ours being a localist model that there are at least as many units in each layer as there are words or sublexical chunks to be learned. The occurrence layer and the production layer have intra-layer excitatory connections, such that activation of one unit in a given layer is capable of sending positive, excitatory input to other units in that layer. By contrast, units in the recognition layer are connected to one another by inhibitory connections, such that activation of one unit results in the suppression of activation at other units to which it is connected. It is this set of inhibitory connections that gives this layer a competitive character.

(a) The occurrence layer

Units in the occurrence layer indicate the degree of occurrence (see later) of the item that they represent. Some of these occurrence units are primary, in the sense that they are solely activated by the activation of units that are outside the model as depicted. Other occurrence units are secondary, in that they are only activated as a result of the activation of other occurrence units. To give an example, occurrence units corresponding to phonemes (that we will, not uncontroversially, take here to be the fundamental building blocks of speech-based word forms), will be activated by units representing something akin to subphonemic features. These subphonemic feature units are not depicted in the current model, as we assume that they are not, by their nature, *sequential* constituents.

Occurrence units, as their name suggests, signal the occurrence of their corresponding item in the world. They do so by 'firing', that is, emitting a pulse of activation, a certain number of times in rapid succession. (The term firing is by loose analogy with the behaviour of neurons.) For example, the unit corresponding to a given phoneme will fire a maximum number of times, say 10 times, in response to the clear unambiguous presentation of that phoneme. In response to a mispronounced version of the phoneme, or perhaps to a stimulus intermediate between two phonemes, that unit would fire fewer than its maximum number of times. That is what was meant above by the unit indicating the 'degree of occurrence' of its associated item.

Secondary units in the occurrence layer will also fire in response to an external stimulus, but only if that external stimulus activates the occurrence units to which they are connected. Let us assume that there
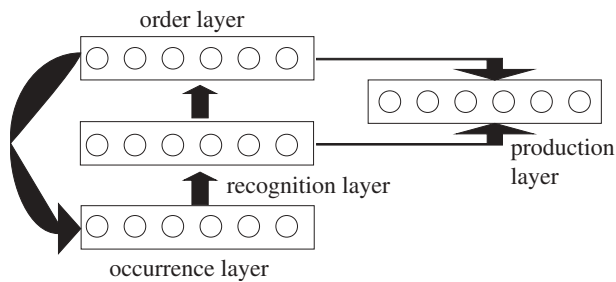
Figure 1. The structure of the model.

is a secondary occurrence unit corresponding to the sequence CAT, which is connected to the primary occurrence units representing C, A and T. (We shall not use a phonetic transcription here, to highlight the fact that the model can be fairly generally applied to sequences of items, whatever those items might be.) Ideally, we will expect that the CAT unit itself activates maximally if its components occur in the correct temporal order. This will entail that the occurrence units to which it is connected fire in the correct temporal order. We would not, however, expect the CAT node to activate maximally if its components occur in an incorrect order, such as in the sequence ACT.

To ensure that the CAT unit activates in the correct way, we make several key assumptions. First, that the CAT unit has differing connection strengths on each of the connections that it receives from the occurrence units corresponding to its components. Specifically, the positive connection from the C unit is stronger than that from the A unit, which in turn is stronger than that from the T unit. Let us assume that these strengths are 10, 9 and 8, respectively. Second, we assume that the pulses of activation that are sent along these connections when a given connected unit itself fires are of unit strength that becomes modulated (e.g. multiplied) by the strength of the connection. A pulse that arrives from the C unit, therefore, arrives with strength 10 at the CAT unit, while a pulse that arrives from the A unit arrives at the CAT unit with strength 9 and so on. Third, we assume that the CAT unit has a threshold for firing: it will itself fire if a pulse arrives with a strength that exceeds the current value of its threshold. As an example, let us assume that the CAT unit has a baseline threshold of 9.5. Finally, we assume that each time the CAT unit fires, its threshold is reduced by a particular small decrement, set to 0.1 for the purposes of this example.

With these assumptions in place, we can now see that the CAT unit behaves in the appropriate fashion. Suppose that the units C, A and T each activate maximally and in the correct order (i.e. C fires 10 times, then A, then T). All the first 10 pulses that arrive at the CAT unit will arrive with strength 10. Each exceeds the firing threshold, that starts at 9.5, so the CAT unit itself fires 10 times, after which its threshold will have been reduced to 8.5. The pulses from the A unit then arrive with strength 9, and given that the threshold has been lowered from 9.5 to 8.5, each is sufficient to cause the CAT unit itself to fire. After the presentation of A, therefore, the CAT unit has fired 20 times and its threshold has been reduced to 7.5. In this state, the CAT unit receives pulses of strength 8 from the T

unit. It fires in response to each of these, such that by the end of the presentation of the sequence CAT, the CAT unit has fired 30 times, which, as a three-item chunk, is as much as it will ever fire in response to any stimulus list. It is, therefore, considered to be maximally activated by the sequence CAT, as required.

Now imagine what happens in response to the near-anagram sequence ACT. Assume again that the baseline threshold is set at 9.5. When the pulses arrive from the A unit, they arrive with strength 9. They are not sufficiently strong, therefore, to lead to firing of the CAT unit. The subsequent 10 pulses from the C unit are sufficient to cause the CAT unit to fire 10 times because they arrive with strength 10. The threshold is reduced from 9.5 to 8.5. Finally, the pulses from the T unit arrive with strength 8; this is below the firing threshold of 8.5, so the CAT unit does not fire in response to any of the pulses arriving from the T unit. By the end of the presentation of the sequence ACT, therefore, the CAT unit has only fired 10 times, as compared with 30 times in response to the sequence CAT.

In summary, secondary occurrence units represent sequences of items (i.e. chunks) where the occurrence of those items is itself signalled by the firing of other occurrence units. Secondary occurrence units respond in an order-sensitive way to the firing of these constituent occurrence units. Although we have presented the threshold-comparison process here as being deterministic, simulations described later include an element of noise here: the threshold is described by a sigmoid function rather than by a step function. Details are given in the appendix, but the basic behaviour outlined here is unaffected.

(b) *The recognition layer*
Let us assume that there are primary occurrence units for the items C, A and T, and secondary occurrence units for the sequences (words) AT and CAT. When the stimulus sequence CAT occurs, each of these five occurrence units will activate maximally. The single-tons will fire 10 times, the AT unit will fire 20 times and the CAT unit will fire 30 times. With all of these units firing, which of the lists (words) should be recognized? Intuition suggests that it should be the word CAT, and the job of the recognition layer is to ensure that this is the case. Activations at the occurrence layer are forwarded in a one-to-one fashion to units in the recognition layer. Given the inhibitory connections within that layer, a competition for activation ensues. This competition for activation is biased in favour of 'larger' occurrence units, that is, those with a higher level of maximal firing. Having said that, a larger unit will only win the competition if the corresponding occurrence unit has received a large proportion (near unity) of its maximal expected firings. For example, in the above example, recognition units for C, A, T, AT and CAT will all be receiving their maximal expected activation; under these circumstances, the competition is biased towards the winning of the competition by the largest unit, CAT, as required. If, however, the sequence CA occurred, then while the units for C and A would be receiving

their maximum expected 10 pulses, the AT unit would only receive 10 of its expected 20, and the CAT unit would only receive 20 of its expected 30: under these circumstances, the best 'parsing' of the sequence CA would be simply as a C followed by an A, and these would win the competition in turn at the recognition layer (see below).

The technicalities of the recognition layer are described in the appendix, but it plays essentially the same competitive parsing role as is implemented in a variety of foregoing models. It echoes Cohen & Grossberg's (1987) masking field, a concept that was extended in different directions by Marshall (1990); Nigrin (1993) and Page (1994). In the word-recognition literature, it resembles the competition for activation in the Norris (1994) Shortlist model, among others. Moreover, the procedure for activating secondary occurrence units as a sequence unfolds bears many similarities to models of the Cohort-type (Marslen-Wilson & Tyler 1980). Notwithstanding these technicalities, for a preliminary understanding of our model, it is sufficient to have a qualitative understanding of the parsing process that this layer implements.

