Modelling spatial immediate serial recall performance in a macaque (*Macaca mulatta*)

Mark Hurlstone
University of Western Australia

Mark J. Hurlstone
School of Psychology
35 Stirling Highway
University of Western Australia
Crawley WA 6009
email: mark.hurlstone@uwa.edu.au
Tel: +61 8 6488 3249
URL: http://mark-hurlstone.github.io

Conflict of interest: The author declares that he has no conflict of interest.

Animal Cognition manuscript No.
(will be inserted by the editor)

Modelling spatial immediate serial recall performance in a macaque (*Macaca mulatta*)

Draft of: 06/07/17

Abstract Botvinick et al (2009) examined the immediate serial recall (ISR) capabilities of a rhesus macaque using a spatial short-term memory task akin to that employed in human memory studies. They found that the monkey's performance exhibited several hallmark effects of human ISR, suggesting that computational mechanisms of human short-term serial order memory may therefore be appropriate for understanding ISR in some nonhuman primate species. Here I applied a computational model of short-term memory that instantiates four core mechanisms of human ISR—competitive queuing, position marking, a primacy gradient, and response suppression—to their monkey's data. The model provided an excellent account of the monkey's ISR performance and its capacity to do so was greatly diminished when one or more of the mechanisms was eliminated from the model, suggesting that all four mechanisms—rather than a restricted set of those mechanisms—is required to explain the data. The results of the current simulation study suggest that common mechanisms may underpin short-term serial order memory across species.

Keywords Serial order · Macaque · Short-term memory

The ability to encode, store, and recall arbitrary fixed sequences of items, actions, and events is a fundamental human cognitive skill. This kind of sequence learning has been studied extensively in humans using the *immediate serial recall* (ISR) task in which participants are given novel sequences of verbal, visual, or spatial items that they must subsequently recall in the correct order. Considerable progress has been made in understanding the cognitive components underpinning this seemingly simple task thanks to the development of several computational models of short-term memory that explain detailed aspects of the human ISR data using explicit mechanisms for the representation of serial order (Botvinick and Plaut 2006; Brown et al 2000; Burgess and Hitch 1999; Farrell and Lewandowsky 2002; Hartley et al 2016; Henson 1998; Lewandowsky and Farrell 2008; Page and Norris 1998). The analysis of recall error data have played an instrumental role in constraining these models. For example, error analyses have ruled out chaining models in which order is stored through item–item associations (Farrell et al 2013; Henson et al 1996), and conferred support instead for positional models in which order is stored through position–item associations (Conrad 1960; Henson 1999; Ryan 1969a,b).

The ability to learn arbitrary fixed sequences is not unique to humans. Indeed, non-human primates (hereafter, 'primates') posses a remarkable capacity for mastering serial tasks. Monkeys can encode and reproduce the order of arbitrary fixed sequences of visual images (Chen et al 1997; Orlov et al 2000; Terrace 2005) numerals (Inoue and Matsuzawa 2007; Matsuzawa 1985; Kawai and Matsuzawa 2000), and spatial targets (Barone and Joseph 1989; Kermadi and Joseph 1995). They can remember sequences containing as many as seven elements (Chen et al 1997) and their performance on serial tasks has sometimes been shown

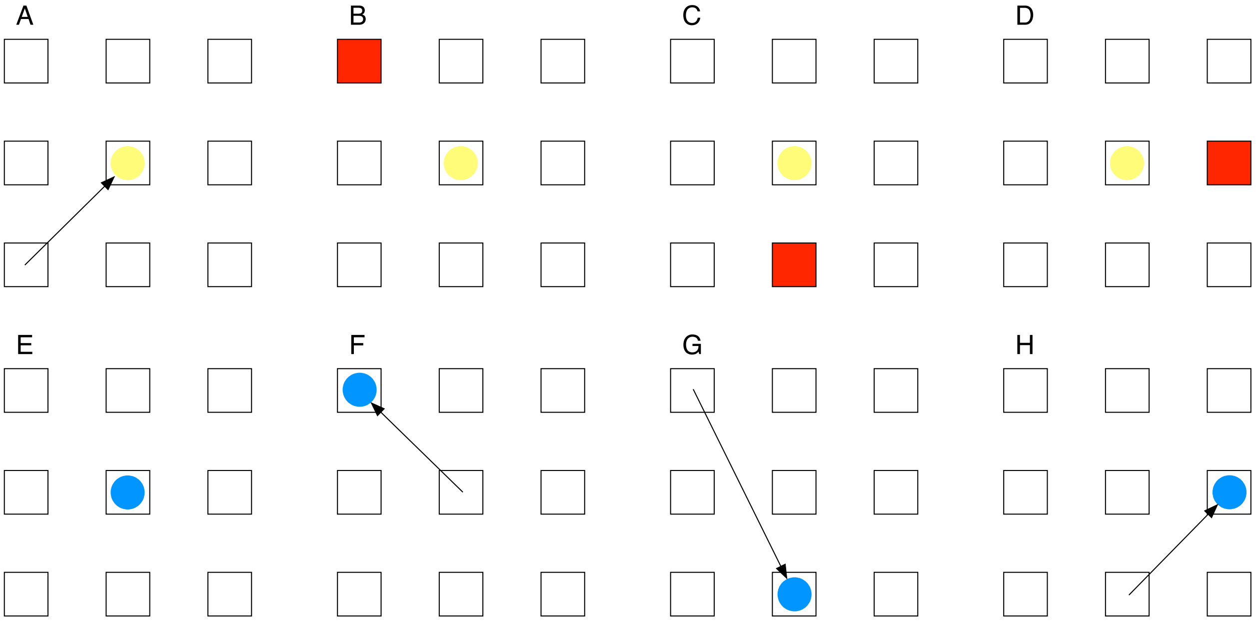

Fig. 1 Schematic of the spatial ISR task used by Botvinick et al (2009). The items were locations contained within a 3 × 3 visually presented grid. (**A**) At the start of a trial, the central location was illuminated and the animal was required to move a blue cursor from another location into the central location using a joystick, which caused the cursor to turn yellow. (**B-D**) Following the temporary disappearance of the cursor, a sequence of three or four locations—depending on the list-length; three in this example—was conveyed to the observing animal by highlighting an arbitrary sub-set of locations red, one location at a time. (**E**) Following a brief delay, the blue cursor re-appeared in the central location cueing the animal to reproduce the just observed sequence. (**F**) The animal used the joystick to move the blue cursor to the location corresponding to the first serial position—holding it in place briefly until the response was registered by the computer—before repeating this process for the locations corresponding to the second (**G**) and third (**H**) serial positions (and the location corresponding to the fourth serial position when the sequence length was four-items). Figure adapted from Botvinick et al (2009).

to exceed that of human participants (Inoue and Matsuzawa 2007). Like humans, monkeys are able to collectively plan short sequences of items prior to enactment (Biro and Matsuzawa 1999; Inoue and Matsuzawa 2007; Kawai and Matsuzawa 2000; Scarf et al 2011), and they appear to represent sequences by forming associations between items and positions, rather than between successive items (Carpenter et al 1999; D'Amato and Colombo 1988; D'amato and Colombo 1989; Orlov et al 2000, 2002, 2006).

Notwithstanding these similarities, as noted by Botvinick et al (2009), until recently comparisons of the sequence processing capabilities of humans and primates had been hampered by the fact that studies with the two groups have employed different tasks. The primate studies have predominantly employed a serial learning paradigm (Terrace 2005) in which monkeys gradually acquire sequences over multiple encoding and recall attempts using a forward training procedure—e.g., to acquire the sequence A → B → C → D → E → F → G, the subject is trained to respond A initially, then A → B, then A → B → C, and so on and so forth until the entire sequence can be produced. As such, these studies probe long-term memory for serial order, rather than short-term memory for serial order—the competency indexed by the ISR task used with humans.

