

# From Numerosity to Ordinal Rank: A Gain-Field Model of Serial Order Representation in Cortical Working Memory

Matthew Botvinick<sup>1</sup> and Takamitsu Watanabe<sup>2</sup>

<sup>1</sup>Psychology Department and Institute for Neuroscience, Princeton University, Princeton, New Jersey, 08540, and <sup>2</sup>Department of Physiology, University of Tokyo School of Medicine, Tokyo 113-0033, Japan

Encoding the serial order of events is an essential function of working memory, but one whose neural basis is not yet well understood. In the present work, we advance a new model of how serial order is represented in working memory. Our approach is predicated on three key findings from neurophysiological research: (1) prefrontal neurons that code conjunctively for item and order, (2) parietal neurons that represent count information through a graded and compressive code, and (3) multiplicative gain modulation as a mechanism for information integration. We used an artificial neural network, integrating across these three findings, to simulate human immediate serial recall performance. The model reproduced a core set of benchmark empirical findings, including primacy and recency effects, transposition gradients, effects of interitem similarity, and developmental effects. The model moves beyond previous accounts by bridging between neuroscientific findings and detailed behavioral data, and gives rise to several testable predictions.

**Key words:** prefrontal cortex; parietal cortex; working memory; serial order; computational models; numerosity

## Introduction

Working memory is a cognitive function that serves to preserve task-relevant information in an active and accessible form over periods of a few seconds (Baddeley, 1986; Jonides et al., 2005). It has long been recognized that one critical feature of working memory is its capacity to encode and maintain information about the serial order of perceived events (Marshuetz, 2005). This capacity is essential in many domains including the comprehension, learning, and production of action sequences, the encoding of causal relationships, and perhaps above all, language processing (Martin and Gupta, 2004).

The ability to recall serial order information from working memory, and the limits of this ability, have been studied by cognitive psychologists for decades, and this research effort has yielded an exceedingly detailed description of human serial recall performance. However, the neural mechanisms underlying the behavioral data are not yet fully understood. Although several neuroscientific models have been proposed previously (Dominey et al., 1995, 1997; Beiser and Houk, 1998; O'Reilly and Soto, 2001), few have made contact with behavioral data at any level of detail. At the same time, where psychologically sophisticated models have been offered, they have rarely made significant contact with evidence from neuroscience (Houghton, 1990; Burgess and Hitch, 1999; Brown et al., 2000; Farrell and Lewandowsky, 2002; Botvinick and Plaut, 2006).

In the present study, we introduce a novel computational

model of working memory for serial order, which bridges between the domains of neuroscience and behavior. The model is based directly on a set of recent neuroscientific findings and shows how these observations, when integrated into a single account, might explain detailed patterns of serial recall performance. In what follows, we begin by reviewing the neuroscientific data on which the model is founded, and then report a series of simulation studies in which the model was tested against empirical benchmarks from the behavioral literature.

## Elements of the account

### (1) *Conjunctive coding of item and rank information in prefrontal cortex*

The first basic finding that our model draws on comes from single-unit recordings in monkeys performing immediate serial recall and related tasks. Across a series of such studies, beginning with Barone and Joseph (1989) and continued most recently by Inoue and Mikami (2006) (see also Kermadi et al., 1993; Kermadi and Joseph, 1995; Funahashi et al., 1997; Ninokura et al., 2003, 2004), a critical and consistent finding has been that sequences are encoded through a conjunctive code, which crosses item with order information.<sup>a</sup> Specifically, within the prefrontal cortex as well as caudate nucleus, single neurons have been found to respond selectively to particular items (shapes or locations), but their response to these items depends on the ordinal position in which the items appear (see Fig. 1A). The representational code carried by these neurons is conjunctive in the sense that the neurons respond maximally to a particular conjunction or combination of item and ordinal position. Such conjunctive coding pro-

Received Feb. 2, 2007; revised June 8, 2007; accepted June 11, 2007.

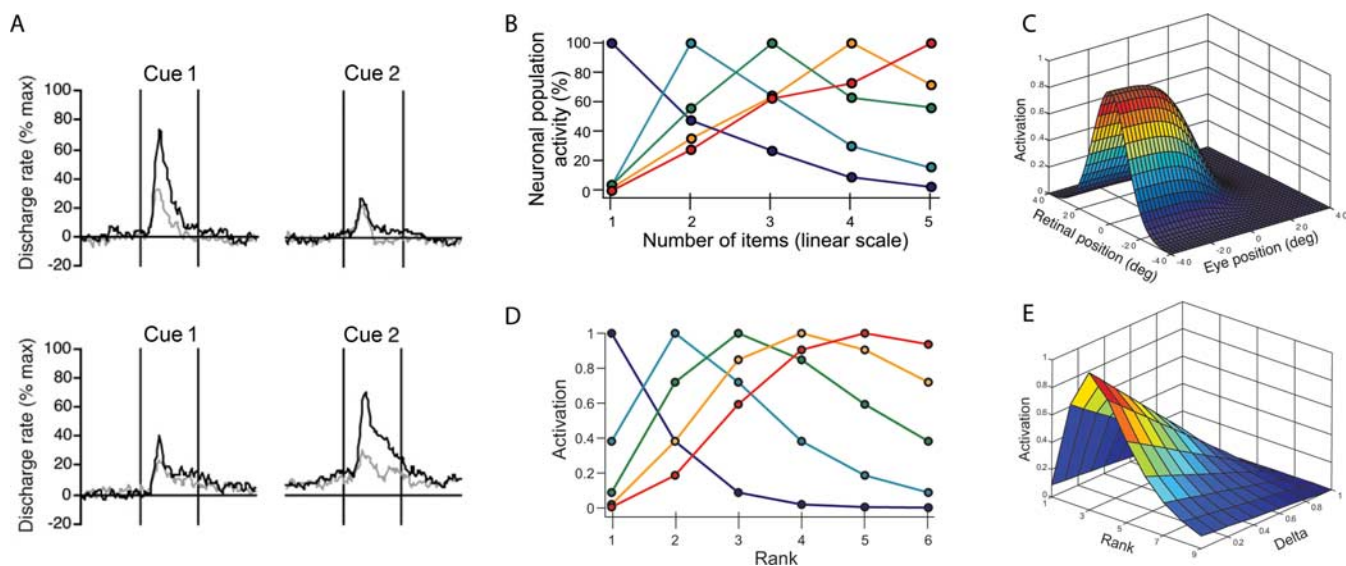
This work was supported by National Institutes of Health Grant MH16804 (M.B.) and the University of Tokyo International Academic Exchange Grant Program (T.W.).