A final point should be made regarding the recognition layer: the competition for activation is established in such a way that there is no effective competition between units that, in activating, never received pulses simultaneously. To give an example, if there are learned occurrence units for the chunks THE and CAT, then in response to the sequence THECAT, each of these will maximally activate. If care were not taken, however, there would also be a competition between recognition units corresponding to THE and CAT, such that only one of them would be able to sustain its activation over time. It is for that reason that units that never simultaneously received a pulse while they activated have no need to compete. By the time that the C occurs in the sequence THECAT, the occurrence unit for THE has already received its maximal input. After this point occurs, its recognition unit is no longer impacted by, nor does it impact, any recognition unit that is not already either fully or partially active. All things being equal, we will expect the recognition unit for THE to win its competition (over T, H, E and perhaps HE) earlier than CAT wins its competition over C, A, T and AT. (In the full model, there is a mechanism that effectively guarantees that a competition that starts earlier always finishes earlier.) As each unit wins a competition at the recognition layer, it is 'forwarded' to the order layer and its recognition activation is reset. This prepares it to respond to a subsequent occurrence of its corresponding list. The consequences of the two 'word' recognition units, THE and CAT, winning their competitions in the appropriate order are felt at the next layer, to which we now turn.

(c) *The order layer*

The order layer implements our primacy model of ISR (Page & Norris 1998a). Its implementational details can be found in the original paper. To characterize briefly, the order layer has units which activate such that the activation of units representing items that occur earlier in a list is higher than the activation of units representing items that occur later in that list. In fact, for a normal (ungrouped) list, the activation of units representing successive items is a linear function of the corresponding item's list position. For the (unlearned) list ABCD, therefore, the competition at the recognition layer will be won by the units representing A, B, C and D in turn, resulting in a gradient of activations across the four corresponding units at the order layer. This might, for example, result in the order unit representing A having an activation of 10, the order unit representing B having an activation of 9, with activations of 8 and 7 for the order units representing C and D, respectively. Finally, for reasons discussed in detail in the original paper, this so-called primacy gradient of activations is modulated (e.g. multiplied) by a value that decays exponentially at all times after the presentation of the list begins. In our model, the decay is quite rapid, such that a gradient of activations will decay to half its original values after a period of around 2 s. This time-based decay (that is itself far from uncontroversial) is responsible for the primacy model's ability to explain a variety of ISR data, including primacy effects, effects of short delays and effects of word length.

To forestall any confusion, we note that the units over which a primacy gradient can be instated represent tokens rather than types. To illustrate the distinction, suppose the model is presented with the list ABAD. This will be represented across four order units, including two different units that each represents a token of type A. There will typically be several units representing each of the known types, but only one will activate to any given occurrence of that type. This does make for some complication in the long-term memory system, and the issue is discussed extensively in Nigrin (1993) and Page (1993, 1994). Space forbids a reprisal of that discussion here—suffice to say that the mechanism prevents the repeated firing of the same incoming connection to a given occurrence unit. Such repeated firing would otherwise cause an occurrence unit that had learned the word PIT to fire maximally in response to the sequence PIP (by firing along the same strong P connection in response to both P tokens; thanks to an anonymous reviewer for the specific example).

The order layer is the model's primary store of order information. As can be seen from the depiction of the model in figure 1, the order-unit activations are copied to two places. First, they are copied back to the occurrence layer. It is here that they play a role in the learning of new chunks. To anticipate discussion of this point below, it is the primacy gradient of activations that one would experience across the order layer in response to the *previously unlearned* word CAT (e.g. C = 10, A = 9, T = 8) that is 'copied' into the strengths of the incoming connections to a new CAT unit as it is being learned (also 10, 9 and 8; see above). For this reason, the short-term order information held at the order layer is necessary for a long-term chunk to be established at the occurrence layer. In the complete absence of a primacy gradient

at the order layer, no new chunk can be learned at the occurrence layer. This is, in a nutshell, our account of the dependence of word-learning on short-term retention of order. For instance, for an extreme example of a short-term memory patient, for whom the order layer is assumed to be entirely non-functional, there would be no way of learning new words, as is consistent with the data.

(d) *The production layer*

The production layer implements, in general terms, the process of serial recall. It has some structure beyond that depicted. For the phonological loop, for example, that extra structure would comprise the machinery needed to implement a speech production process that unpacks lexical items into their constituent phonemes/syllables (see Page & Norris (1998*b*) for an implemented integration of the model with those of speech production pioneered by Gary Dell and colleagues). Nevertheless, for current purposes, the production layer's function can be summarized fairly simply: when sequence recall is required, the production layer implements the selection of its most activated unit, the recall of the corresponding item and the suppression of the selected unit to prevent its repeated selection (cf. Grossberg 1978; Houghton 1990). For example, in response to the unlearned list ABCD, the order layer will contain the decaying primacy gradient described in the previous section. When recall (or rehearsal) is required, that primacy gradient will be copied to the production layer where the cyclic process of selection will begin. Although the process of selecting the most active unit is subject to 'noise', we would expect the production unit representing A to be most active and selected first. Its production unit (or, more likely, its ability to win the selection process) is then suppressed, allowing the unit representing B to win the next round of selection. This is followed in turn by the units representing C and then D. Thus, in the absence of noise, the recall cycle applied to the production layer will result in the recall of the items A, B, C and D, in that temporal order—a correct recall of the stimulus list. Although the details can be found in the original description of the model, it should be clear that in the presence of selection noise, the most likely errors will be the exchange of positions between adjacent items. For example, if any item is going to be chosen ahead of A, then it is likely to be the next most active item B, and so on. Because of the decay of the primacy gradient with time, the 'gaps' between activations of adjacent items will be smaller for items recalled later in the lists. These later items will therefore be recalled more poorly, as is characteristic of the behavioural data. Finally, items whose activation decays too far, risk their activation falling below a recall threshold, resulting in an omission.

Having introduced the fundamentals of the model, in the remainder of the paper we will focus on the learning that underlies the Hebb repetition effect and how it results in performance improvements that are consistent with those seen in the data. We will then present the results of some specific simulations

before indicating how the framework that we have developed would apply to word-learning and to word recognition.

(e) *Learning, and the genesis of the Hebb repetition effect*

We will illustrate our account of the Hebb effect by using as an example the repeated presentation of the previously unlearned list ABCD. We will assume that the individual list items (A, B, C and D) are themselves familiar, and are represented fully at all other layers. (The meaning of 'full' representation will become clear shortly.) Shortly after the presentation of the list ABCD, the last item will have been forwarded from the recognition layer to the order layer, leaving the recognition layer devoid of strong competitors. A primacy gradient of activation at the order layer will also have resulted as will the potential for it to be copied to the production layer when recall is required. We now introduce the process by which an occurrence unit, and its associated units, comes to learn the chunk ABCD over repeated presentations. To do this, we will differentiate between uncommitted, engaged and committed occurrence units. An uncommitted occurrence unit is a secondary occurrence unit that has not yet become even partially committed to any chunk. An uncommitted unit is, therefore, one that has not seen any learning at its incoming connections from other occurrence units. By contrast, a committed unit is one that has completely learned a chunk; it competes fully and strongly at the recognition layer when its learned sequence is presented. Finally, an engaged unit is one that has started the process of moving between being an uncommitted unit towards being a committed unit. This process necessarily begins on the first presentation of a given list.