Spatial ISR in primates

This gap in the literature was recently filled by two studies that examined the sequence processing abilities of a single rhesus macaque (Botvinick et al 2009) and two baboons (Fagot and De Lillo 2011) using an ISR task for sequences of spatial locations similar to that used to investigate serial spatial short-term memory in humans (Jones et al 1995; Smyth and Scholey 1996). Although both studies showed that the ISR performance of monkeys exhibits attributes of human ISR—notably list length and serial position effects (see below)—the study of Botvinick et al (2009) is especially noteworthy, since these authors evaluated their monkey's performance with reference to several benchmark characteristics of human ISR, which included a detailed analysis of recall error data.

The spatial ISR task employed by Botvinick et al (2009) is illustrated graphically in Fig 1. The to-be-remembered items were spatial locations (square icons) contained within a 3 × 3 visually presented grid. On each ISR trial, a sequence of either three or four locations was conveyed by sequentially highlighting a random sub-set of the locations, after which the animal's task was to reproduce the sequence using a joystick controlled cursor. The monkey's perfor-

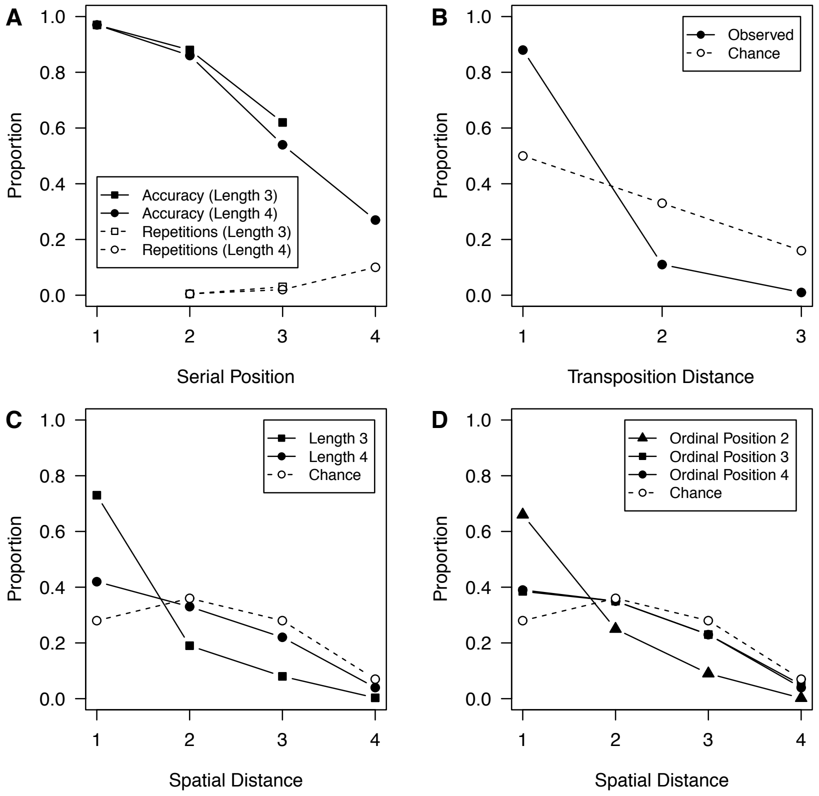

Fig. 2 The monkey's results for four key behavioural measures: (**A**) accuracy and repetition error serial position curves, (**B**) transposition error gradient for four-item lists, (**C**) aggregate spatial error gradients, and (**D**) spatial error gradients for the second, third, and fourth output position in four-item lists.

mance on this task exhibited several hallmarks of human ISR, which are described in the foregoing sub-sections.

List length effect

In humans, ISR performance for lists composed of verbal items (Crannell and Parrish 1957; Maybery et al 2002) and spatial items (Jones et al 1995; Smyth and Scholey 1996) decreases as the length of the target sequence to-be-recalled increases—a result dubbed the *list length effect*.

Fig 2A shows the accuracy serial position curves for three- and four-item lists for the monkey's performance. Consistent with the list length effect in human ISR, it can be seen that recall accuracy was lower for four- than three-item lists.

Primacy and recency effects

A hallmark feature of human ISR is the presence of serial position effects on performance. Recall accuracy decreases sharply from the first position onwards (viz. the *primacy effect*), with an upturn in performance for the final item (viz. the *recency effect*), yielding an asymmetrically bowed serial position curve. Primacy and recency effects are a characteristic of both verbal ISR (Baddeley 1968; Henson et al 1996) and spatial ISR (Jones et al 1995; Smyth and Scholey 1996) performance in humans.

Consistent with the human ISR data, it can be seen from Fig 2A that the monkey exhibited a pronounced primacy effect. However, at variance with those data it did not exhibit a positive recency effect. Indeed, the monkey actually showed a negative recency effect, with the accuracy of recall of the final item on three- and four-item lists being lower than for the penultimate item.

Transposition error gradients

A large proportion of errors in human ISR are order errors involving the recall of items in the wrong serial positions. These transposition errors exhibit a tendency to cluster around their correct serial positions—an empirical regularity known as the *locality constraint* (Henson 1996). Accordingly, when the probability of a transposition is plotted as a function of transposition distance—viz. the ordinal distance of the transposed item from its correct position—

the resulting gradients peak for one-apart transpositions with the number of transpositions decreasing monotonically with increasing transposition distance. The locality constraint is a robust feature of both verbal ISR (Henson 1996; Lewandowsky and Farrell 2008) and spatial ISR (Hurlstone and Hitch 2015; Parmentier et al 2006; Smyth and Scholey 1996) performance in humans.

The transposition gradient for four-item lists associated with the monkey's performance can be inspected in Fig 2B from which it can be seen that transpositions obeyed the locality constraint observed in humans.

Item similarity effect

Errors in human ISR sometimes involve confusions between items that share similar characteristics. In verbal ISR, such confusions are based on phonology as evidenced by the phonological similarity effect (Baddeley 1966, 1968; Conrad and Hull 1964)—lists of phonologically similar sounding items (e.g., B D G P T V) are recalled less accurately than lists of phonologically dissimilar sounding items (e.g., e.g., F K L R X Y). In spatial short-term memory, item confusion errors occur on the basis of the spatial proximity of items (Hitch, 1974). Accordingly, when errors are plotted as a function of spatial proximity to the correct item, the probability of an error decreases monotonically with increasing spatial distance—viz. a locality constraint over the spatial, as opposed to temporal, distance between items (Hitch 1974; Rerko et al 2014). In human spatial ISR, the magnitude of such spatial errors is known to interact with serial position (Farrand and Jones 1996; Farrand et al 2001).

Fig 2C plots the aggregate spatial error gradients for three- and four-item lists for the monkey's performance (in the figure, spatial proximity is represented using a Manhattan distance metric). Consistent with the locality constraint, the gradients peak for one-apart spatial errors and decrease monotonically with increasing spatial distance. The gradient is also steeper for three- than four-item lists. Fig 2D plots the individual spatial error gradients for the second, third, and fourth output positions in four-item lists. It can be seen that the distribution of spatial errors interacts with serial position, with the error gradients becoming shallower at later output positions.

In brief, consistent with the data from human spatial ISR the monkey's performance was sensitive to the spatial similarity of items, and additionally this spatial similarity effect interacted with serial position.

Repetition errors

Repetition errors—viz. the repeated report of an item presented only once on the study list—are very rare in human verbal (Henson 1996; Vousden and Brown 1998) and spatial ISR (Hurlstone and Hitch 2015), accounting for around 1% (Hurlstone and Hitch 2015) to 5% (Henson 1996) of all responses. The frequency of repetitions in human ISR is known to increase with serial position (Henson 1996).