Correspondence should be addressed to Matthew Botvinick, Princeton University, Psychology Department, 3-C-10 Green Hall, Princeton, NJ 08540. E-mail: matthewb@princeton.edu.

DOI:10.1523/JNEUROSCI.2110-07.2007

Copyright © 2007 Society for Neuroscience 0270-6474/07/278636-07\$15.00/0

<sup>a</sup>Although our focus in the present work is on activation-based mechanisms for serial order memory centered in the prefrontal cortex, it is important to acknowledge evidence that memory for serial order information may also depend on long-term memory mechanisms housed in medial temporal lobe structures (Fortin et al., 2002).



**Figure 1.** *A*, Response profiles for two prefrontal neurons, reported by Inoue and Mikami (2006), during sequential presentation of two visual shape cues. Both neurons displayed differential responses to preferred (black) and nonpreferred (gray) shapes, as well as differential responses across ordinal positions. The neuron contributing to the top panels responded preferentially to first-rank items, and the neuron in the bottom panels to second-rank items. Although these particular units did not display sustained activation, such activation was observed in other units within the same region. *B*, Response profiles of IPS neurons with graded, compressed responses to number, from Nieder (2005). Individual traces correspond to neurons with different preferred numerosities, as indicated by the legend. *C*, Illustration of the gain-field representations used by units in the computational model of Pouget and Sejnowski (1997). Activity is plotted for a single unit, with multiplicatively interacting receptive fields for eye-centered stimulus position and eye position [redrawn from Pouget and Snyder (2000)]. *D*, Graded, compressive representation of rank in the model. Shown are the response profiles of the first five rank units (preferred ranks 1–5) to items presented at ordinal positions 1–6. *E*, Response profile of an internal unit in the model. Delta indexes the degree of dissimilarity between the current input item and the unit's most preferred item, with zero being a precise match. The unit displays multiplicatively interacting receptive fields for item and rank, responding maximally when its preferred item occurs at rank three.

vides an answer to the question of how the brain may solve the binding problem inherent to sequence encoding, the need to link individual items with individual serial positions.

### (2) Information integration through gain-field encoding

The second key finding derives from single-unit recording studies that suggest how, in general, the brain may compute conjunctive codes. Starting from studies on spatial coordinate transformations in vision, Salinas et al. (Salinas and Thier, 2000; Salinas and Abbott, 2001) have proposed that information from multiple domains is commonly integrated, at the neural level, through multiplicative gain modulation. For example, in spatial processing, information about retinotopic location and eye position is integrated to yield head- or eye-centered representations. Single-unit recording data suggest that this mapping is mediated by parietal neurons whose response profiles can be modeled as the product of two receptive fields, one for retinotopic position and one for eye position (Brotchie et al., 1995). The sufficiency of this mechanism was demonstrated in a neural network model by Pouget and Sejnowski (1997) (see Fig. 1*C*). Additional computational studies have indicated how the same kind of gain modulation might support information integration in additional domains, including object recognition and sensorimotor mapping (Pouget and Snyder, 2000; Salinas and Thier, 2000; Salinas and Abbott, 2001; Salinas, 2004).

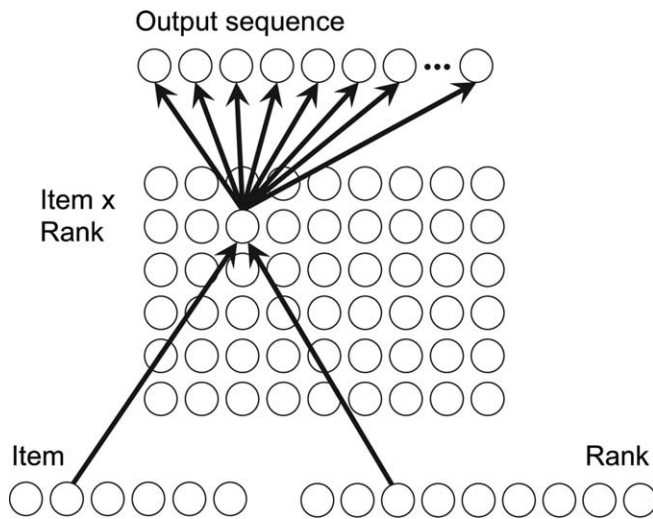
### (3) Graded, compressive representations of sequential numerosity in intraparietal sulcus

The third finding of interest bears on the question of how serial position or rank may be represented at the neural level. It has been proposed, based in part on neuroimaging data, that serial order processing may draw on representations of number arising within the intraparietal sulcus (IPS) (Marshuetz et al., 2000, 2006; Marshuetz, 2005; Nieder, 2005). Previous single-unit recording

work by Nieder et al. (2006) provides additional motivation for this idea, by demonstrating that neurons in the IPS respond selectively to the number of occurrences of a repeating event, with a distinct subset of neurons responding preferentially to the event's first occurrence, another subset to its second occurrence, and so forth. Nieder et al. (2006) described these neurons as coding for "sequential numerosity" or "sequential quantity." In what follows, for brevity, we describe such neurons as coding for rank.