An uncommitted occurrence unit has not carried out any learning in its incoming connections. In our model, the weights to uncommitted occurrence units are initialized to values that are randomly distributed, with small variance (e.g. 0.05), around unity. The threshold of each occurrence unit is always set to a proportion (0.95, as above) of the highest connection strength that a unit receives. The decrement by which the threshold is lowered each time the occurrence unit fires is set to an appropriate value that is also a proportion (0.01 above) of the highest connection strength to the relevant unit. Finally, the threshold is prevented from falling below a given value (e.g. 0.95). These settings will ensure that the unit continues to behave as described above, even as its incoming weights are strengthened during learning. More to the point, the settings also mean that for any given list, a large number of uncommitted occurrence units will fire in response to every firing of each primary occurrence unit associated with a given stimulus list. In a sense, therefore, uncommitted occurrence units respond strongly to all lists—they are not order selective. They are, however, prevented from winning any competition at the recognition layer because that competition is modulated by what we might call a unit's degree of learning. Fully committed nodes compete fully at the recognition layer, but a unit that is just

starting to learn a given list cannot outcompete an active recognition unit corresponding to a fully learned list. Because uncommitted occurrence units are not order-selective and because many of them will therefore fire strongly in response to any given list, there is a potential for this to lead to a proliferation of weak competitors slowing up competitions at the recognition layer. To prevent this from happening, it is ensured that only the single most active uncommitted occurrence unit is permitted to forward its activation to the recognition layer, where it can begin to compete for activation.

Immediately after the first presentation of the list ABCD, therefore, recognition units corresponding to each of the elements (A, B, C and D) will have activated in turn, leading to the corresponding primacy gradient at the order layer. While they are activating, no uncommitted unit will be able to gain activation at the recognition layer owing to the strong competitive advantage of these fully learned items. However, as the final list item, D, wins its competition, is forwarded to the order layer and has its recognition unit reset, there comes an opportunity for an uncommitted unit to activate, in the absence of competitors, at the recognition layer. The most active uncommitted occurrence unit is very likely to be a secondary unit that is connected to each of the occurrence units for the list items, and which has therefore fired in response to each of them as they occurred. It only remains to specify what learning takes place for this uncommitted occurrence unit to change its status from being uncommitted to being engaged to the stimulus list.

As a previously uncommitted node gains activation at the occurrence layer (owing to its lack of order sensitivity) and at the recognition layer (owing to the lack of competition there), it can begin to learn the sequence ABCD that is represented at the order layer and copied down to the occurrence layer. In short, the incoming connection strengths from the occurrence units for A, B, C and D are adjusted to make them somewhat more like the primacy gradient of activation representing the list. The equation for this learning is given in the appendix, but essentially the connection strengths take a step towards some multiple (e.g. 10) times the primacy gradient values. To be specific, let us assume that the primacy gradient representing the list ABCD is A = 5, B = 4.5, C = 4, D = 3.5, with zero gradient activation for all other items. Let us also assume that the initial connection strengths to the most active uncommitted unit from the A, B, C and D occurrence units are 1.20, 1.13, 1.12 and 1.09, respectively (remember that although initial values randomly distributed around unity, the most active uncommitted unit is likely to be one that, by chance, has incoming strengths that are approximately rank ordered in a manner appropriate for the to-be-learned list). On the initial learning trial, we might say that the connection strengths are adjusted to move them a proportion of the way from where they are towards values that are 10 times the primacy gradient values: for this example, the connection strengths will move a small step towards values of 50, 45, 40 and 35. The size of the step that the connection strengths take in this direction can be set by a value that we call the learning rate. If the learning rate were 0.1, then the connection strength

from the A occurrence unit to the to-be-engaged ABCD unit would move from an initial value of 1.20 to a new value equal to (0.9 × 1.20) + (0.1 × 50) = 6.08. Performing the same calculation for each of the connection strengths, they would move from being 1.20, 1.13, 1.12 and 1.09 (from A, B, C and D units, respectively) to being 6.08, 5.52, 5.01 and 4.48. In other words, the weights increase and, as a result of them becoming more like a primacy gradient, the previously uncommitted unit becomes more order-selective. This means that, after learning, it will tend to activate more specifically in response to the list ABCD.

An occurrence unit that has taken a step towards learning a given list is said to be engaged (neither uncommitted nor fully committed), and it is very likely to activate more strongly than any other uncommitted or engaged unit on the next occasion that the list ABCD is experienced. On this second occasion, more learning will ensure that its incoming connection strengths become yet larger and even more parallel to the corresponding primacy gradient. Thus, the occurrence unit becomes progressively more order-selective and a progressively stronger competitor for activation at the recognition layer. Eventually, after a number of learning trials, the engaged unit becomes sufficiently committed to its learned list that it is can be termed fully committed—this is signalled by its largest incoming weight reaching a particular large value. At this point, it can compete on an equal basis with other fully committed units at the recognition and production layers.

The learning process described above depends heavily on both the presence and the strength of the primacy gradient. If the primacy gradient were absent, no learning could occur (cf. short-term memory patients). If, on the other hand, the primacy gradient were only weakly present (e.g. heavily decayed, as might be associated with poor ISR performance), then list-learning will be slowed accordingly. Word-learning would therefore be slower for low-span versus high-span participants, as the data suggest.

(f) Learning at the production layer

The process of learning at the occurrence layer allows a previously uncommitted occurrence unit to become engaged to, and eventually fully committed to, a repeatedly presented list. That unit will, in a sense, come to recognize that list. This does not, however, explain how repeated presentation of a list can result in the progressively better recall seen in, for example, the Hebb repetition effect. This improvement depends on an analogous process of learning that takes place at the production layer. Remember that at the end of the first presentation of the list ABCD, there is a primacy gradient representing the list present at the order layer, which is hence able to be copied to the production layer (figure 1). There is also a previously uncommitted recognition unit active at the recognition layer. Under these circumstances, the production layer unit associated with this previously uncommitted recognition unit engages in some learning by changing some of its *outgoing* weights in the direction of the primacy gradient across the A, B, C and D production

units. Outgoing weights, in this case from the to-be-engaged ABCD production unit to each of the A, B, C and D production units, can be thought of as starting from values close to zero. When learning occurs, these values take a step in the direction of some multiple (e.g. 10, as for the occurrence layer learning) of the primacy gradient values copied to the A, B, C and D production units from the order layer. (This copying might even be thought to be contingent on a demand to recall the list—see below for a comment regarding the importance of list recall in the Hebb repetition effect.) So, for example, using the same values for the ABCD primacy gradient given in the previous section, after the first presentation of the list ABCD, the production unit corresponding to the occurrence and recognition units that are becoming engaged to that list will acquire outgoing connection strengths that take a step towards values of 50, 45, 40 and 35 to the A, B, C and D production units, respectively. The size of this step will be determined by the production learning rate: if that learning rate were 0.1, then the outgoing connection strengths would end up close to 5, 4.5, 4.0 and 3.5.

We are now in a position to explain why recall performance is improved for the second presentation of ABCD relative to its first presentation and, by extension, why performance continues improving until it becomes perfect (i.e. ABCD becomes a fully learned chunk in its own right—like the acronym 'YMCA'). When ABCD is presented a second time, the recognition units corresponding to A, B, C and D activate in turn and lead to the corresponding primacy gradient's being instated at the order layer. Immediately after the last of the items has passed through the recognition layer, the engaged ABCD unit can activate at the recognition layer. When it does so, its associated production unit activates at the production layer. The activation of the engaged ABCD production unit in turn activates the production units for A, B, C and D via the primacy gradient that it has partially learned in its outgoing weights. (Note: the engaged unit cannot itself win a competition at the production layer, as the competition takes place between fully committed units.) On account of this primacy gradient, it activates the production unit for A more than that for B, which in turn is activated more than that for C and so on. These activations are added to the primacy gradient activations that are being sent from the order layer, and they therefore make the primacy gradient at the production layer steeper and higher than it would have been after the first presentation of the list ABCD. A steeper primacy gradient is associated with fewer order errors (because the gaps between the activations of successive items are increased) and fewer omission errors (because the gap between activation of each item and some assumed omission threshold is also increased). Repeated presentation of a list should, therefore, result in fewer omissions and fewer order errors, exactly the pattern that is observed.