The repetition error serial position curves for three- and four-item lists for the monkey's performance are shown alongside the accuracy serial curves in Fig 2A. Consistent with the human ISR data, erroneous repetitions were rare and their frequency increased across output positions.

Fill-in errors

Transposition errors in human ISR exhibit a particular pattern of sequential dependency. If an item i is recalled a position ahead of its correct position (e.g., recalling B*xx* following the list ABC), item $i-1$ is more likely to be recalled at the next output position (e.g., BA*x*; known as a *fill-in error*) than item $i+1$ (e.g., BC*x*; known as an *infill error*). Specifically, fill-in errors outweigh infill errors by a ratio of approximately 2:1 in both verbal ISR (Farrell et al 2013; Henson 1996; Page and Norris 1998; Surprenant et al 2005) and spatial ISR (Guérard and Tremblay 2008).

This fill-in tendency is also a feature of the monkey's ISR performance, which contained 46 fill-in errors and 10 infill errors, yielding a ratio of 4.6:1.

Protrusion errors

In human ISR, items that were not part of the study list sometimes intrude into participant's recalls. These errors often take the form of immediate intrusions, which occur when a participant recalls an item from trial $N-1$ on trial N despite the item not being present on the study list for that trial. When these immediate intrusion errors occur they often maintain their serial position from trial $N-1$ (Conrad 1960; Henson 1996, 1999). These position-preserving intrusions are known as *protrusion* errors (Henson 1996) and their frequency of occurrence is greater than would be expected by chance alone.

The monkey's ISR performance was also characterised by protrusions. Out of 157 immediate intrusion errors on four-item lists, 51 (32.5%) were protrusion errors that maintained their position from the preceding trial—a rate greater than would be expected by chance (25%; viz. $1/n$—where n is the list length).

Current study

As noted by Botvinick et al (2009), the correspondence between the monkey's ISR performance and that of humans is striking and suggests that the mechanisms of serial order embodied in computational models of serial order in human short-term memory might be extensible to some primate species. The aim of the current study is to provide a formal test of this possibility by applying a computational model that implements the representational mechanisms that have been found to underpin human ISR to their monkey's data to establish if these mechanisms—or a restricted set of these mechanisms—can reproduce his behavioural results.

The model is based on competitive queuing models of serial behaviour (Glasspool 2005; Houghton 1990)—the dominant class of models of human ISR (Brown et al 2000; Burgess and Hitch 1999; Farrell and Lewandowsky 2002; Henson 1998; Page and Norris 1998). In competitive queuing models, items are activated in parallel and the item with the strongest activation level is selected for output. There is already direct evidence that primates utilise the competitive queuing mechanism based on electrophysiological recording data obtained whilst rhesus monkeys performed a serial imitation task (Averbeck et al 2002, 2003a,b). In the competitive queuing model examined here, serial order is represented by forming associations between items and a representation of the their ordinal sequence position (viz. position marking), by encoding each item with progressively less strength (viz. a primacy gradient), and by suppressing items in memory once they have been recalled (viz. response suppression). There is considerable evidence that these four mechanisms are implicated in the representation of serial order in short-term memory in humans (Hurlstone et al 2014; Lewandowsky and Farrell 2008).

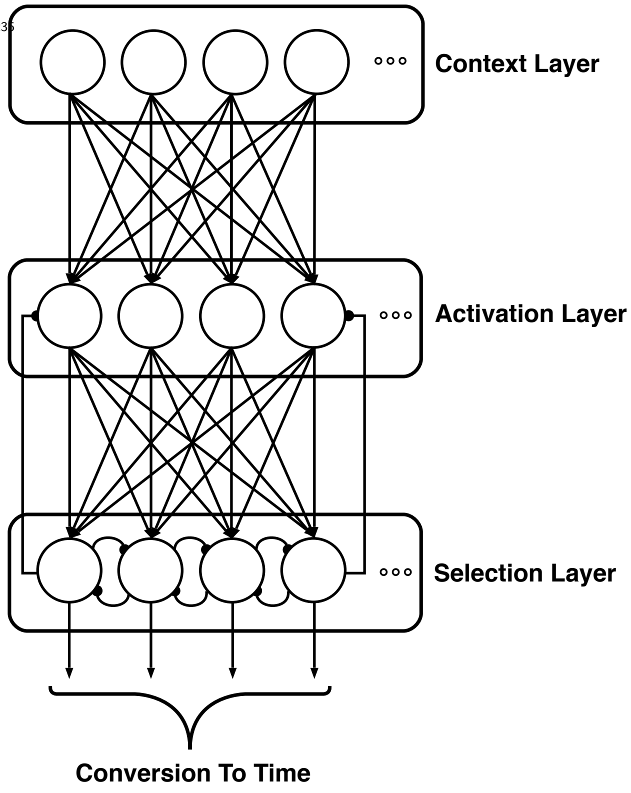

Fig. 3 Schematic of the architecture of the competitive queuing model. Note that to avoid visual clutter only a subset of the units in each layer are shown; inhibitory connections from the selection layer to the activation layer are included only for two units; self-excitatory connections in the selection layer are omitted; and only adjacent-neighbour lateral inhibitory connections are illustrated in the selection layer. See main text for further details.

A generic competitive queuing model of spatial ISR

Model architecture

Fig 3 shows a schematic of the architecture of the competitive queuing model, which comprises three distinct layers of units—a context layer, an activation layer, and a selection layer. The context layer maintains a distributed representation of the current position in the list, whereas the activation and selection layers implement a localist coding scheme, whereby each unit corresponds to a different item (location) in the 3 × 3 visually presented grid (Fig 1). The context layer and activation layer are connected by a Hebbian weight matrix within which short-term associations between items and a representation of their position are stored. Each unit in the activation layer has a connection to all other units in the selection layer. The strength of these connections reflect the similarity between each location with itself, and all other locations in the visual grid. Accordingly, the connection is strongest between each unit in the activation layer and its corresponding unit in the selection layer, with the strength of connections between all other units decreasing as an exponential function of their spatial distance from one another. Formally, the strength of the connection w_{ij}^{AS} between each unit j in the activation layer and each unit i in the selection layer is given by:

$$w_{ij}^{AS} = e^{-cd_{ij}}, \tag{1}$$

where d_{ij} is the Manhattan distance between item i and item j, which is given by:

$$d_{ij} = \sum_{l=1}^{k} |a_l - b_l|, \tag{2}$$

where the two-element vectors a and b represent item i's and j's coordinates within the spatial array (Fig. 1), respectively, and l indexes the k dimensions of each vector. The parameter c in equation 1 is a sensitivity parameter that governs the rate at which similarity between items decreases with increasing spatial distance.

In addition to the weighted connections from the activation layer, each unit in the selection layer has a self-excitatory connection and lateral inhibitory connections to all other units in the selection layer. Each unit in the selection layer also has an inhibitory connection to its corresponding unit in the activation layer.

Encoding

The position of each item in a serial list is represented by a 16-element distributed context vector (with values between -1 and +1) constructed so that the cosine similarity between any pair of vectors decreases as an exponential function of their absolute ordinal distance (Farrell 2006):

$$\cos\left(\mathbf{c}_p,\mathbf{c}_q\right) = \phi^{(|p-q|)}, \tag{3}$$

where $\mathbf{c}_p$ and $\mathbf{c}_q$ are context vectors for positions p and q, respectively, and ϕ is a parameter controlling the rate at which similarity drops off with increasing ordinal distance.