Importantly, the study by Nieder et al. (2006), together with closely related work (Nieder et al., 2002; Nieder and Miller, 2003, 2004; Nieder, 2005), provides detailed information concerning the format of rank representations within the IPS. First, Nieder et al. (2006) found that IPS neurons code for rank in a graded manner; individual neurons responded maximally to a specific rank, but also responded more weakly to other ranks, with the response dropping off in intensity with distance from the preferred rank. Second, closely related work on numerosity representation (Nieder and Miller, 2003) indicates that IPS neurons represent count information using a compressive code, reflected in more broadly tuned receptive fields for larger numbers (see Fig. 1*B*). As Nieder and Miller (2003) and others have noted, such compressive coding provides an explanation for the so-called scalar property, an instance of Weber's law according to which better discrimination is shown between small numerosities than between larger ones.

Our central proposal is that these three findings (conjunctive coding of item and rank, information integration through multiplicative gain modulation, and graded, compressive coding of count information) can be fit together to provide a satisfying account of how serial order is represented in working memory. According to this account, during sequence encoding, graded and compressive rank representations arising within the IPS feed



**Figure 2.** Structure of the model showing connections to and from one internal unit.

forward to the prefrontal cortex, where rank information is integrated with item information through multiplicative gain modulation. The resulting graded conjunctive representation in the prefrontal cortex provides the basis for serial recall.

### Neural network implementation

To make this account explicit, and to evaluate its ability to account for human recall performance, we implemented the account in the form of a runnable neural network model. The structure of the network was based directly on the gain field model of visual processing proposed by Pouget and Sejnowski (1997). Like that model, ours was composed of interconnected processing units, which assumed scalar activation values representing the time-averaged spike rates of individual neurons.<sup>b</sup> These were organized into four layers or groups (see Fig. 2). There were two input layers, one representing item (e.g., shape, location, or verbal item), and the other representing ordinal position or rank. As detailed below, each unit in the item layer responded maximally to a specific and unique item, but also responded submaximally to other items, to an extent determined by those items' similarity to the unit's optimal stimulus. The response profiles for units in the rank layer were chosen so as to resemble those reported by Nieder et al. (Nieder, 2005; Nieder et al., 2006). Specifically, each unit responded maximally to a unique rank, but also showed graded responses to surrounding ranks. This, as well as the compressive quality of empirically observed encodings of number, was captured by making each unit's response a scaled log-normal function of rank (see Fig. 1D).

Both input layers sent projections to an internal layer. Each unit within this layer received connections from one unit in the item layer and one unit in the rank layer, and assumed a level of activation equal to the product of the activations of these two input units (see Fig. 1E). All units in the internal layer sent projections to each unit in an output layer, within which each unit coded for a specific response sequence (see Materials and Methods).

The model was used to simulate immediate serial recall for six-item sequences. The first item in the target sequence was presented by imposing the appropriate patterns of activation over

the item and rank input layers. Activations in the internal layer were updated, based on these inputs. The second item in the target sequence was presented on the next time step by imposing new patterns of activation over the input layers. The pattern of internal-layer activation induced by these inputs was added to the pattern induced by the first item in the sequence. This summation implemented the assumption that sequence elements are represented through a superpositional, activation-based code, as argued by Botvinick and Plaut (2006) (Beiser and Houk, 1998; O'Reilly and Soto, 2001). Empirical support for such a superpositional code proceeds from neurophysiological studies such as those by Inoue and Mikami (2006) and Mushiaki et al. (2006), both of which reported representation of sequences through concurrent activation of prefrontal neurons coding conjunctively for item and rank.

Subsequent presentation of the third through sixth items resulted in a distributed pattern of activation in the internal layer that contained information pertaining to all six items in the target sequence (see Fig. 3A,B). With this pattern in place, activation fed forward from the internal layer to the output layer. The synaptic weights connecting these layers were trained, using supervised gradient-descent learning, to activate the output unit representing the target sequence (see Materials and Methods).

On each step of processing, random noise was added to the activation value of each unit in the input and internal layers, modeling the intrinsic variability of activation codes in biological neurons. The introduction of this noise meant that the model's internal representation for any given target sequence might "accidentally" end up looking like the pattern usually used to represent a different sequence, causing the model to commit a recall error (see Fig. 3C).

Using procedures detailed in the following section, we used the model to simulate immediate serial recall under a range of conditions, evaluating its ability to capture a key set of behavioral benchmarks.

### Materials and Methods

Simulations were implemented using Matlab (Mathworks, Natick, MA).

**Model specifications.** The model comprised six item units, nine rank units, 54 internal units, and 720 output units. Each item unit was associated with an optimal stimulus ( $\psi$ ) and unit activation was determined according to a function of this item and the item actually occurring as a stimulus,  $s$ :

$$I_{\psi}(s) = \begin{cases} 1, & \text{if } s = \psi \\ 1 - \delta, & \text{if } s \neq \psi \end{cases} \quad (1)$$

where  $I_{\psi}(s)$  is the activation of the input unit with optimal stimulus  $\psi$  in response to stimulus item  $s$ , and  $\delta$  is a model parameter controlling the degree of dissimilarity between item representations (range, 0–1). For the simulation involving mixed confusable and nonconfusable items, input items were divided into two groups: for alternating lists, one group of three confusable items and one group of three nonconfusable items; for isolate lists, one group of five confusable items and one separate nonconfusable item. Confusable items were assumed to differ by  $\delta_C$  and nonconfusable by  $\delta_N$ . The separation between of confusable and nonconfusable items was determined by a third parameter,  $\delta_{NC}$ .