(g) Basic simulations and some characteristics of the Hebb effect model

The model described in its basic form above has been implemented by the authors as a computer simulation.

Although the reporting of a major set of simulations is beyond the scope of this paper, we will present the results of simulations that will confirm that the model is capable of simulating the Hebb repetition effect itself, while remaining consistent with each of the general characteristics (discussed above) that makes the Hebb repetition effect a good laboratory analogue of word-form learning. The simulated model was exposed to a set of lists for ISR. Sets of lists were established that could demonstrate Hebb repetition learning under various conditions. Serial order recall errors were recorded as they would be for a participant in an experiment; although the model produces full serial position curves, we will use the value of mean proportion correct to illustrate points we wish to make.

In our first simulation, we repeatedly presented a given seven-item list, call it ABCDEFG, with no intervening filler lists, to show that repeated presentation leads to an improvement in recall. We then inserted filler lists in various ways. First, we inserted either two, five, eight or nine filler lists between each consecutive pair of repetitions of the repeating list, where those filler lists were generated randomly from a set comprising the items H, I, J, K, L, M and N. Because this item set does not overlap at all with the item set used to generate the repeating list, the increasing number of fillers (i.e. the increasing repetition spacing) should have no effect on the size of the Hebb repetition effect. As expected, the size of the repetition effect as shown in figure 2 does not depend on repetition spacing.

In addition, we have added to figure 2 a line indicating the performance for Hebb repetitions spaced at six lists apart, starting with the sixth list, but with filler lists derived from the same item set. These were the conditions under which Melton (1963) and we (Cumming *et al.* submitted) were unable to see a Hebb repetition effect. Consistent with these data, no Hebb effect is observed. It is worth taking some time to understand, in general terms, why this is the case. In this paradigm, all filler lists are anagrams of the repeating list. As is required by the logic of the Hebb repetition effect, a separate occurrence unit becomes engaged to each of these *filler* lists on their one (and only) occurrence—the requirement for this first-trial learning is consequent on the fact that the system cannot know in advance which lists are going to repeat and which are not. When the sixth list occurs (i.e. the first appearance of what will be the repeatedly presented list), the uncommitted unit that will become engaged to that list has to suppress the activation of five units that are weakly engaged to anagram lists. At least some of these weakly engaged units activate relatively strongly to the first occurrence of the repeating list, because they are not yet very order-selective and each is engaged to a perfect anagram of the to-be-repeated list. The consequent competition for activation at the recognition layer can take sufficiently long that by the time it resolves in favour of the uncommitted unit, the primacy gradient has decayed sufficiently to ensure that learning can only be very weak. In an experiment in which the Hebb repeating list appears only every six lists, by the time
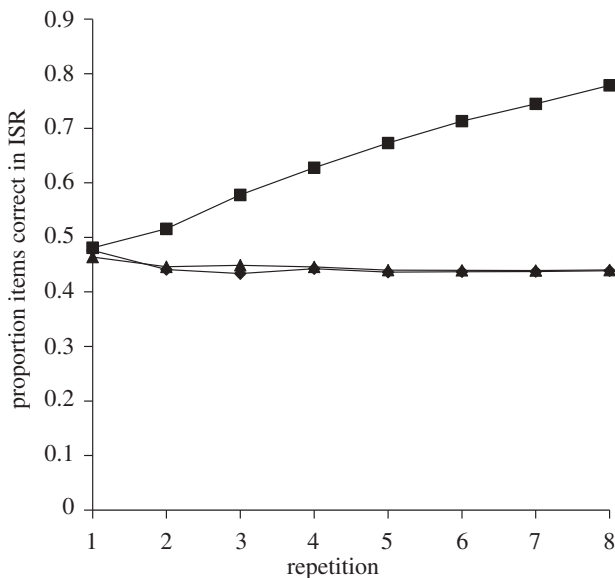
Figure 2. A graph showing ISR performance under various conditions of the Hebb repetition effect. For the sake of clarity, various of the conditions are portrayed using a single line representing Hebb repetition learning. The individual simulations showed that learning in these conditions followed essentially the same course (as per the data in Cumming *et al*. submitted); the line depicts one simulated condition rather than an average across conditions. A positive repetition effect is seen under all conditions, up to and including 12-apart repetitions, except where there is item overlap between the repeating-list and filler-list item sets. When there is overlap of this sort, there is repetition learning for 2-apart repetition, but not for 6-apart, in accordance with Melton (1963). Squares, repeat: 12-apart, 9-apart, 6-apart, 3-apart (all non-overlap), 2-apart (overlap); diamonds, repeat: 6-apart (overlap); triangles, non-repeating fillers.

its first repetition occurs, there are 10 units representing anagrams, whose activation needs to be suppressed at the recognition layer before effective learning can proceed. On the second repetition, there are 15 such units, and so on. It becomes progressively more difficult to ensure relatively rapid learning under circumstances in which large numbers of anagram lists are present and being partially learned. Naturally, as the spacing is reduced to, say, a repetition on every other list, then the situation is ameliorated, as is shown in figure 2.

Provided that their items are not drawn from the same item set, several lists should be learnable simultaneously. To demonstrate this, presentation was alternated between the lists, ABCDEFG, and HIJKLMN, as shown in figure 3. Both lists were learned in exactly the same manner and at the same rate as those shown in figure 2. Moreover, there was no transfer from the learning of the list ABCDEFG to the list ADCFEBG: even though the latter list shares four letters-in-position with the learned list, it was no better recalled than a filler list, in accordance with the results of Cumming *et al*. (2003).

Finally, we consider the speed of learning, the longevity of learning and the importance to the Hebb repetition effect of list recall. First, as must be the case, our model shows first-trial learning and exhibits rates of learning (shown in figure 2) that are designed
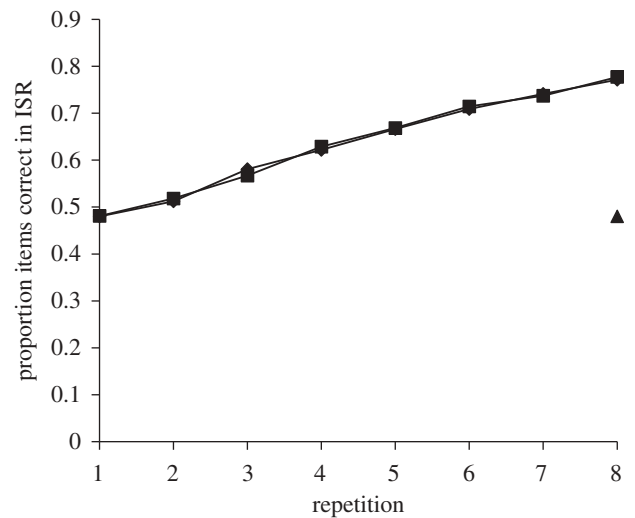

Figure 3. A graph showing a positive repetition effect for each of two alternating and repeating lists. There was no overlap in item set between the lists. The triangle indicates performance, after learning has occurred, for a list derived from the first repeating list, in which alternating items (starting with the first) are maintained in the same positions as in the learned list, with the remaining items moving (cf. Cumming *et al*. 2003). Diamonds, repeating 1; squares, repeating 2; triangles, alternating items maintained from r1.

to be in accordance with those seen in experiments employing seven-item lists—we based our simulations on the approximately 4 per cent performance per repetition seen in all of the non-overlap conditions of Cumming *et al*. (submitted). Second, we note that the learning that is involved has the potential to be extremely long lasting: while weak initial learning may decay over time, perhaps in accordance with a 'use-it-or-lose-it' schedule, there is no *necessary* reason why it should. There is no catastrophic interference of subsequently learned items on previously learned items in this model, as there is in some other classes of the connectionist model (Page 2000). Last, the distinction in our model between learning in the occurrence layer over repeated presentations and learning in the production layer over repeated recalls is entirely consistent with data relating to the Hebb repetition effect when recall of the repeating list is not required.