The encoding of a list of items within the network is carried out by imposing the context vector representing the current list position over the context layer, activating the just presented item in the activation layer—viz. setting its unit's activation to 1, and the activation of all other units to 0—and modifying the strength of the weight w_{ij}^{CA} connecting each unit i in the activation layer and each unit j in the context layer via Hebbian learning:

$$\Delta w_{ij}^{CA} = \eta_p\, a_i^A\, a_j^C, \tag{4}$$

where a_i^A is the activation of unit i in the activation layer, a_j^C is the activation of unit j in the context layer, and η is a learning rate parameter that governs the strength of the context-item associations. Specifically, the strength of these associations decreases exponentially across list positions according to a primacy gradient of encoding strength:

$$\eta_p = \theta^{p-1}, \tag{5}$$

where θ is a parameter determining the steepness of the primacy gradient over list positions.

The process just described encodes serial order through position–item associations (viz. position marking). However the primacy gradient in encoding strength means that the representation of order contains an ordinal, as well as a positional, component. Evidence that rhesus monkeys use position–items associations has been obtained from positional intrusion errors in a delayed sequence-recall task (Orlov et al 2000, 2002, 2006).

Retrieval

At retrieval the context vector for the current recall position is reinstated in the context layer and presented to the weight matrix connecting it with the activation layer. The net input to each item unit i in the activation layer net_i^A is given by:

$$net_i^A = \textstyle\sum_j a_j^C\, w_{ij}^{CA} + \varepsilon(0,\delta^A), \tag{6}$$

where a_j^C and w_{ij}^{CA} are the same as before. This generates an activation gradient over the item units such that the unit corresponding to the item presented at the position being cued will be activated strongest, with the activation of units corresponding to neighbouring list-items falling off as their ordinal distance from the cued item on the presented list increases. The activations elicited by the reinstated context vector are augmented with zero-mean Gaussian noise ε with standard deviation δ^A to simulate transposition errors.

A key feature of the competitive queuing approach is that items are temporarily inhibited once they have been output by setting their activations to negative values. Accordingly, some mechanism is required to ensure that suppressed items are briefly removed from the response competition. To accomplish this, the activation of each item unit i in the activation layer a_i^A is subject to the following activation function:

$$a_i^A(p) = \begin{cases} net_i^A(p) & \text{if } a_i^A\ (p\text{-}1) \geq 0, \\ net_i^A(p) + \mathrm{a}_i^A(p\text{-}1)\exp(\text{-}\lambda) & \text{otherwise,} \end{cases} \tag{7}$$

To explain, if the activation of unit a_i^A at position p-1 is equal to or greater than zero—viz. if that unit is not in a suppressed state—then its activation at position p is based on the net input elicited by the reinstated context vector only, otherwise its activation is the joint combination of the net input elicited by the reinstated context vector and its current negative activation value. The activation values of suppressed items gradually recover from inhibition over recall positions, with the extent of this release from inhibition being governed by the parameter λ.

The most active item unit in the activation layer propagates its activation along its weighted connections to each unit in the selection layer. The initial input to each unit i in this layer a_i^S is given by:

$$a_i^S = a_{win}^A\, w_{iwin}^{AS} + \varepsilon(0,\delta^S), \tag{8}$$

where a_{win}^A is the activation of the winning item unit win in the activation layer and w_{iwin}^{AS} is the weight of its connection with each unit i in the selection layer, which is governed by equations 1 and 2. This results in the item unit in the selection layer corresponding to the winning item in the activation layer being activated the strongest, with the activation of other item units decreasing as their spatial distance from this item increases. The item unit activations in the selection layer are supplemented with zero-mean Gaussian noise ε with standard deviation δ^S in order to simulate spatial errors. The activations of item units in the selection layer are iteratively updated over time according to the following equation:

$$int_i^S(t) = a_i^S(t-1)\,\alpha + \beta \sum_{j \neq i} a_j^S(t-1) + \varepsilon(0, \delta^S), \qquad (9)$$

where $int_i^S(t)$ is the internal activation—viz. the net input a unit receives from within the selection layer—of item unit i at time t, $a_i^S(t-1)$ is its activation at the preceding time step, $a_j^S(t-1)$ is the activation of each other item unit j at the preceding time step, α is the strength of the recurrent self-excitatory connections ($\alpha = 1.1$), β is the strength of the lateral-inhibitory connections (β = -0.1), and ε and δ^S are as before. Equation 9 implements a winner-takes-all "competitive filter" (Houghton 1990), which selects the most active item for output. The initially most active item unit has the advantage that it will send more activation to itself than any other item unit, and will also receive the least lateral inhibition. As the unit activations are iteratively updated over time this results in a gradual increase in the activation of the strongest unit, and a gradual decrease in the activations of the weaker units as they receive more lateral inhibition.

The competitive filter is applied for a fixed duration of 20 iterative cycles—sufficient time to identify an unambiguous response. The item with the strongest activation level at the end of this period constitutes the network's response for the current recall position. The recall of an item is followed by the suppression of its unit's activation in the activation layer by setting its activation level to a fixed suppressed value τ of -1. This implements the mechanism of response suppression, which is a core ingredient in competitive queuing models. It was also assumed that the process of generating an item for output adds noise to the representations of yet to-be-recalled items—an assumption known as output interference. Consistent with other network models of serial recall (Brown et al 2000; Lewandowsky and Farrell 2008), this output-contingent interference was modelled by adding zero-mean Gaussian noise with standard deviation δ^{wCA} to each of the context-item weights linking the activation and context layers following the output of each item (δ^{wCA} = .04).

To model protrusion errors, the context-item weights were not refreshed (i.e., set to zero) at the start of each new simulation trial. Instead, the context-item associations established on each trial were stored on a common weight matrix. At the end of each simulation trial, the context-item weights were normalised so that recent weight changes were more influential than earlier weight changes.

Modelling negative recency

One challenge in modelling the monkey's data is capturing the negative recency effect. This effect is difficult to reproduce using the model just described because competitive queuing models are by design configured to generate a recency (along with a primacy) effect. There are several factors that promote recency in the current model. One factor is "edge effects"—because the last item only has neighbouring competitors on one side of the list it will be recalled with a higher level of accuracy than items at medial serial positions, which have neighbouring competitors on both sides of the list. Another contributing factor is response suppression—as recall progresses, the cohort of recall candidates gradually winnows down due to the suppression of items already emitted, which increases the likelihood that the final item will be assigned to its correct position.

Modelling the negative recency effect is necessary not just in terms of providing an explanation for its presence in the monkey's data but because it is impossible to recreate the distributions of the various different errors without modelling the sharp decline in recall accuracy toward the end of the list. How then to explain this result? One explanation offered by Botvinick et al (2009) is that it might be due to interference caused by irrelevant events—including a tone and a juice reward—that were interspersed between each recall episode. This interference is not dissimilar to the output interference implemented in the current model and provides an additional precedent for its incorporation. However, output interference alone was found to be insufficient to generate the sharp drop in recall over the final two serial positions. Another possibility—and the explanation pursued here—is that this effect reflects the action of a *selective encoding strategy* adopted during the encoding of serial lists. This account supposes that the monkey allocated most of his attention to encoding the first and second item in each list, with a subsequent sharp drop in the attention allocated to encoding the third item (in three- and four-item lists), and a further abrupt drop in the attention allocated to encoding the fourth item (in four-item lists).