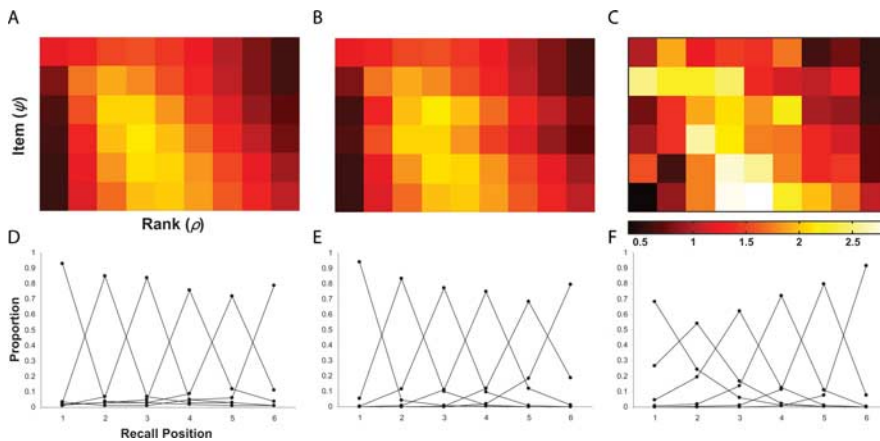
Each rank unit was assumed to be activated maximally by a specific rank  $\rho$ , and to assume an activation based on this rank and the rank actually being encoded ( $r$ ), according to the scaled log normal function:

$$R_{\rho}(r) = \exp\left(-\frac{(\ln r - \ln \rho)^2}{2\sigma^2}\right), \quad (2)$$

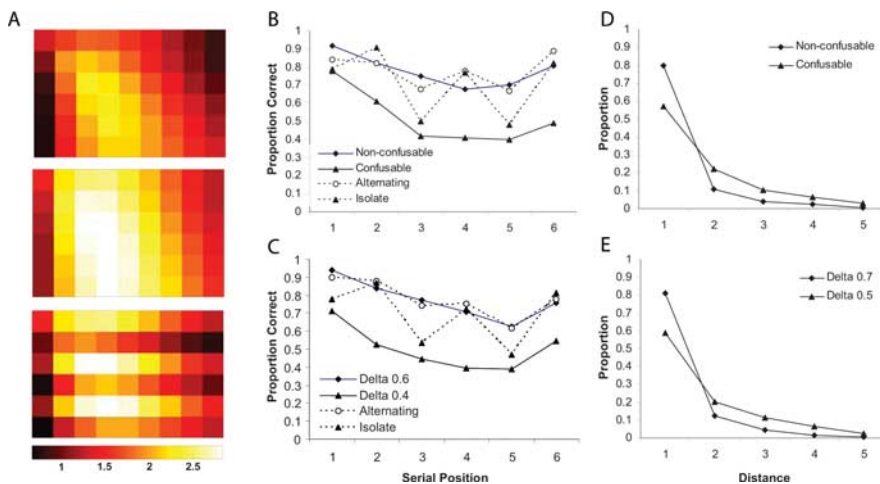
where  $R_{\rho}(r)$  is the activation of the rank unit with preferred rank  $\rho$ , during presentation of the item at rank  $r$ . As shown in Figure 1D, this

<sup>b</sup>As in the model of Pouget and Sejnowski (1997), no effort was made to capture differences in overall firing rates between cortical regions (e.g., between the IPS and prefrontal cortex). Such an undertaking would face the problem that spike rates in the relevant empirical studies have tended to be reported only in normalized form.





**Figure 3.** *A*, Pattern of activation over the internal units of the model, representing the sequence 123456, where the numbers correspond to the items preferred by item units 1–6. Each cell corresponds to a single unit in the model, with units in each row sharing a preferred item (1–6, counting from the top) and units in each column sharing a preferred rank (1–9, counting from the left). For clarity, the contribution of noise is omitted. *B*, Pattern of activation representing the sequence 124356 (noise omitted). *C*, Pattern generated by the input sequence 123456, including noise, on a trial when the response was 124356. *D*, Positional recall data from an empirical study by Henson (1998). Each trace shows the proportion of trials on which items from a single input position were recalled at each output position. *E*, Positional recall from the model. Root mean squared error of fit, 0.036. Parameters for all data shown are  $\sigma = 0.5$ ,  $\delta = 0.6$ , and  $\nu = 0.09$ . *F*, Pattern of positional recall from the model, when invariant and symmetrical Gaussian response profiles were used for the rank units.



**Figure 4.** *A*, Activation patterns representing sequences of dissimilar items ( $\delta = 0.6$ , top), similar items ( $\delta = 0.4$ , center) and a combination (similar items at positions 1, 2, 3, 5 and 6,  $\delta_N = 0.6$ ,  $\delta_C = 0.4$ ,  $\delta_{NC} = 0.65$ ; bottom). Remaining parameters are  $\sigma = 0.5$ ,  $\nu = 0$ . *B*, Positional accuracy data from an experiment by Farrell and Lewandowsky (2003) for pure lists of phonologically confusable and nonconfusable items, alternating lists with confusable items at odd ranks, and lists of confusable items with one distinctive item (“isolate”) at position 2, 4, or 6. Data from the latter list type is summarized by showing recall for isolate items from all positions in one data series. *C*, Corresponding performance pattern from the model. Parameters as listed for left panels, with  $\nu = 0.08$ . Root mean squared error (RMSE), 0.049. *D*, Transposition gradients from an empirical study by Henson (1996), showing the proportion of transposition errors involving displacements of one to five positions for six-item lists of phonologically confusable or nonconfusable items. *E*, Corresponding simulation data. Parameters are  $\sigma = 0.3$ ,  $\delta = 0.5, 0.7$ ,  $\nu = 0.2$ . RMSE, 0.011.

function leads to graded, compressive response profiles resembling those reported by Nieder et al. (2002, 2006; Nieder and Miller, 2003, 2004) (see Fig. 1*B*), graded in the sense that rank units respond maximally to a particular rank, but also submaximally to other ranks, and compressive in the sense that unit tuning curves broaden with increasing rank.