(h) *The extension to word-learning and to word recognition*

In this paper, we have given a fairly generic account of a model of sequence learning. There is good evidence (Page *et al*. 2006; Mosse & Jarrold 2008) that Hebb repetition learning is observable in many domains (e.g. phonological, spatial, visual). Nonetheless, the model has been designed with the learning of phonological word forms in mind. The data reviewed above suggest that there is some common sequencing mechanism underpinning ISR, the Hebb repetition effect and the learning of phonological word forms. That mechanism is what our model seeks to elucidate. The primacy model gives a good account of ISR data that have now been extended to a preliminary account of the Hebb repetition effect. The production

side of the primacy gradient has already been brought into line with models of speech production (Page & Norris 1998b; Page *et al.* 2007). By incorporating what is essentially a shortlist-like, cohort-like competitive mechanism into the model, we are confident that it can be applied to the recognition of words from continuous speech.

(i) *Comparison with other models*

In this section, we will briefly discuss the revised model presented by Burgess & Hitch (2006) and the computational model of Gupta (2009). These two models are those that are closest in scope to ours, the former being applied specifically to ISR and the Hebb effect and the latter to the link between ISR and NWR. As we hope to show in this very brief account, neither model is able to play the unifying role that we intend for the model presented above.

Burgess & Hitch's (2006) model is based on the idea of position–item association. Each item in a list is associated with a set of context units, the activation of which represents its position in the list. Recall is achieved by reactivating the positional codes in the correct sequence, recalling the most strongly associated item at each position. As Burgess & Hitch (2006) themselves noted, some revision was necessary to their developing model, specifically with regard to the Hebb repetition effect. The previous version (Burgess & Hitch 1999) was unable either to learn multiple different lists over the same set of ISR trials or to cope with data (e.g. Cumming *et al.* 2003) which had showed that there was no transfer of learning from a previously repeated list to a test list in which only some items maintained their (learned) list position. Both these deficiencies were attributable to the use of a single set of positional context units to drive recall of a variety of lists. Accordingly, in their revision, Burgess & Hitch (2006) outlined a solution in which different sets of positional context units are recruited for lists that differ sufficiently. Sufficient difference is taken to obtain if, at any point in the presentation of a list, the match between that list and the pattern learned by a given context set drops below a particular value (0.6 in their simulations). By involving different context sets, several sufficiently different lists can be learned over the same period and without interference. This is because only one item (or perhaps one in a very limited set) becomes strongly associated with any given list position for any given context set. Moreover, there is no transfer from a learned list to one that retains only half of its items in the same position because, on presentation of the transfer list, the lack of sufficient match will lead to the turning-off of the context set associated with the prior learning. Space does not permit a full critique of the Burgess & Hitch (2006) model, but some points merit discussion.

First, the measures taken by Burgess & Hitch (2006) do indeed solve some of the problems with their account of the Hebb repetition effect but only by abandoning the notion of a general positional context signal. In certain, not uncommon, experimental contexts, a given 'positional' context set can, according to the new model, be applied to only a single one of the

lists presented; in the limit, this requires as many context sets as lists. The original motivation for having a common positional context set applied across lists was to account for positional errors in ISR, such as positional protrusions. A positional protrusion is when an item from a given position in one list is erroneously recalled *in the same position* in the following list. The new version of the model, however, suggests that two consecutive lists made up of different list items (or, indeed, two lists differing in their first item) would necessarily attach to different context sets; the model could not therefore show positional protrusions under such circumstances, contrary to the data.

Second, and on a related note, there is nothing *necessarily* positional in character about Burgess & Hitch's (2006) modified mechanism. In their description of the process by which items in a particular list become associated with a given context set, and of the subsequent match process that takes place on later re-presentation of that list, the authors make explicit comparison with the way in which items associate with a whole-list representation in an adaptive resonance theory (ART) network. But in an ART network (or at least in the sequence-processing equivalent of Cohen & Grossberg (1987)), the representation of the whole list is actually itself a localist representation, a chunk node that associates in an ordinal way, not a positional way, with the items in its associated list. The model that we present above is exactly of this type, and crucially allows for an acquired chunk itself (e.g. a word) to be recalled as part of a 'higher level' list (e.g. a list of words). Where Burgess and Hitch associate items in a novel list with a newly recruited positional context set, we (and Cohen & Grossberg 1987; Nigrin 1993; Page 1993, 1994; etc) associate items with a newly recruited unit that comes to represent a new chunk. For Burgess and Hitch, the positional nature of a newly recruited context set is difficult to attest precisely because its positions are not related (by any mechanism in the model) to the equivalent positions in any other context set associated with a different list or lists. In short, if you are going to associate individual lists with list-specific representations, there is nothing obvious to be gained from making that association positional, as opposed to ordinal as in a number of prior models.

Third, in the Burgess & Hitch (2006) model, the mechanism for simulating the Hebb effect cannot be as closely related as the data suggest to the mechanism by which sequences of phonemes come to be learned as a new word. In their model, a learned sequence comprises strong connections between list items and a particular positional context set. By contrast, a word in the model is represented by a lexical (localist) node connected in an appropriate manner to its input and output phonemes. The two are incommensurate in terms of the manner in which they encode serial order: The strengthening of item-to-context-set connections cannot result in the acquisition of anything resembling a new lexical item. Burgess and Hitch concede that their model 'requires further elaboration' with regard to the sequential ordering of sublexical items within to-be-acquired words. Given the strong

and accumulating data regarding a link between ISR, Hebb effects and the learning of phonological word forms, this is an important restriction. Our model is precisely aimed at explaining the links between performance in these different tasks and, to this extent, it clearly differs in its ambitions from Burgess and Hitch's model.

Finally, even within the restricted domain of the Hebb repetition effect, the Burgess and Hitch model might encounter problems. For example, it has been shown (e.g. Szmalec *et al.* 2009; referred to earlier) that for grouped lists, individual groups can be learned provided that the groups themselves repeat across lists, even if the order in which the groups appear changes across presentations. For example, groups G1, G2 and G3 might be learned by exposure to the different lists G1–G2–G3, G2–G3–G1 and G3–G1–G2 (where the dash represents, say, a pause). This begs questions of the Burgess and Hitch model: Given that the lists as a whole are not similar enough to recruit the same context set (assuming groups contain different items, as they usually do), then what is the mechanism for group learning? If the mechanism is held to be via some association with a context set representing a within-group position, then how can three (or more) different items become associated with each within-group position? If, alternatively, different context sets are used to mark the within-group position in different groups, then why do we see a surfeit of between-group transpositions that preserve the within-group position (the group equivalent of positional protrusions)?

Turning now to the Gupta (2009) model, we note that this model continues to use, as its primary short-term sequential memory mechanism, a common set of positional codes with which list items are associated. For this reason, it would suffer the same problems (described above) as did the model of Burgess & Hitch (1999), if this mechanism were invoked in any account of the Hebb effect. In fact, Gupta's model makes no claim to account for the Hebb effect and, in this regard, is of somewhat more limited scope than the model presented above. Gupta's model is, however, intended to illustrate what is shared between ISR and NWR, and it is in the simulation of this relationship that its strengths lie. Word-learning, in Gupta's model, involves a somewhat complex mechanism that introduces two further representations of implied serial order. Representations at the lexical level comprise both a localist representation of a given word and a distributed representation of the same word, the latter comprising a list of syllables slotted into a word frame. These two representations of a given word are linked together by connections established by Hebbian learning. Similar dual representations are present at a syllable level, with the distributed syllable representation comprising phonemes slotted into a syllable template. The slot-based word frames and syllable templates therefore constitute one long-term representation of serial order. In addition, linking the word level to the syllable level and, equivalently, the syllable level to the phoneme level are two simple recurrent networks (SRNs) that learn, over the long term, to convert a distributed word representation into a sequence of syllables and a distributed syllable representation into a sequence of phonemes. These SRNs thus constitute a second and distinct long-term memory for serial order.