This selective encoding strategy was implemented through the following modification to equation 2 that generates the primacy gradient governing the encoding strength of the context-item associations across serial positions:

$$\eta_p = \begin{cases} \theta^{p-1} & \text{if } p \leq 2, \\ \theta^{p+1} & \text{if } p = 3, \\ \theta^{p+3} & \text{if } p = 4, \end{cases} \qquad (10)$$

The encoding strength of the first two items in the list is calculated in the same way as in equation 2. However, the encoding strength of the third item (on three- and four-item lists) is calculated as though this were the fifth item on a longer sequence, whilst the encoding strength of the fourth item (on four-item lists) is calculated as though this were the eighth item on a longer sequence. As desired, this produces a sharp drop in encoding strength from input position 2 to 3, and another sharp drop in encoding strength from input position 3 to 4.

Model fitting procedure

The to-be-fitted data consisted of the following behavioural measures (27 data points in total): (1) accuracy serial position curves (seven data points), (2) repetition serial position curves (seven data points), (3) transposition gradient (three data points), (4) aggregate spatial error gradients (eight data points), (5) proportion of fill-in to infill errors (one data point), and (6) proportion of immediate intrusion errors that were protrusions (one data point). Data for both list-lengths were used for behavioural measures 1, 2, and 4, whereas only the data for four-item lists were used for the other behavioural measures (consistent with the reporting of these data by Botvinick et al 2009).

To fit the model, predictions were generated for the above measures for a set of starting model parameter values and the discrepancy between the data and model predictions was evaluated using the Pearson chi-square statistic:

$$\chi^2 = N \sum_i \frac{p_i - \pi_i}{\pi_i}, \tag{11}$$

where p_i is the observed proportion for data point i, π_i is the corresponding proportion predicted by the model, and N is the number of observations. Since the chi-square statistic was calculated using observed and predicted proportions, rather than frequencies, the value of N was simply set to 100. The parameters of the model were varied systematically using the SIMPLEX function minimisation algorithm (Nelder and Mead 1965) until a minimum value of the chi-square statistic was obtained. Each parameter vector explored by the minimisation algorithm involved 2500 simulation trials of three- and four-item lists.

The parameters that were varied during the fitting were the degree of similarity of the context vectors ϕ; the steepness of the primacy gradient in encoding strength θ; the amount of decay from response suppression λ; the sensitivity parameter controlling the degree of confusability of items in the selection layer c; the amount of noise in the activation layer δ^A; and the amount of noise in the selection layer δ^S (six free parameters in total).

Model comparisons

The general model described above instantiates several different mechanisms. To ensure that the incorporation of each of these mechanisms is warranted and that the model is not over-specified, model comparisons were performed in which the fit of the general model was compared with that of several restricted model variants in which one mechanism (or several) in the general model were eliminated. Specifically, the fit of the general model was compared with that of three restricted model variants: (a) a model in which response suppression was eliminated (removing two parameters; viz. τ, and λ), (b) a model in which the primacy gradient was eliminated (removing one parameter; viz. θ), and (c) a model in which the primacy gradient and response suppression was eliminated (removing three parameters; viz. θ, τ, and λ). These models were fit to the behavioural measures in exactly the same way as described for the general model, except that the parameters associated with the eliminated mechanisms were rendered inactive.

In addition to evaluating the models in terms of their best fitting chi-square statistics, the Bayesian information criterion (BIC, Schwarz, 1978) was calculated. The BIC is an index that takes into consideration both a model's goodness of fit and its number of model parameters. The BIC was calculated as:

$$\mathrm{BIC}_i = V_i \ln(n) + n \ln\left(\frac{\mathrm{RSS}_i}{n}\right), \tag{12}$$

where V is the number of free model parameters, n is the number of data points being fitted, RSS is the residual sum of squares, and i indexes the model for which BIC is being calculated.

To aid interpretation, the raw BIC scores were converted into BIC weights (Burnham and Anderson 2002; Wagenmakers and Farrell 2004), which express the degree of support for each model on a continuous measure of evidence. The BIC weight for model i was calculated by:

$$w\mathrm{BIC}_i = \frac{\exp(-0.5\,\Delta\mathrm{BIC}_i)}{\sum_{k=1}^{K} \exp(-0.5\,\Delta\mathrm{BIC}_k)}, \tag{13}$$

where $\Delta\mathrm{BIC}_i$ is the difference in BIC between model i relative to the best model, and each $\Delta\mathrm{BIC}_k$ is the difference in BIC between a specific model k in the candidate set K and the best model.

Simulation Results

The goodness-of-fit quantities for the general model and the three restricted model variants can be scrutinised in Table 1. It can be seen by inspection that the general model obtained the smallest χ^2 value, followed by the restrict^{PG+PM} model, then the restrict^{PM} model, with the restrict^{PM+RS} model obtaining the largest χ^2 value.

Table 1 Goodness-of-fit quantities for the general model and the three restricted model variants.

Model	V	χ^2	BIC	ΔBIC	wBIC
general	**7**	**23.05**	**-168.79**	**0**	**1**
restrict^{PG+PM}	5	104.91	-102.22	66.57	0
restrict^{PM+RS}	6	120.69	-93.93	74.86	0
restrict^{PM}	4	114.51	-99.28	69.51	0

Note: V = number of free model parameters; χ^2 = Pearson chi-square statistic; BIC = Bayesian information criterion; ΔBIC = difference in BIC with respect to the best fitting model; wBIC = BIC weight. The bold item indicates the best fitting model.

Before scrutinising the BIC scores, note that an extra parameter was added to the value of V (the number of free model parameters) prior to the computation of the BIC for the general model and the restrict^{PM+RS} model—taking the value of V for these models from six to seven and five to six, respectively. This extra parameter reflects the degree of response suppression (τ). Although this parameter was not varied in the fitting (it is by definition a fixed parameter), the models nevertheless need to be penalised for containing an extra parameter that is not present in the other two models.

Turning now to the BIC scores, the scores shown in Table 1 confirm that the advantage of the general model over the restricted model variants stands after controlling for model complexity, and additionally the BIC weights indicate that the evidence in favour this model is decisive. The best fitting parameters for the general model are given in Table 2. As the model comparisons confirm that the general model is the preferred model of the data, for brevity, in what follows I will only present the simulation results for this model.

List length, primacy, and recency

The accuracy serial position curves predicted by the general model are shown in Fig 4A. It is apparent from inspection of this figure that the general model predicted a small list length effect, an extensive primacy effect, and a negative recency effect consistent with the empirical data (Fig 2A). The list-length effect arises in the model because the more items in the list, the greater the probability there will be at least one error. The primacy effect is attributable to the primacy gradient in the encoding strength of items, whilst the negative recency effect is the result of the selective encoding strategy built into the primacy gradient, which causes an abrupt drop in encoding strength toward the end of the list thereby counteracting the mechanisms that promote recency.

Transposition error gradient

Consistent with the empirical transposition gradient for four-item lists (Fig 2B), the transposition gradient predicted by the general model for the same list length shows a positional locality constraint—the gradient peaks for one-apart transpositions and then decreases monotonically with increasing transposition distance (Fig 4B). Thus, when an item was recalled in the wrong position, it tended to be recalled in a position close to its correct position. In the model, the positional locality constraint arises due to the local self-similarity of the context signal—neighbouring states (viz. adjacent serial positions) of context are more similar to one another than states that are separated in ordinal distance (viz. non-adjacent serial positions). This means that when a context pattern is presented to the activation layer, items that are close in ordinal distance to the target item will be activated more strongly than items that are far from the target item in ordinal distance. Accordingly, when the item unit activations are augmented with noise, near-neighbour transpositions will be more likely than distant-neighbour transpositions.