Each unit in the internal layer took inputs from a unique pair of item and rank units, and assumed an activation value based on the product of their activation values:

$$\Delta h_{\rho\psi} = R_{\rho}(r)I_{\psi}(s), \quad (3)$$

where  $h_{\rho\psi}$  is the activation of the internal unit receiving input from the rank unit with preferred rank  $\rho$  and the item unit with preferred stimulus

$\psi$ . The  $\delta$  symbol indicates that internal unit activation was augmented by the indicated activation product on each step of encoding. At each step of encoding, multiplicative noise, with SD  $\nu$ , was applied to input and internal layers.

Each output unit represented a unique ordering of the six items represented in the item input layer. Every output unit received inputs from all internal units. At the end of encoding, the activation of each output unit was set according to the softmax function:

$$o_i = \frac{\exp(a_i)}{\sum_j \exp(a_j)}, \quad (4)$$

where  $a_i$  is the net input to unit  $i$ , determined by the activations of the internal units ( $h_j$ ) and the intervening connection weights  $w_{ij}$ :

$$a_i = \sum_j h_j w_{ij}. \quad (5)$$

The task simulated was immediate forward recall for six-item sequences.<sup>c</sup> The target sequences always included the same six items. At the onset of each new trial, all unit activations were set to zero. Presentation of target items then proceeded as described above. After presentation of the sixth list item, the output layer was updated, and its most active output unit identified the output sequence.

**Training.** Internal to output weights were set initially to 0. All 720 possible target lists were presented in random order, without replacement, and after each trial the internal to output weights were adjusted using the  $\delta$  rule:

$$\Delta w_{ij} = \alpha(t_i - o_i)h_j, \quad (6)$$

where  $\alpha$  is a learning rate, and  $t_i$  is the target value for output unit  $i$  for the present target list (1 for the output unit representing the target list, otherwise 0). The learning rate was dynamically adjusted to minimize the training duration, which was truncated at 500 cycles through the training set. However, essentially identical results were obtained with a fixed learning rate of 0.001 and a fixed training duration of 2500 cycles. The noise parameter  $\nu$  was set to zero during training.

**Testing.** To evaluate performance under a given set of parameters, the model was tested 50 times on each sequence in the training set, and average positional accuracy was computed. In addressing each behavioral benchmark, parameters minimizing root mean squared error were sought through grid search over the model’s three free parameters  $\sigma$ ,  $\delta$ , and  $\nu$  (in the mixed-list simulation, the five parameters  $\sigma$ ,  $\delta_C$ ,  $\delta_N$ ,  $\delta_{NC}$  and  $\nu$ ).

## Results

### Positional accuracy

In behavioral studies of serial recall, plotting recall accuracy by serial position typically results in a “bow-shaped” curve (Fig. 3*D*), reflecting a recall advantage for initial items (the primacy effect) and a smaller advantage for the last one or two items (the recency

<sup>c</sup>Rank units with preferred ranks larger than six were included in the model because, given the graded nature of the rank code, such units naturally contribute to the representation of six-item sequences.

effect). The positional recall accuracy of the model displayed this same profile, as shown in Figure 3*E*. This pattern of performance stems from two factors. Both the primacy and recency effects derive from edge effects, because there are fewer opportunities for items at the boundaries of the sequence to exchange positions with near neighbors. The primacy effect derives, additionally, from the greater distinctiveness of items at the beginning of the list, driven by the compressive rank code of the model.<sup>d</sup> The contribution of this factor can be seen by comparing Figure 3, *E* and *F*. Figure 3*F* illustrates the performance of the model when ordinary Gaussian rather than log-normal rank codes are used, eliminating the broadening of tuning curves with increasing rank. As comparison with *E* makes clear, this change to the model significantly reduces the magnitude and extent of the primacy effect.

### Transposition gradients

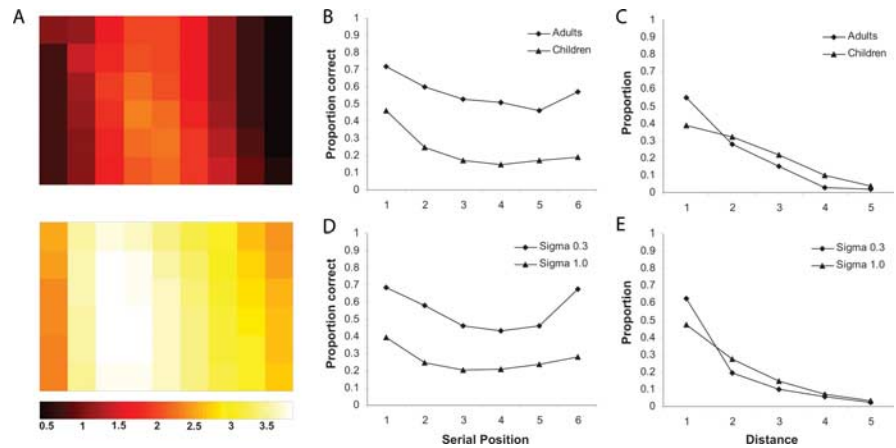
Another consistent finding from behavioral studies of serial recall is that when an item is recalled at the incorrect serial position (a transposition error), its recall position is likely to lie near its original position. As shown in Figures 3 and 4, the model's recall performance displayed this same property. This aspect of the model's behavior derives from the similarity structure of its internal representations. As a result of the form of the rank representations of the model, items in nearby ordinal positions are represented more similarly than items in more widely separated positions, a factor that makes it relatively common for the model to confuse the locations of closely spaced items.