It is difficult to see how any of the three ordering mechanisms in Gupta's (2009) model (short-term positional, long-term slot-based and long-term SRN) could be invoked in a successful account of the Hebb repetition effect. If this effect is as closely associated with word-form learning as we (and others) claim, then the model is likely to need modification. Furthermore, as it stands, the model does not account for the apparently strong link between ISR and word-form learning—the word-form learning both within-levels (Hebbian) and between-levels (SRN) is largely independent of the distinct (position-based) short-term memory system. By contrast, in the model that we describe above, there is a single ordinal representation of serial order (the primacy gradient) that drives both short-term and long-term sequential memory. Any deficiencies in this representation will, therefore, make themselves evident in ISR, NWR, Hebb effects and word-learning, as the data suggest.

5. CONCLUSION

We have presented a framework that attempts to unite ISR, repetition learning, word-form learning and the recognition of words from continuous speech. We have provided quantitative simulations of the first two and have, we hope, offered reason to believe that the model can be applied successfully to the latter two. It is important that the modelling of particular tasks (e.g. ISR) takes place in a context that emphasizes the larger, in this case linguistic, context within which the responsible mechanisms developed.

APPENDIX A

Here we describe the implementation of the model that was used in the simulations of the Hebb repetition effect reported above.

A.1. The occurrence layer

Each item from a pool of possible list items was allocated a primary occurrence unit. To simulate the presentation of a given list of seven items, the corresponding primary occurrence units fired 10 times each in the order indicated. Each firing constituted a pulse of unit strength that was transmitted along weighted connections to the secondary occurrence units. The weight on the connection from the ith occurrence unit to the jth occurrence unit was denoted w_{ij}. The pulse that arrived at the ith occurrence unit from the jth occurrence unit was modulated by the weight of the connection between the two, and therefore arrived with strength w_{ij} at its target unit. In response, an occurrence unit that received a pulse would itself fire with a probability

$$p(\text{firing}) = (1 + e^{-s(f-\theta)})^{-1},$$

where f is the strength of the pulse received, θ is the current threshold of the receiving unit and s is the slope of

this sigmoid threshold function ($s = 8$). The threshold, θ, for a given unit had a baseline value of a constant multiple of the largest weight incoming to that unit. That multiple had a value equal to $(P - 0.5)/P$, where P is the peak parameter of the primacy gradient, which is discussed below. In the simulations reported here, $P = 11.5$ (as in Page & Norris 1998a), implying a threshold of 0.957 times the largest incoming weight.

Every time an occurrence unit fired, its threshold was reduced by a value equal to a multiple of its largest incoming weight. This multiple was equal to 1/115 here, derived as the reciprocal of the number of firings of each primacy occurrence unit representing a list item (10) multiplied by the peak value, P, of the primacy gradient (11.5). The threshold of any occurrence unit was not allowed to fall below 0.95.

Once an occurrence unit had started firing (i.e. it is active), the cumulative number of its firings was recorded. This recorded value continued to accumulate throughout the presentation of a list. If, however, an active unit that had not reached a point near to its maximal activation failed to fire in response to a certain number (10) of consecutive firings of any primary occurrence unit, then that unit was reset. Resetting involved setting the cumulative firings back to zero and returning its threshold, θ, to the appropriate baseline level. For certain applications (Page 1994), it may be advantageous to have a primary occurrence unit that signals a definite pause—this will allow the fast resetting of incomplete words, e.g. PANDA, in response to the stimulus PAN.

Before any learning had taken place, the weights of connections between primary occurrence units and secondary units (and, potentially, between secondary units and other such units, though not here) were initialized to values, normally distributed around 1.05 with a small standard deviation equal to 0.04.

A.2. The recognition layer
The competition at the recognition layer was driven by the cumulative firings of the corresponding occurrence units. For the simulations reported here, this competition was assumed to be won in turn by each of the recognition units corresponding to the individual list items. These are the only strongly learned items, so they compete strongly at the recognition layer, though not with each other. By the end of the presentation of the list, the last list item will have just won the competition for activity at the recognition layer; it will therefore forward activity to the order layer, where a primacy gradient will have been established to represent the whole list (as per Page & Norris 1998a; see below).

The activation of the recognition units, r_i, evolved according to

$$\frac{dr_i}{dt} = -r_i + \frac{a_i}{s_i}(B - r_i)(r_i^2 + E) - \frac{I}{s_i}r_i N_i + n(\sigma),$$

where a_i is the activation forwarded from the occurrence unit, s_i is the size of the recognition unit, B is the maximum activation that could be attained at the recognition layer ($B = 1.5$), E is a parameter scaling the excitatory input to the recognition unit ($E = 3$), I

is a scaling parameter scaling the inhibitory input to the unit ($I = 60$) and $n(\sigma)$ is a sample (at each time step) from zero-mean Gaussian noise with standard deviation $\sigma = 0.001$. N_i is the inhibitory input to a competitive unit from other competitive units, indexed by j

$$N_i = \sum_{j \neq i} \frac{a_j}{s_j} r_j^2 \sqrt{w_{\max j}}.$$

So the competitive signal sent out by a given recognition unit was proportional to its occurrence unit activation scaled by its size, the square of its own activation and a measure of the amount of learning that has been undergone, equal to the square root of the maximum weight to its corresponding occurrence unit, denoted $w_{\max j}$. It was this last term that ensured that recognition units corresponding to well-learned lists were able to compete strongly.

For an uncommitted recognition unit, the size of a recognition unit, s_i, tracked its input a_i. For an engaged or committed unit, s_i was set to the value of a_i at the time of first learning. (Alternatively, it may be thought of as having tracked the cumulative maximum input to the recognition unit—these were the same value.) The single uncommitted unit that is allowed to compete at the recognition layer always had a scaled input equal to 1. This would not allow it to win any competition against a fully activated unit that is engaged or fully committed: Although they would have the same, maximal scaled input, the higher value of $w_{\max}$ associated with the engaged or committed unit would ensure that that unit would win out at the recognition layer.

The competition for activation at the recognition layer was run for a simulated time equal to 1 s, or until one of the recognition units achieved an activation of 1.05, whichever occurred sooner. For these simulations, the time period of 1 s was sufficient to allow activations r_i to reach equilibrium, or to allow one to reach 1.05, at which point the competition was curtailed. The first recognition unit that reached an activation of 1.05 was nominated the 'winning unit'. If no unit reached that level of activation, there was no winning unit for that list.

A.3. The order layer
Activations at the order layer were set to correspond with a primacy gradient with parameters equal to typical values used by Page & Norris (1998a). The gradient had a peak, P, equal to 11.5 (as above) and a decay rate of 0.78 (such that an activation of x would become an activation of $0.78x$ 1 s later).

A.4. The production layer
The production unit activations were equal to the order-unit activations, plus the relevant values of the output projected by the production unit associated with the winning unit at the recognition layer. That is, under circumstances in which there was a winning unit at the recognition layer, the corresponding production unit projected activation to other production units, where this activation was equal to the product of the activation of the winning production unit (1.05) and the weight of the outgoing connection

from this winner to the various production units, denoted $v_{\mathcal{J}i}$. For committed units, the net production unit activations were subjected to a repeated cycle of noisy pick-the-biggest, recall and suppression, with activations being compared with an omission threshold in each case (Page & Norris 1998a). The zero-mean choice noise had a standard deviation of 0.3, and the zero mean omission noise was 0.05. The omission threshold had a value of 0.1.