Spatial error gradients

As well as predicting a locality constraint for transposition errors, the general model predicted a locality constraint for spatial errors. In accordance with the monkey's data (Fig 2C), the aggregate spatial error gradients predicted by the model peak for one-apart spatial errors, with the frequency of errors decreasing monotonically with increasing spatial distance, and additionally, the aggregate spatial error gradient is steeper for three- than for four-item lists (Fig 4C). Furthermore, it can be seen from inspection of Fig 4D that the general model captured the reduction in slope of the individual spatial error gradients with increasing output position for four-item lists that is a feature of the monkey's ISR performance (Fig 2D). In the model, the spatial locality con-

Table 2 Fixed and free parameter values for the general model.

Parameter	Description	Value
Fixed parameters		
α	Recurrent self-excitation	1.1
β	Lateral inhibition	–0.1
τ	Level of response suppression	–1
δ^{wCA}	Degree of output interference	0.04
Free parameters		
ϕ	Similarity of context vectors	0.6755
θ	Steepness of primacy gradient	0.7827
λ	Decay from response suppression	0.1631
c	Item confusability in selection layer	0.0713
δ^A	Noise applied to activation layer	0.0480
δ^S	Noise applied to selection layer	0.0055

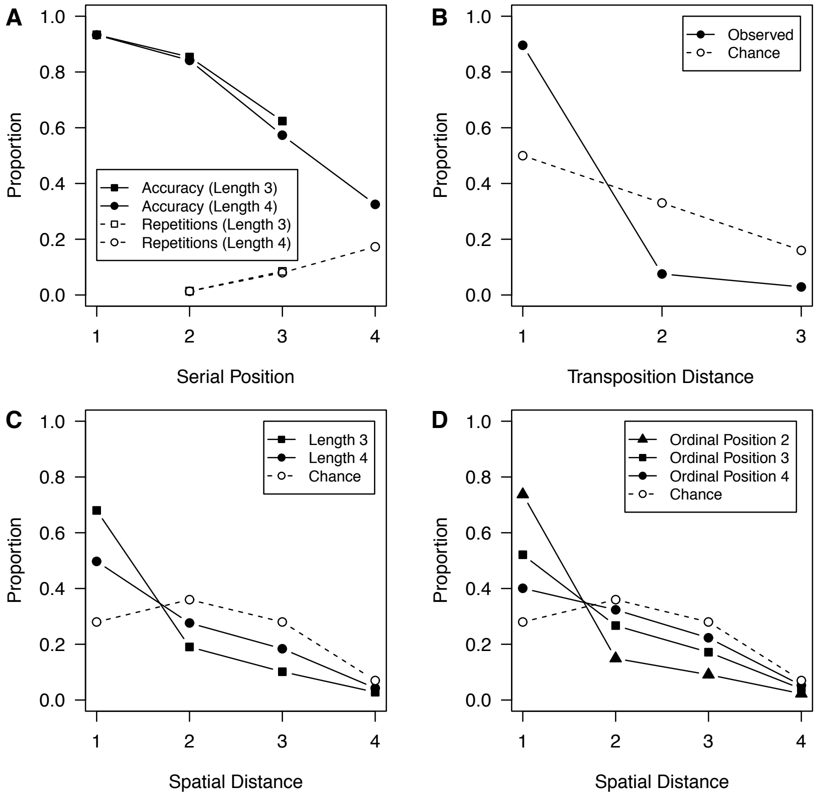

Fig. 4 Predictions of the general model for the four key behavioural measures: (**A**) accuracy and repetition error serial position curves, (**B**) transposition error gradient for four-item lists, (**C**) aggregate spatial error gradients, and (**D**) spatial error gradients for the second, third, and fourth output position in four-item lists.

straint is a consequence of the exponential relation between similarity and spatial distance embodied in the weights linking item units in the activation and selection layers (equations 1 and 2). This means that when activation from a winning item in the activation layer propagates through to the selection layer, items that are spatially near to the target item will be activated more strongly than items that are spatially far from the target item. Accordingly, when the item unit activations are augmented with noise, near-neighbour spatial errors will be more likely than distant-neighbour spatial errors.

Repetition errors

The repetition error serial position curves predicted by the general model are shown in Fig 4A. Consistent with the empirical data (Fig 2A), repetitions were infrequent and their probability increased with output position. The scarcity of repetitions in the model is due to the operation of response suppression—the suppression of a recalled item's activation in the activation layer removes it briefly from the response competition, rendering it unlikely the item will be reported twice. The probability of repetitions increases with output position due to the greater time available for a suppressed item to recover from inhibition.

Protrusion errors

The general model predicted a protrusion rate of 31.24% (289 protrusions out of 925 immediate intrusions), which compares favourably with the rate of 32.5% (51 protrusions out of 150 immediate intrusions) exhibited by the monkey. Immediate intrusions arise in the model because the context-item associations formed on different trials are stored in a common Hebbian weight matrix—the associations formed on trial *n* will be superimposed on the associations established on all trials so far. Thus, when a context cue for a given position is presented to the activation layer on trial *n*, as well as activating the target item on trial *n* and neighbouring list competitors, the context cue will activate the target item and neighbouring list competitors from trial $n - 1$ (and to a lesser degree items from earlier trials whose associations are less strongly represented in the weight matrix due to the weight normalisation that occurs at the end of each trial, which gradually washes out the influence of these associations over time). With the addition of random noise to item activations, sometimes the discrepancy between the activations of items from trial n and trial $n-1$ will be bridged, triggering an intrusion. However, because the context-item association of the item on trial $n-1$ that occupied the position being cued will bear a stronger similarity to the current context cue than the associations of items from different positions on trial n-1, intrusions will most often be protrusions.

Fill-in errors

The general model produced a fill-in ratio of 3.88:1 (101 fill-in errors vs. 26 infill errors), which is slightly smaller than the observed ratio of 4.6:1 for the monkey's data (46 fill-in errors vs. 10 infill errors). The predominance of fill-in events in the general model is attributable to the representation of serial order by a primacy gradient and response suppression. This means that when an item i is reported a position too soon and then suppressed, item $i-1$ will be a stronger recall competitor at the next output position than item $i+1$ because the former item—by virtue of occurring earlier on the study list—will have been encoded with greater strength on the primacy gradient.

Discussion

Summary of findings

Using computational simulations, the current study sought to establish whether the spatial ISR performance of a rhesus monkey could be explained in terms of the seriating mechanisms embodied in computational models of human ISR. The results suggest that four core mechanisms of human ISR—viz. competitive queuing, position marking, a primacy gradient, and response suppression—are also implicated in the representation and control of serial order in this primate species. The model comparisons confirmed that all four mechanisms are necessary to accurately reproduce the monkey's performance. That is, the correspondence between the data and simulation results was severely compromised when one or more of the seriating mechanisms was eliminated from the general model. In what follows, I will describe those aspects of the monkey's data that confer support for the different seriating mechanisms, consider some limitations of the current modelling exercise, and identify directions for future work.

Evidence for the four mechanisms

At the outset, it was noted that strikingly direct evidence has been obtained from electrophysiological recording data that rhesus monkeys use the competitive queuing mechanism to plan and produce sequences (Averbeck et al 2002, 2003a,b). The results of the current simulation study provide converging evidence for this proposition by showing that a model based on competitive queuing principles can reproduce the key behavioural features of a rhesus monkey's spatial ISR performance. To my knowledge, this is the first computational application of the competitive queuing mechanism to serial order behaviour in a primate.