### Effects of interitem similarity

In behavioral studies, when sequence items are highly confusable (e.g., phonologically similar in verbal recall), recall performance is undermined (Fig. 4*B*). Conrad (1965) showed that this is attributable in part to an increase in the number of transposition errors when items are confusable. Moreover, transpositions in confusable lists are prone to span wider lags than in nonconfusable lists (Henson, 1996) (Fig. 4*D*). The performance of the model displayed these same effects (Fig. 4*C,E*). Variations in interitem similarity were simulated by varying the degree of overlap between activation patterns in the model's item input layer (see Materials and Methods) (Fig. 4*A*). Increasing interitem similarity reduced recall by increasing the number of transpositions, and increased the tendency of items to transpose across relatively wide lags. Once again, the model's performance can be understood in terms of the similarity relations among its internal representations. The internal representations of two different list orderings are more similar, and therefore more confusable, when items are relatively highly overlapping than when they overlap less.

### Mixed lists

Another behavioral finding that has received a great deal of recent emphasis involves recall for sequences of highly similar items



**Figure 5.** *A*, Representations of sequence 123456 in simulations addressing performance of adults (top;  $\sigma = 0.3$ ) and children (bottom;  $\sigma = 1.0$ ). Remaining parameters are  $\delta = 0.3$ ,  $\nu = 0$ . *B*, *C*, Positional recall accuracy (*B*) and transposition gradients (*C*) for children and adults, from a study by McCormack et al. (2000). *D*, *E*, Corresponding pattern from the model with parameters as just listed, except  $\nu = 0.09$ . Root mean squared error for all 22 data points, 0.059.

(e.g., in verbal recall, the phonologically related letters B, P, T, C, G) that contain one or more distinctive items, for example, BPTRCG or BRPMTL. The general finding is that the distinctive or “nonconfusable” items within such mixed lists are recalled as well or better than when the same items appear among other nonconfusable items (e.g., JRYMQL) (Fig. 4*B*). Varying the degree of overlap among the model's item representations to simulate the presentation of mixed lists (see Materials and Methods) (Fig. 4*A*) yielded a comparable pattern of recall performance (Fig. 4*C*).

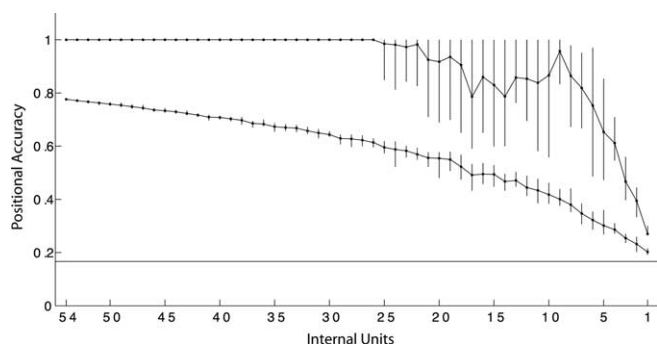
### Development

Another benchmark behavioral finding pertains to recall performance among children versus adults. Not surprisingly, recall accuracy improves with age. A more informative finding is that the transposition curve becomes steeper with age, that is, transpositions tend to span smaller lags (McCormack et al., 2000) (Fig. 5*B,C*). This effect has been proposed to derive from a progressive sharpening of neural rank representations over the course of development (Lipton and Spelke, 2003). We simulated this by varying the breadth of tuning among the rank input units in the model (see Materials and Methods) (Fig. 5*A*). Relatively broad tuning yielded recall performance resembling that observed among children (Fig. 5*D,E*).

### Representational capacity

One possible objection to the account implemented in the model is that it would seem, in the general case, to require a prohibitively large number of processing units. It is often assumed that conjunctive representational regimes scale poorly, because of the problem of combinatorial explosion. However, O'Reilly et al. (O'Reilly and Busby, 2002; O'Reilly et al., 2003) have demonstrated that this assumption is not generally warranted. In the present model, the use of conjunctive representations of item and order might appear to require at least  $I \times R$  units, where  $I$  is the number of distinct items to be represented and  $R$  is the number of distinct ranks. However, as shown in Figure 6 (blue data series), the present model can recall six-item lists when equipped with  $<36$  internal units. As in the theoretical account provided by O'Reilly et al. (2002, 2003), the present model's ability to function with only a subset of its internal units is attributable to its use of coarse conjunctive representations, within which any given unit

<sup>d</sup>Another consequence of this factor is that exchanges between adjacent items become more frequent with increasing rank. Thus, although the format of the data in Figure 3*F* does not make it evident, the model is less prone to exchange items at positions 2 and 3 than it is to exchange items 4 and 5.



**Figure 6.** Mean positional accuracy displayed by the model when trained with a varying number of internal units and tested both without the injection of noise (top series) and with noise (bottom series). In each simulation at a given unit count, units were selected at random and removed until the target unit count was attained. Error bars show the range of accuracies across 10 simulations. The horizontal line indicates the level of chance performance. Parameters as listed in the caption to Figure 1.

carries information about a range of item-rank pairings. The redundancy inherent in the use of such coarse coding also means that the model can continue to perform accurately if a small number of units are removed after training (data not shown). Another important consequence is that, although the model can function correctly with relatively few internal units, increasing the number of internal units results in performance that is more robust to noise. This is shown in Figure 6 (lower data series), which shows the model's performance under noise across a range of internal layer sizes.

### Alternative implementations

Very similar results were obtained with an implementation of the model in which additive noise was used, an implementation in which activation in each input layer was normalized to sum to 1, and an implementation in which separate output groups were used for each ordinal position, with output item at each position represented by the most active unit in the relevant group. However, as noted previously, use of straight Gaussian rank representations with fixed variance, in place of the original log-normal representations, changed the behavior of the model considerably, yielding a pattern of recall accuracy inconsistent with the empirical data (Fig. 3*F*).<sup>e</sup>

### Discussion

We have presented a computational model addressing how serial order is represented in cortical working memory. The model is integrative in two senses. First, the model integrates across three basic findings from single-unit neurophysiology, indicating how they may fit together to subservise a single, critical cognitive function. Second, the model bridges across the domains of neuroscience and behavior, starting from formally specific and highly constraining neuroscientific findings, and leveraging these to explain detailed patterns of recall behavior.

