# Integral Supplementary Material for "How Is the Serial Order of a Visual Sequence Represented? Insights From Transposition Latencies"

Mark J. Hurlstone
University of Western Australia

Graham J. Hitch
University of York

Hurlstone and Hitch (2017) used computational simulations to compare the transposition error latency predictions of five different models of serial order. This integral supplementary document reports detailed information about the generic network architecture used for the simulations; how the different representational principles and models were implemented; and the model fitting and evaluation procedures. This document is not meant as a stand-alone paper (the sections contained herein are effectively appendices to the main article)—please refer to Hurlstone and Hitch (2017) for further explanation.

## Formal Description of the Generic Network Architecture Used to Model Transposition Latencies

Following Farrell and Lewandowsky (Farrell & Lewandowsky, 2004; Lewandowsky & Farrell, 2008) and our own earlier modeling (Hurlstone & Hitch, 2015), we did not utilize a fully implemented competitive queuing architecture for our simulations, but instead employed a single layer lateral inhibition network corresponding to the competitive choice layer in competitive queuing models. For each of the representational principles being modeled, we specified the profile of activations that would be expected initially at each output position in the parallel planning layer, before feeding that pattern of activations into the lateral inhibition network in order to generate an unambiguous response and an associated recall latency. Thus, we did not simulate the process of encoding serial order, since the selection mechanism is insensitive to the exact mechanisms generating the initial activations used to drive recall.

### A Common Lateral Inhibition Response Selection Network

A schematic of the response selection network employed for the simulations is illustrated in Figure 1. It consists of a single competitive layer of localist item nodes corresponding to the pool of response elements from which sequences can be generated. Each node has a recurrent self-excitatory connection, plus lateral inhibitory connections to all other nodes. The excitatory and inhibitory weights are a hardwired property of the network and were set to constant values of 1.1 and -0.1, respectively. This network acts as a "competitive filter" that selects a single response from amongst a set of parallel activated representations. As noted earlier, retrieval involves first establishing starting activation values for the item nodes based on the representational principles being modeled (see later). The activations are then iteratively passed through the weights until only one node remains active above a response threshold. The netinput $\text{net}_j$ a node $j$ receives from within the network is determined by the following equation:

$$\text{net}_j(t) = a_j(t-1)\,\alpha + \beta \sum_{i \neq j} a_i(t-1) + \epsilon(0, \sigma), \quad (1)$$

Where $a_j$ is the activation of node $j$, $a_i$ is the activation of all other nodes in the layer, $\alpha$ and $\beta$ represent the excitatory and lateral inhibitory weight values, respectively, and $t$ corresponds to time. To model transposition errors, the node activations are supplemented with zero-mean Gaussian noise $\epsilon$ with standard deviation $\sigma$ ($\sigma = .04$ for all simulations).

The first term on the right hand side of equation 1 represents the recurrent self excitation, whereas the second term represents the lateral inhibition received from

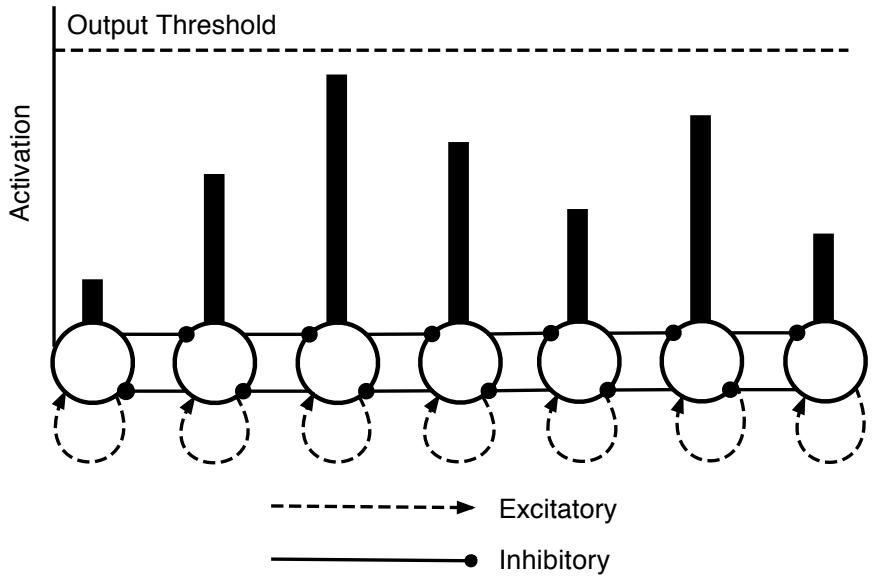


*Figure 1*. A schematic of the lateral inhibition neural network architecture employed for the simulations. The network is modeled on the competitive choice layer employed in competitive queuing models of serial behavior (Bullock, 2004; Rhodes & Bullock, 2003). Each localist item node possesses a single recurrent excitatory connection as well as lateral inhibitory connections to all other item nodes. Nodes are fully interconnected, but only adjacent-neighbor inhibitory connections are shown to prevent visual cluttering. Note—the number of nodes in the network is dependent on the sequence length being modeled.

all other nodes in the layer. This sets up a *winner-takes-all* response competition over the item nodes, and the initially most active node has the advantage that it will send more activation to itself than any other node, and will also receive the least lateral inhibition. Equation 1 is iteratively applied by passing the activation pattern at time $t$-1 through the weights to determine the activation pattern at time $t$. Thus, after a starting activation pattern has been established over the item nodes the competitive dynamics of the network will result in a gradual increase in the activation of the strongest node, and a gradual decrease in the activations of the weaker nodes as they receive more lateral inhibition. The iterations stop when the strongest node exceeds a response threshold $T$ ($T$ = 1.0 for all simulations) and the number of iterations required to determine the response is taken as the network's recall latency.