It is the model's capacity to reproduce the pattern of recall errors seen in the monkey's spatial ISR performance that is most emblematic of the operation of the competitive queuing mechanism. Like all competitive queuing models, a defining feature of the current model is that items are (re)activated in parallel at retrieval. There are two sources of parallel response activation in the present model. The first originates from the cueing of items in the activation layer by the context vectors in the context layer. This results in the target item being activated maximally, with neighbouring items being activated based on their ordinal distance from the target item in the input list. The second originates from the retrieval of an item in the activation layer, which activates items in the selection layer based on their similarity to the retrieved item. Since the retrieved item will be most similar to itself, its unit will be activated maximally, whereas the activation of other units will fall off with increasing spatial distance from the retrieved item. The addition of moderate random noise to item units in the activation and selection layers is the basis by which the model generates transposition and spatial errors, respectively. Since the activations of items in both layers follow a gradient that is centred on the target item, transposition and spatial errors are most likely to involve items that are close in ordinal or spatial distance to the target item. It is this parallel gradient-based response activation in the activation and selection layers—and the two-stage sequence production process—that enables the model to reproduce the positional and spatial locality constraints that are key features of the monkey's data.

The use of a two-stage sequence production process to model spatial errors might seem unparsimonious. Why not model spatial errors in the activation layer and do away with the selection layer? Empirically, the observation of both a positional and a spatial locality constraint suggests that the effects of positional and spatial uncertainty arise independently, rather than additively. Indeed, an attempt to model the data within a single stage by superimposing an activation gradient representing the spatial uncertainty of items over the activation gradient representing the positional uncertainty of items in the activation layer was unsuccessful. Specifically, this single-stage model generated violations of both the positional and spatial locality constraints—the transposition error and spatial error gradients both exhibited marked non-monotonicities, at variance with the empirical data. Note also that in modelling phonological similarity effects in human verbal STM, the current dominant view is that a two-stage sequence production process is required, with the detrimental effects of phonological similarity occurring downstream of an initial serial ordering stage (Burgess and Hitch 1999; Henson 1998; Page and Norris 1998).

Although all errors generated by the model depend upon the parallel sequence dynamics of the competitive queuing mechanism, the occurrence of certain errors is a consequence of the specific way serial order is represented. Protrusion errors are a consequence of the representation of serial order via position marking. The fill-in effect is a consequence of the representation of serial order by a primacy gradient coupled with response suppression. If the latter two mechanisms are eliminated, then the resulting model (restrict^{PM}) predicts the same frequency of fill-in and infill events (this is because after item i has been recalled a position too soon, items $i - 1$ and $i + 1$ will be activated to an equivalent degree by the position marker for the next output position), at variance with the monkey's performance. The primacy gradient is also necessary to explain the extensive primacy effect and flattening of the spatial error gradients over output positions. When this mechanism is eliminated (restrict^{PM+RS} and restrict^{PM}), the primacy effect is greatly abated and the spatial error gradients over output position become superimposed on one other. The main source of evidence for response suppression is the scarcity of repetitions exhibited by the model, and the increase in these errors across output positions. The latter result arises due to the gradual wearing off of response suppression over output positions. When this mechanism is eliminated from the model (restrict^{PG+PM} and restrict^{PM}), much higher levels of recall accuracy are observed at the third and fourth output positions than observed empirically. This is because to compensate for the omission of response suppression, higher levels of accuracy at these positions are required to prevent the model from generating excessive levels of erroneous repetitions.

Potential limitations and future directions

Although the model provides an excellent account of the observed hallmark effects, there are some limitations of the current modelling exercise. First, the data and modelling are based on a single animal. The theoretical inferences drawn here will therefore need to be verified by showing that the animal's ISR performance is representative of other animals of the same species. Of particular interest is whether the absence of recency is a defining feature of the ISR performance of macaque monkeys in general. It is noteworthy that Fagot and De Lillo (2011) also failed to observe recency in their study of the ISR performance of two baboons, tentatively suggesting that recency in ISR is a uniquely human attribute. However, in their study recency was merely absent, whereas in Botvinick et al (2009) recency was present but in a negative, rather than a positive, direction. Second, the explanation of negative recency in terms of selective encoding is speculative. As noted by Botvinick et al (2009) it could have been the consequence of the relatively long duration of recall (7.5 s on average for three-item lists and 10.9 s on average for four-item lists) providing the opportunity for time-based decay or interference of memory representations. In modelling the present data, I considered this as well as several other competing explanations, however, in practice the selective encoding account was the only mechanism by which I was able to reproduce negative recency and the correct underlying distributions of recall errors.

Although the monkey's ISR performance is qualitatively consistent with that of humans—negative recency notwithstanding—the maximum number of items it can retain is evidently much smaller. The sharp drop in recall performance after the second serial position implies a memory span of around two-items, an empirical estimate that harmonizes well with the results obtained in other serial memory studies with primates (Barone and Joseph 1989; Funahashi et al 1997; Ninokura et al 2003; Scarf et al 2011). Human adults, by comparison, have a memory span in the region of six-items or greater. One question for future work is how to explain these different memory limits in the model outlined here? In some competitive queuing models—known as normalised competitive queuing models (Bullock 2004)—the recurrent excitation and lateral inhibition implemented in the selection layer is also instantiated in the activation layer but in a way that encourages the maintenance of an activation pattern, rather than a winner-takes-all competition. The effect of this is to impose a neural bandwidth limit on the activation layer's capacity to maintain the relative priority of a set of parallel activated memory representations. It is possible that the smaller capacity limit seen in primates is the result of less available neural bandwidth with which to support parallel activated representations, perhaps due to weaker levels of self-excitation or stronger levels of lateral inhibition.

Concluding remarks

The present computational simulation study provides provisional evidence that four mechanisms of human ISR—competitive queuing, position marking, a primacy gradient, and response suppression—are implicated in the representation and generation of serial order in rhesus monkeys. The observation of these mechanisms in a species of old world monkey implies that they are ancient features in the evolution of primates rather than bespoke components of human cognition. This in turn suggests that the same mechanisms might support sequential memory processing in other primate species.

Compliance with ethical standards

Conflict of interest

The author declares that he has no conflict of interest.

Ethical approval

This article does not contain any studies with human participants or animals performed by the author.

References

Averbeck BB, Chafee MV, Crowe DA, Georgopoulos AP (2002) Parallel processing of serial movements in prefrontal cortex. Proceedings of the National Academy of Sciences 99(20):13,172–13,177

Averbeck BB, Chafee MV, Crowe DA, Georgopoulos AP (2003a) Neural activity in prefrontal cortex during copying geometrical shapes: I. single cells encode shape, sequence, and metric parameters. Experimental Brain Research 150(2):127–141

Averbeck BB, Crowe DA, Chafee MV, Georgopoulos AP (2003b) Neural activity in prefrontal cortex during copying geometrical shapes: II. decoding shape segments from neural ensembles. Experimental Brain Research 150(2):142–153

Baddeley AD (1966) Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. The Quarterly Journal of Experimental Psychology 18(4):362–365

Baddeley AD (1968) How does acoustic similarity influence short-term memory? The Quarterly Journal of Experimental Psychology 20(3):249–264

Barone P, Joseph JP (1989) Prefrontal cortex and spatial sequencing in macaque monkey. Experimental Brain Research 78(3):447–464

Biro D, Matsuzawa T (1999) Numerical ordering in a chimpanzee (*Pan troglodytes*): Planning, executing, and monitoring. Journal of Comparative Psychology 113(2):178–185

Botvinick MM, Plaut DC (2006) Short-term memory for serial order: A recurrent neural network model. Psychological Review 113(2):201–233

Botvinick MM, Wang J, Cowan E, Roy S, Bastianen C, Mayo JP, Houk JC (2009) An analysis of immediate serial recall performance in a macaque. Animal Cognition 12(5):671–678

Brown GD, Preece T, Hulme C (2000) Oscillator-based memory for serial order. Psychological Review 107(1):127–181

Bullock D (2004) Adaptive neural models of queuing and timing in fluent action. Trends in Cognitive Sciences 8(9):426–433

Burgess N, Hitch GJ (1999) Memory for serial order: a network model of the phonological loop and its timing. Psychological Review 106(3):551–581