Together, this combination of attributes represents a significant step beyond previous models of serial order processing. A number of psychological models have engaged behavioral data in detail (Page and Norris, 1998; Burgess and Hitch, 1999; Brown et al., 2000; Farrell and Lewandowsky, 2002; Botvinick and Plaut,

2006). In fact, the model we proposed has important features in common with some of these models, most notably the use of overlapping rank representations (Houghton, 1990; Burgess and Hitch, 1999; Brown et al., 2000; Botvinick, 2005; Botvinick and Plaut, 2006). However, in contrast to the present model, most models addressing detailed behavioral benchmarks have not made meaningful contact with neuroscientific data.

Our model also shares basic features with a number of previous models addressing the neural basis of serial order processing, including the use of conjunctive, superpositional sequence representations (Dominey et al., 1995; Dominey, 1997; Beiser and Houk, 1998; O'Reilly and Soto, 2001). The model we presented goes beyond this previous work by making contact with detailed behavioral data.

### Predictions of the model

Like other work proposing the dependence of serial order memory on rank representations in the IPS (Marshuetz, 2005; Nieder, 2005), our model predicts that any disruption of these representations should specifically impair immediate serial recall performance. This appears consistent with neuropsychological evidence associating left parietal damage with impairments in memory span (Vallar and Shallice, 1990). A more distinctive prediction of the model is that there should exist neocortical neurons whose response properties take the form of gain fields combining item and order information in a graded and compressive manner. Although the data suggest that such neurons may occur in the inferior prefrontal cortex (Inoue and Mikami, 2006; see Fig. 1*A*), at least for visual stimuli, gain field representations might well arise first more posteriorly. Indeed, receptive fields resembling those predicted by the model have been observed in the context of motor production, located in the superior parietal lobule (Sawamura et al., 2002).

### Directions for additional evaluation and development

To focus on the issue of representation, our model abstracted over several mechanisms and processes, which could be addressed in a fuller implementation. For example, the internal units in our model were assumed to display persistent activation, a key property of active memory widely believed to underpin working memory function (Fuster, 2001; Miller and Cohen, 2001). One way of elaborating the model would be to incorporate specific mechanisms giving rise to sustained activation, along the lines proposed by Compte et al. (2000) or Zipser et al. (1993). Our implementation also did not address the mechanism by which multiplicative codes might be computed. A more explicit account of this might be drawn work such as that of Mehaffey et al. (2005). Another simplification in our model was to abstract, like some previous neuroscientific models (Beiser and Houk, 1998), over the process of recall. This is another area where the model calls for further development, and where previous models once again provide useful precedents (Dominey, 1997; O'Reilly and Soto, 2001; Botvinick and Plaut, 2006).

There also remain a large number of interesting behavioral phenomena to which the present theory might also be applied. Findings not addressed in the present work include list length effects, suffix and modality effects, grouping effects, and effects of irrelevant speech, as well as effects of prior probability (Botvinick and Bylsma, 2005). Testing the applicability of the model to such additional phenomena presents a worthwhile direction for future work.

<sup>e</sup>The strong recency effect in the figure reflects the fact that early list items are more subject to the cumulative effects of noise. If this factor is equalized across items (as might be justified given that in the laboratory task items are recalled one by one), the straight Gaussian implementation yields a symmetric recall accuracy curve, still inconsistent with the empirical pattern.