A.5. Learning

If, by 1 s after the forwarding of the final list item to the order layer, there was a winning unit at the recognition layer, learning was carried out at both the occurrence and production layers. First, the primacy gradient at the order layer was decayed by an estimated 'time since rehearsal' (4 s, for the seven-item lists here) plus a time equal to the time taken for the recognition layer competition to resolve (i.e. some time less than 1 s). This representation of order was then partially learned by the incoming weights *to* the winning occurrence unit, and the outgoing weights *from* the winning production unit.

The weights from other occurrence units, i, to the winning occurrence unit, $\mathcal{J}$, were adjusted such that

$$\Delta w_{i\mathcal{J}} = \lambda_o(Lg_i - w_{i\mathcal{J}}),$$

where λ_0 is the occurrence learning rate (0.1), L is a scaling factor equal to 10 and g_i is the primacy gradient activation at the order layer. The weights from the winning production unit, denoted $v_{\mathcal{J}i}$, were simply accumulated such that

$$\Delta v_{\mathcal{J}i} = \lambda_p g_i,$$

where $\lambda_p = 0.2$ is the production learning rate. In both cases, learning is assumed to cease once the largest incoming/outgoing weight reaches a particular large value, at which the associated units are deemed to be fully committed. In the simulations of the Hebb repetition effect presented here, no new unit reached this threshold for full commitment.

Simulations of ISR performance assumed that each item took 0.8 s to recall, and performance was assessed over 10 000 simulated trials.

REFERENCES

Acheson, D. J. & MacDonald, M. C. 2009 Twisting tongues and memories: explorations of the relationships between language production and verbal working memory. *J. Mem. Lang.* **60**, 329–350. (doi:10.1016/j.jml.2008.12.002)

Baddeley, A. D. 1968 How does acoustic similarity influence short-term memory? *Q. J. Exp. Psychol.* **20**, 249–263.

Baddeley, A. 1986 *Working memory*. New York, NY: Clarendon Press/Oxford University Press.

Baddeley, A. D. & Hitch, G. J. 1974 Working memory. In *Recent advances in learning and motivation*, vol. 8 (ed. G. H. Bower), pp. 47–90. New York, NY: Academic Press.

Baddeley, A., Gathercole, S. E. & Papagno, C. 1998 The phonological loop as a language learning device. *Psychol. Rev.* **105**, 158–173. (doi:10.1037/0033-295X.105.1.158)

Basso, A., Spinnler, H., Vallar, G. & Zanobio, M. E. 1982 Left hemisphere damage and selective impairment of auditory verbal short-term memory. A case study. *Neuropsychologia* **20**, 263–274. (doi:10.1016/0028-3932(82)90101-4)

Brown, G. D. A. & Hulme, C. 1996 Nonword repetition, STM, and word age-of-acquisition. In *Models of short-term memory* (ed. S. E. Gathercole), pp. 129–148. Hove, UK: Psychology Press.

Brown, G. D. A., Preece, T. & Hulme, C. 2000 Oscillator-based memory for serial order. *Psychol. Rev.* **107**, 127–181. (doi:10.1037/0033-295X.107.1.127)

Brown, G. D. A., Neath, I. & Chater, N. 2007 A temporal ratio model of memory. *Psychol. Rev.* **114**, 539–576. (doi:10.1037/0033-295X.114.3.539)

Burgess, N. & Hitch, G. J. 1992 Toward a network model of the articulatory loop. *J. Mem. Lang.* **31**, 429–460. (doi:10.1016/0749-596X(92)90022-P)

Burgess, N. & Hitch, G. J. 1999 Memory for serial order: a network model of the phonological loop and its timing. *Psychol. Rev.* **106**, 551–581. (doi:10.1037/0033-295X.106.3.551)

Burgess, N. & Hitch, G. J. 2006 A revised model of short-term memory and long-term learning of verbal sequences. *J. Mem. Lang.* **55**, 627–652. (doi:10.1016/j.jml.2006.08.005)

Cohen, M. A. & Grossberg, S. 1987 Masking fields: a massively parallel neural architecture for learning, recognizing, and predicting multiple groupings of patterned data. *Appl. Opt.* **26**, 1866–1891. (doi:10.1364/AO.26.001866)

Colle, H. A. & Welsh, A. 1976 Acoustic masking in primary memory. *J. Verbal Learn. Verbal Behav.* **15**, 17–31. (doi:10.1016/S0022-5371(76)90003-7)

Conrad, R. & Hull, A. J. 1964 Information, acoustic confusion and memory span. *Brit. J. Psychol.* **55**, 429–432.

Cumming, N., Page, M. P. A. & Norris, D. 2003 Testing a positional model of the Hebb effect. *Memory* **11**, 43–63. (doi:10.1080/741938175)

Cumming, N., Page, M. P. A., Norris, D., Hitch, G. & McNeil, A. M. Submitted. Repetition spacing and order competition effects in the Hebb repetition task.

De Renzi, E. & Nichelli, P. 1975 Verbal and nonverbal short-term memory impairment following hemispheric damage. *Cortex* **11**, 219–353.

Dollaghan, C. 1985 Child meets word: 'Fast mapping' in preschool children. *J. Speech Hear. Res.* **28**, 449–454.

Dollaghan, C. A. 1987 Fast mapping in normal and language-impaired children. *J. Speech Hear. Disord.* **52**, 218–222.

Ellis, A. W. 1980 Errors in speech and short-term memory: the effects of phonemic similarity and syllable position. *J. Verbal Learn. Verbal Behav.* **19**, 624–634. (doi:10.1016/S0022-5371(80)90672-6)

Farrell, S. & Lewandowsky, S. 2002 An endogenous distributed model of ordering in serial recall. *Psychon. Bull. Rev.* **9**, 59–79.

Farrell, S. & Lewandowsky, S. 2003 Dissimilar items benefit from phonological similarity in serial recall. *J. Exp. Psychol. Learn. Mem. Cogn.* **29**, 838–849. (doi:10.1037/0278-7393.29.5.838)

Gathercole, S. E. & Baddeley, A. D. 1989 Evaluation of the role of phonological STM in the development of vocabulary in children: a longitudinal study. *J. Mem. Lang.* **28**, 200–213. (doi:10.1016/0749-596X(89)90044-2)

Gathercole, S. E., Willis, C., Emslie, H. & Baddeley, A. D. 1992 Phonological memory and vocabulary development during the early school years: a longitudinal study.

Dev. Psychol. **28**, 887–898. (doi:10.1037/0012-1649.28.5.887)

Gathercole, S. E., Hitch, G. J., Service, E. & Martin, A. J. 1997 Short-term memory and new word learning in children. *Dev. Psychol.* **33**, 966–979. (doi:10.1037/0012-1649.33.6.966)

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A.-M. & Martin, A. J. 1999 Phonological short-term memory and vocabulary development: further evidence on the nature of the relationship. *Appl. Cogn. Psychol.* **13**, 65–77. (doi:10.1002/(SICI)1099-0720(199902)13:1<65::AID-ACP548>3.0.CO;2-O)

Grossberg, S. 1978 Behavioral contrast in short-term memory: serial binary memory models or parallel continuous memory models. *J. Math. Psychol.* **17**, 199–219. (doi:10.1016/0022-2496(78)90016-0)

Gupta, P. 1996 Word learning and verbal short-term memory: a computational account. In *Proc. of the Eighteenth Annual Meeting of the Cognitive Science Society* (ed. G. W. Cottrell), pp. 189–194. Mahwah, NJ: Lawrence Erlbaum.

Gupta, P. 2002 Are nonwords lists? The effects of syllable serial position on repetition of individual nonwords. In *Abstracts of the Quebec'02 Conference on Short-term/Working Memory, Quebec City*: Quebec.

Gupta, P. 2003 Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *Q. J. Exp. Psychol. (A)* **56**, 1213–1236.