Burnham KP, Anderson DR (2002) Model selection and multimodal inference: A practical information-theoretic approach. New York, NY: Springer-Verlag

Carpenter AF, Georgopoulos AP, Pellizzer G (1999) Motor cortical encoding of serial order in a context-recall task. Science 283(5408):1752–1757

Chen S, Swartz KB, Terrace H (1997) Knowledge of the ordinal position of list items in rhesus monkeys. Psychological Science 8(2):80–86

Conrad R (1960) Serial order intrusions in immediate memory. British Journal of Psychology 51(1):45–48

Conrad R, Hull AJ (1964) Information, acoustic confusion and memory span. British Journal of Psychology 55(4):429–432

Crannell CW, Parrish JM (1957) A comparison of immediate memory span for digits, letters, and words. The Journal of Psychology 44(2):319–327

D'Amato M, Colombo M (1988) Representation of serial order in monkeys (*Cebus apella*). Journal of Experimental Psychology: Animal Behavior Processes 14(2):131–139

D'amato M, Colombo M (1989) Serial learning with wild card items by monkeys (*Cebus apella*): Implications for knowledge of ordinal position. Journal of Comparative Psychology 103(3):252

Fagot J, De Lillo C (2011) A comparative study of working memory: Immediate serial spatial recall in baboons (*Papio papio*) and humans. Neuropsychologia 49(14):3870–3880

Farrand P, Jones D (1996) Direction of report in spatial and verbal serial short term memory. The Quarterly Journal of Experimental Psychology: Section A 49(1):140–158

Farrand P, Parmentier FB, Jones DM (2001) Temporal–spatial memory: Retrieval of spatial information does not reduce recency. Acta Psychologica 106(3):285–301

Farrell S (2006) Mixed-list phonological similarity effects in delayed serial recall. Journal of Memory and Language 55(4):587–600

Farrell S, Lewandowsky S (2002) An endogenous distributed model of ordering in serial recall. Psychonomic Bulletin & Review 9(1):59–79

Farrell S, Hurlstone MJ, Lewandowsky S (2013) Sequential dependencies in recall of sequences: Filling in the blanks. Memory & Cognition 41(6):938–952

Funahashi S, Inoue M, Kubota K (1997) Delay-period activity in the primate prefrontal cortex encoding multiple spatial positions and their order of presentation. Behavioural Brain Research 84(1):203–223

Glasspool DW (2005) Modelling serial order in behaviour: Evidence from performance slips. In: Connectionist models in cognitive psychology, Hove, England: Psychology Press, pp 241–270

Guérard K, Tremblay S (2008) Revisiting evidence for modularity and functional equivalence across verbal and spatial domains in memory. Journal of Experimental Psychology: Learning, Memory, and Cognition 34(3):556–569

Hartley T, Hurlstone MJ, Hitch GJ (2016) Effects of rhythm on memory for spoken sequences: A model and tests of its stimulus-driven mechanism. Cognitive psychology 87:135–178

Henson RN (1998) Short-term memory for serial order: The start-end model. Cognitive Psychology 36(2):73–137

Henson RN (1999) Positional information in short-term memory: Relative or absolute? Memory & Cognition 27(5):915–927

Henson RNA (1996) Short-term memory for serial order Unpublished doctoral dissertation, University of Cambridge

Henson RNA, Norris DG, Page MPA, Baddeley AD (1996) Unchained memory: Error patterns rule out chaining models of immediate serial recall. The Quarterly Journal of Experimental Psychology: Section A 49(1):80–115

Hitch GJ (1974) Short-term memory for spatial and temporal information. The Quarterly Journal of Experimental Psychology 26(3):503–513

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

Hurlstone MJ, Hitch GJ (2015) How is the serial order of a spatial sequence represented? Insights from transposition latencies. Journal of Experimental Psychology: Learning, Memory, and Cognition 41(2):295–324

Hurlstone MJ, Hitch GJ, Baddeley AD (2014) Memory for serial order across domains: An overview of the literature and directions for future research. Psychological Bulletin 140(2):339–373

Inoue S, Matsuzawa T (2007) Working memory of numerals in chimpanzees. Current Biology 17(23):R1004–R1005

Jones D, Farrand P, Stuart G, Morris N (1995) Functional equivalence of verbal and spatial information in serial short-term memory. Journal of Experimental Psychology: Learning, Memory, and Cognition 21(4):1008–1018

Kawai N, Matsuzawa T (2000) Numerical memory span in a chimpanzee. Nature 403(6765):39

Kermadi I, Joseph J (1995) Activity in the caudate nucleus of monkey during spatial sequencing. Journal of Neurophysiology 74(3):911–933

Lewandowsky S, Farrell S (2008) Short-term memory: New data and a model. Psychology of Learning and Motivation 49:1–48

Matsuzawa T (1985) Use of numbers by a chimpanzee. Nature 315(6014):57–59

Maybery MT, Parmentier FBR, Jones DM (2002) Grouping of list items reflected in the timing of recall: Implications for models of serial verbal memory. Journal of Memory and Language 47(3):360–385

Nelder JA, Mead R (1965) A simplex method for function minimization. The Computer Journal 7(4):308–313

Ninokura Y, Mushiake H, Tanji J (2003) Representation of the temporal order of visual objects in the primate lateral prefrontal cortex. Journal of Neurophysiology 89(5):2868–2873

Orlov T, Yakovlev V, Hochstein S, Zohary E (2000) Macaque monkeys categorize images by their ordinal number. Nature 404(6773):77–80

Orlov T, Yakovlev V, Amit D, Hochstein S, Zohary E (2002) Serial memory strategies in macaque monkeys: Behavioral and theoretical aspects. Cerebral Cortex 12(3):306–317

Orlov T, Amit DJ, Yakovlev V, Zohary E, Hochstein S (2006) Memory of ordinal number categories in macaque monkeys. Journal of Cognitive Neuroscience 18(3):399–417

Page M, Norris D (1998) The primacy model: A new model of immediate serial recall. Psychological Review 105(4):761–781

Parmentier FB, Andrés P, Elford G, Jones DM (2006) Organization of visuo-spatial serial memory: Interaction of temporal order with spatial and temporal grouping. Psychological Research 70(3):200–217

Rerko L, Oberauer K, Lin HY (2014) Spatial transposition gradients in visual working memory. The Quarterly Journal of Experimental Psychology 67(1):3–15

Ryan J (1969a) Grouping and short-term memory: Different means and patterns of grouping. The Quarterly Journal of Experimental Psychology 21(2):137–147

Ryan J (1969b) Temporal grouping, rehearsal and short-term memory. The Quarterly Journal of Experimental Psychology 21(2):148–155

Scarf D, Danly E, Morgan G, Colombo M, Terrace HS (2011) Sequential planning in rhesus monkeys (*Macaca mulatta*). Animal Cognition 14(3):317–324

Smyth MM, Scholey KA (1996) Serial order in spatial immediate memory. The Quarterly Journal of Experimental Psychology: Section A 49(1):159–177

Surprenant AM, Kelley MR, Farley LA, Neath I (2005) Fill-in and infill errors in order memory. Memory 13(3-4):267–273

Terrace HS (2005) The simultaneous chain: A new approach to serial learning. Trends in Cognitive Sciences 9(4):202–210

Vousden JI, Brown GDA (1998) To repeat or not to repeat: The time course of response suppression in sequential behaviour. In: 4th Neural Computation and Psychology Workshop, London, 9–11 April 1997, Springer, pp 301–315

Wagenmakers EJ, Farrell S (2004) AIC model selection using Akaike weights. Psychonomic Bulletin & Review 11(1):192–196