## References

- Baddeley A (1986) Working memory. New York: Clarendon.
- Barone P, Joseph JP (1989) Prefrontal cortex and spatial sequencing in macaque monkey. *Exp Brain Res* 78:447–464.
- Beiser DG, Houk JC (1998) Model of cortical-basal ganglionic processing: encoding the serial order of sensory events. *J Neurophysiol* 79:3168–3188.
- Botvinick M (2005) Effects of domain-specific knowledge on memory for serial order. *Cognition* 97:135–151.
- Botvinick M, Bylsma LM (2005) Regularization in short-term memory for serial order. *J Exp Psychol Learn Mem Cogn* 31:351–358.
- Botvinick M, Plaut DC (2006) Short-term memory for serial order: a recurrent neural network model. *Psychol Rev* 113:201–233.
- Brotchie PR, Anderson RA, Snyder LH, Goodman SJ (1995) Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature* 375:232–235.
- Brown G, Preece T, Hulme C (2000) Oscillator-based memory for serial order. *Psychol Rev* 107:127–181.
- Burgess N, Hitch GJ (1999) Memory for serial order: a network model of the phonological loop and its timing. *Psychol Rev* 106:551–581.
- Compte A, Brunel N, Goldman-Rakic PS, Wang XJ (2000) Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cereb Cortex* 10:910–923.
- Conrad R (1965) Order error in immediate recall of sequences. *J Verbal Learn Verbal Behav* 4:161–169.
- Dominey PF (1997) An anatomically structured sensory-motor sequence learning system displays some general linguistic capacities. *Brain Lang* 59:50–75.
- Dominey PF, Arbib MA, Joseph JP (1995) A model of corticostriatal plasticity for learning oculomotor associations and sequences. *J Cogn Neurosci* 7:311–336.
- 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.
- Fortin NJ, Agster KL, Eichenbaum HB (2002) Critical role for the hippocampus in memory for sequences of events. *Nat Neurosci* 5:458–462.
- Funahashi S, Inoue M, Kubota K (1997) Delay-period activity in the primate prefrontal cortex encoding multiple spatial positions and their order of presentation. *Behav Brain Res* 84:203–223.
- Fuster JM (2001) The prefrontal cortex—an update: time is of the essence. *Neuron* 30:319–333.
- Henson RN (1996) Short-term memory for serial order. In: *MRC applied psychology unit*. Cambridge, UK: Cambridge UP.
- Henson RN (1998) Short-term memory for serial order: the start-end model. *Cogn Psychol* 36:73–137.
- Houghton G (1990) The problem of serial order: a neural network model of sequence learning and recall. In: *Current research in natural language generation* (Dale R, Nellish C, Zock M, eds), pp 287–318. San Diego: Academic.
- Inoue M, Mikami A (2006) Prefrontal activity during serial probe reproduction task: encoding, mnemonic and retrieval processes. *J Neurophysiol* 95:1008–1041.
- Jonides J, Lacey SC, Nee DE (2005) Processes of working memory in mind and brain. *Curr Dir Psychol Sci* 14:2–5.
- Kermadi I, Joseph JP (1995) Activity in the caudate nucleus of monkey during spatial sequencing. *J Neurophysiol* 74:911–933.
- Kermadi I, Jurquet Y, Arzi M, Joseph JP (1993) Neural activity in the caudate nucleus of monkeys during spatial sequencing. *Exp Brain Res* 94:352–356.
- Lipton JS, Spelke ES (2003) Origins of number sense: large-number discrimination in human infants. *Psychol Sci* 14:396–401.
- Marshuetz C (2005) Order information in working memory: An integrative review of evidence from brain and behavior. *Psychol Bull* 131:323–339.
- Marshuetz C, Smith EE, Jonides J, Degutis J, Chenevert TL (2000) Order information in working memory: fMRI evidence for parietal and prefrontal mechanisms. *J Cogn Neurosci* 12:130–144.
- Marshuetz C, Reuter-Lorenz PA, Smith EE, Jonides J, Noll DC (2006) Working memory for order and the parietal cortex: an event-related fMRI study. *Neuroscience* 139:311–316.
- Martin N, Gupta P (2004) Exploring the relationship between word processing and verbal short-term memory: evidence from associations and dissociations. *Cogn Neuropsychol* 21:213–228.
- McCormack T, Brown GD, Vousden JI (2000) Children's serial recall errors: implications for theories of short-term memory development. *J Exp Child Psychol* 76:222–252.
- Mehaffey WH, Doiron B, Maler L, Turner RW (2005) Deterministic multiplicative gain control with active dendrites. *J Neurosci* 25:9968–9977.
- Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202.
- Mushiaki H, Saito N, Sakamoto K, Itoyama Y, Tanji J (2006) Activity in lateral prefrontal cortex reflects multiple steps of future events in action plans. *Neuron* 50:631–641.
- Nieder A (2005) Counting on neurons: the neurobiology of numerical competence. *Nat Rev Neurosci* 6:177–189.
- Nieder A, Miller EK (2003) Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron* 37:149–157.
- Nieder A, Miller EK (2004) A parieto-frontal network for visual numerical information in the monkey. *Proc Natl Acad Sci USA* 101:7457–7462.
- Nieder A, Freedman DJ, Miller EK (2002) Representation of the quantity of visual items in the primate prefrontal cortex. *Science* 297:1708–1711.
- Nieder A, Diester I, Tudusciuc O (2006) Temporal and spatial enumeration processes in the primate parietal cortex. *Science* 313:1431–1435.
- Ninokura Y, Mushiaki H, Tanji J (2003) Representation of the temporal order of visual objects in the primate lateral prefrontal cortex. *J Neurophysiol* 89:2868–2873.
- Ninokura Y, Mushiaki H, Tanji J (2004) Integration of temporal order and object information in the monkey lateral prefrontal cortex. *J Neurophysiol* 91:555–560.
- O'Reilly RC, Busby RS (2002) Generalizable relational binding from coarse-coded distributed representations. In: *Advances in neural information processing systems (NIPS)* (Dietterich TG, Becker S, Ghahramani Z, eds). Cambridge, MA: MIT.
- O'Reilly RC, Soto R (2001) A model of the phonological loop: generalization and binding. In: *Advances in neural information processing systems* (Dietterich TG, Ghahramani Z, eds), pp 83–90. Cambridge, MA: MIT.
- O'Reilly RC, Busby RS, Soto R (2003) Three forms of binding and their neural substrates: alternatives to temporal synchrony. In: *The unity of consciousness: binding, integration and dissociation* (Cleeremans A, ed), pp 168–192. Oxford: Oxford UP.
- Page M, Norris D (1998) The primacy model: a new model of immediate serial recall. *Psychol Rev* 105:761–781.
- Pouget A, Sejnowski TJ (1997) Spatial transformations in the parietal cortex using basis functions. *J Cogn Neurosci* 9:222–237.
- Pouget A, Snyder AZ (2000) Computational approaches to sensorimotor transformations. *Nat Neurosci* 3:1192–1198.
- Salinas E (2004) Fast remapping of sensory stimuli onto motor actions on the basis of contextual modulation. *J Neurosci* 24:1113–1118.
- Salinas E, Abbott LF (2001) Coordinate transformations in the visual system: how to generate gain fields and what to compute with them. *Prog Brain Res* 130:175–190.
- Salinas E, Thier P (2000) Gain modulation: a major computational principle of the central nervous system. *Neuron* 27:15–21.
- Sawamura H, Shima K, Tanji J (2002) Numerical representation for action in the parietal cortex of the monkey. *Nature* 415:918–922.
- Vallar G, Shallice T (1990) Neuropsychological impairments of short-term memory. Cambridge, UK: Cambridge UP.
- Zipser D, Kehoe B, Littlewort G, Fuster J (1993) A spiking network model of short-term active memory. *J Neurosci* 13:3406–3420.