Gupta, P. 2005 Primacy and recency in nonword repetition. *Memory* **13**, 318–324. (doi:10.1080/09658210344000350)

Gupta, P. 2009 A computational model of nonword repetition, immediate serial recall, and nonword learning. In *Interactions between short-term and long-term memory in the verbal domain* (eds A. Thorn & M. Page). Hove, UK: Psychology Press.

Gupta, P. & MacWhinney, B. 1997 Vocabulary acquisition and verbal short-term memory: computational and neural bases. *Brain Lang.* **59**, 267–333. (doi:10.1006/brln.1997.1819)

Gupta, P. & Tisdale, J. 2009 Word learning, phonological short-term memory, phonotactic probability and long-term memory: towards an integrated framework. *Phil. Trans. R. Soc. B* **364**, 3755–3771. (doi:10.1098/rstb.2009.0132)

Gupta, P., Lipinski, J., Abbs, B. & Lin, H. 2005 Serial position effects in nonword repetition. *J. Mem. Lang.* **53**, 141–162. (doi:10.1016/j.jml.2004.12.002)

Hartley, T. & Houghton, G. 1996 A linguistically constrained model of short-term memory for nonwords. *J. Mem. Lang.* **35**, 1–31. (doi:10.1006/jmla.1996.0001)

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

Henson, R. N. A. 1998 Short-term memory for serial order: the start-end model. *Cogn. Psychol.* **36**, 73–137. (doi:10.1006/cogp.1998.0685)

Henson, R. N. A., Norris, D. G., Page, M. P. A. & Baddeley, A. D. 1996 Unchained memory: error patterns rule out chaining models of immediate serial recall. *Q. J. Exp. Psychol.* **49A**, 80–115.

Hitch, G. J., Fastame, M. C. & Flude, B. 2005 How is the serial order of a verbal sequence coded? Some comparisons between models. *Memory* **13**, 247–258. (doi:10.1080/09658210344000314)

Houghton, G. 1990 The problem of serial order: a neural network model of sequence learning and recall. In *Current research in natural language generation* (eds R. Dale, C. Mellish & M. Zock). New York, NY: Academic Press.

Johnson, N. F. 1970 The role of chunking and organization in the process of recall. In *Psychology of learning and motivation* (ed. G. H. Bower), New York, NY: Academic Press.

Jones, D. M. & Macken, W. J. 1995 Phonological similarity in the irrelevant speech effect. Within- or between-stream similarity? *J. Exp. Psychol. Learn. Mem. Cogn.* **21**, 103–133. (doi:10.1037/0278-7393.21.1.103)

Majerus, S., Poncelet, M., Elsen, B. & Van der Linden, M. 2006 Exploring the relationship between verbal short-term memory for serial order and item information and new word learning in adults. *Eur. J. Cogn. Psychol.* **18**, 848–873. (doi:10.1080/09541440500446476)

Marshall, J. A. 1990. In A self-organizing scale-sensitive neural network. *Proc. Int. Joint Conf. on Neural Networks, June 1990, San Diego*, pp. 649–654.

Marslen-Wilson, W. & Tyler, L. K. 1980 The temporal structure of spoken language understanding. *Cognition* **8**, 1–71. (doi:10.1016/0010-0277(80)90015-3)

McKelvie, S. J. 1987 Learning and awareness in the Hebb digits task. *J. Gen. Psychol.* **114**, 75–88.

Melton, A. W. 1963 Implications of short-term memory for a general theory of memory. *J. Verbal Learn. Verbal Behav.* **2**, 1–21. (doi:10.1016/S0022-5371(63)80063-8)

Miller, G. A. 1956 The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* **63**, 81–97. (doi:10.1037/h0043158)

Mosse, E. K. & Jarrold, C. 2008 Hebb learning, verbal short-term memory, and the acquisition of phonological forms in children. *Q. J. Exp. Psychol.* **61**, 505–514. (doi:10.1080/17470210701680779)

Neath, I. 2000 Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin and Review* **7**, 403–423.

Nigrin, A. 1993 *Neural networks for pattern recognition*. Boston, MA: MIT Press.

Norris, D. 1994 Shortlist: a connectionist model of continuous speech recognition. *Cognition* **52**, 189–234. (doi:10.1016/0010-0277(94)90043-4)

Page, M. P. A. 1993. *Modelling aspects of music perception using self-organizing neural networks*. Unpublished PhD thesis, University of Cardiff, Wales.

Page, M. P. A. 1994 Modelling the perception of musical sequences with self-organizing neural networks. *Connect. Sci.* **6**, 223–246. (doi:10.1080/09540099408915725)

Page, M. P. A. 2000 Connectionist modelling in psychology: a localist manifesto. *Behav. Brain Sci.* **23**, 443–512. (doi:10.1017/S0140525X00003356)

Page, M. P. A. & Norris, D. 1998a The primacy model: a new model of immediate serial recall. *Psychol. Rev.* **105**, 761–781. (doi:10.1037/0033-295X.105.4.761-781)

Page, M. P. A. & Norris, D. 1998b Modeling immediate serial recall with a localist implementation of the primacy model. In *Localist connectionist approaches to human cognition* (eds A. M. Jacobs & J. Grainger), pp. 227–255. Mahwah, NJ: Lawrence Erlbaum Associates.

Page, M. P. A. & Norris, D. 2009 Is there a common mechanism underlying word-form learning and the Hebb repetition effect? Experimental data and a modelling framework. In *Interactions between short-term and long-term memory in the verbal domain* (eds A. Thorn & M. Page). Hove, UK: Psychology Press.

Page, M. P. A., Cumming, N., Norris, D., Hitch, G. J. & McNeil, A. M. Submitted. Exploring the importance of list recall in the Hebb repetition effect.

Page, M. P. A., Cumming, N., Norris, D., Hitch, G. J. & McNeil, A. M. 2006 Repetition learning in the immediate serial recall of visual and auditory materials. *J. Exp. Psychol. Learn. Mem. Cogn.* **32**, 716–733. (doi:10.1037/0278-7393.32.4.716)

Page, M. P. A., Madge, A., Cumming, N. & Norris, D. G. 2007 Speech errors and the phonological similarity

effect in short-term memory: evidence suggesting a common locus. *J. Mem. Lang.* **56**, 49–64. (doi:10.1016/j.jml.2006.09.002)

Saffran, J. R., Aslin, R. N. & Newport, E. L. 1996 Statistical learning by 8-month old infants. *Science* **274**, 1926–1928. (doi:10.1126/science.274.5294.1926)

Salamé, P. & Baddeley, A. D. 1982 Disruption of short-term memory by unattended speech: implications for the structure of working memory. *J. Verbal Learn. Verbal Behav.* **21**, 150–164. (doi:10.1016/S0022-5371(82)90521-7)

Salamé, P. & Baddeley, A. D. 1986 Phonological factors in STM: similarity and the unattended speech effect. *Bull. Psychon. Soc.* **24**, 263–265.

Stadler, M. A. 1993 Implicit serial learning: questions inspired by Hebb (1961). *Mem. Cogn.* **21**, 819–827.

Szmalec, A., Duyck, W., Vandierendonck, A., Mata, A. B. & Page, M. P. A. 2009 The Hebb repetition effect as a laboratory analogue of novel word learning. *Q. J. Exp. Psychol.* **62**, 435–443. (doi:10.1080/17470210802386375)

Tremblay, S., Macken, W. J. & Jones, D. M. 2000 Elimination of the word length effect by irrelevant sound revisited. *Mem. Cogn.* **28**, 841–846.

Trojano, L. & Grossi, D. 1995 Phonological and lexical coding in verbal short-term memory and learning. *Brain Cogn.* **21**, 336–354.

Warrington, E. K. & Shallice, T. 1969 The selective impairment of auditory verbal short-term memory. *Brain* **92**, 885–896. (doi:10.1093/brain/92.4.885)