In order to bring the predicted recall times of the network within the range of the observed latencies in the experiments they were multiplied by a scaling factor $S$ ($0 < S \leq 200$; where $S = 50$ for the initial simulations—1 iterative cycle = 50 ms).

**Implementation of Representational Principles**

The representational principles were implemented through different settings of the starting activations at each output position, which were computed as follows:

**Position marking.** Position marking was implemented by specifying activations for item nodes that reflected the distances between item positions. Specifically, the activation $a$ of the item node $j$ at output position $p$ was strongest, whilst the activations of item nodes from neighboring serial positions decreased as an accelerating function of their distance from the target item:

$$a_j = \Omega\theta^{(|j-p|)}, \tag{2}$$

Where $\theta$ is a parameter controlling the distinctiveness of the position marking activations ($0 < \theta \leq 1$; $\theta$ = .65 for the initial simulations) and $\Omega$ is a weighting parameter that determines the distance of each item's initial activation from the response threshold ($0 < \Omega \leq 1$; $\Omega = 1$ for the initial simulations). For each output position, the activations generated by $\theta$ were rescaled to sum to 1—calculated by dividing each node's activation

by the sum of the activations of all nodes—before they were multiplied by Ω. This representational scheme produces gradients of activations akin to those generated by the positional context signals in the Burgess and Hitch (1999) and OSCAR (Brown et al., 2000) models. Figure 2A shows example starting activations for position marking for the fourth output position in a six-item sequence.

**Primacy gradient.** The primacy gradient was implemented as a decrease in activations across input positions. The activation of each node was determined by:

$$a_j = \phi\rho^{(j-1)}, \tag{3}$$

Where $\phi$ is the activation of the item node corresponding to the first input position ($0 < \phi \leq 1$; $\phi = .6$ for the initial simulations) and $\rho$ is a parameter controlling the steepness of the primacy gradient ($0 < \rho \leq 1$; $\rho = .85$ for the initial simulations). Retrieval commenced by imposing the entire primacy gradient over the item nodes at the first output position and allowing activation to accumulate towards a response. This process was then repeated for each subsequent output position by imposing the same primacy gradient over the item nodes but with suppression (see below) of those nodes corresponding to previously recalled items. Example starting activations for a primacy gradient for the first output position are shown in Figure 2B.

**Primacy gradient + position marking.** In line with the seriating mechanisms instantiated in several theories of serial recall (Burgess & Hitch, 1999; Brown et al., 2000; Lewandowsky & Farrell, 2008), in some simulations serial order was represented through the combination of a primacy gradient and position marking by calculating starting activations as follows:

$$a_j = (1 - \omega)\,\phi\,\rho^{(j-1)} + \omega\,\Omega\,\theta^{(|j-p|)}, \tag{4}$$

Equation 4 integrates equations 2 and 3 above and incorporates an additional weighting parameter $\omega$ ($0 < \omega \leq 1$; $\omega = .5$ for the initial simulations) that governs the relative importance of the two representations of serial order. When $\omega = .5$ the two representations of order are weighted equally. However, when $\omega < .5$ more weight is given to the primacy gradient representation of order; conversely, when $\omega > .5$ more weight is given to the positional representation of order. Figure 2C shows example starting activations for the combination of a primacy gradient and position marking for the fourth output position.

**Response suppression.** Response suppression was implemented by reducing an item's activation once it had been recalled. For each output position, starting activation values were first calculated based on the other representational principles being modeled. The activations of nodes corresponding to items that had already been recalled were then multiplied by 1-$\tau$, where $\tau$ represents the extent of response suppression ($0 < \tau \leq 1$; $\tau = .95$ for the initial simulations). Example starting activations for a primacy gradient complemented by response suppression for the fourth output position are illustrated in Figure 2D.

**Output interference.** Output interference was modeled by assuming that recall of an item added noise to the activations of yet to be recalled items. Accordingly, random Gaussian noise with a standard deviation that increased as a function of output position was applied to the starting activations generated by the serial ordering principles being modeled (e.g., position marking) and was determined by $\delta \times \sigma \times p$, where $\delta$ is a parameter controlling the weighting of output interference across output positions ($0 < \delta \leq 1$; $\delta = .5$ for the initial simulations) and $\sigma$ is the standard deviation of noise applied to activations during the iterative updating process (see earlier). An example of the increase in the standard deviation of Gaussian noise applied to the starting activations across output positions is shown in Figure 2E.

## Five Models of Serial Order

The response probability and recall latency predictions of five models of serial order—built from different combinations of the four principles—were compared: (1) position marking (PM); (2) position marking and response suppression (PM + RS); (3) position marking and output interference (PM + OI); (4) a primacy gradient and response suppression (PG + RS); and (5) a primacy gradient, position marking, and response suppression (PG + PM + RS). These are the same set of models as those examined by Farrell and Lewandowsky (Farrell & Lewandowsky, 2004; Lewandowsky & Farrell, 2008) and Hurlstone and Hitch (2015), and they are representative of the range of mechanisms instantiated in contemporary theories of serial recall (see Table 1 of Hurlstone & Hitch, 2015 and Table 2 of Hurlstone,

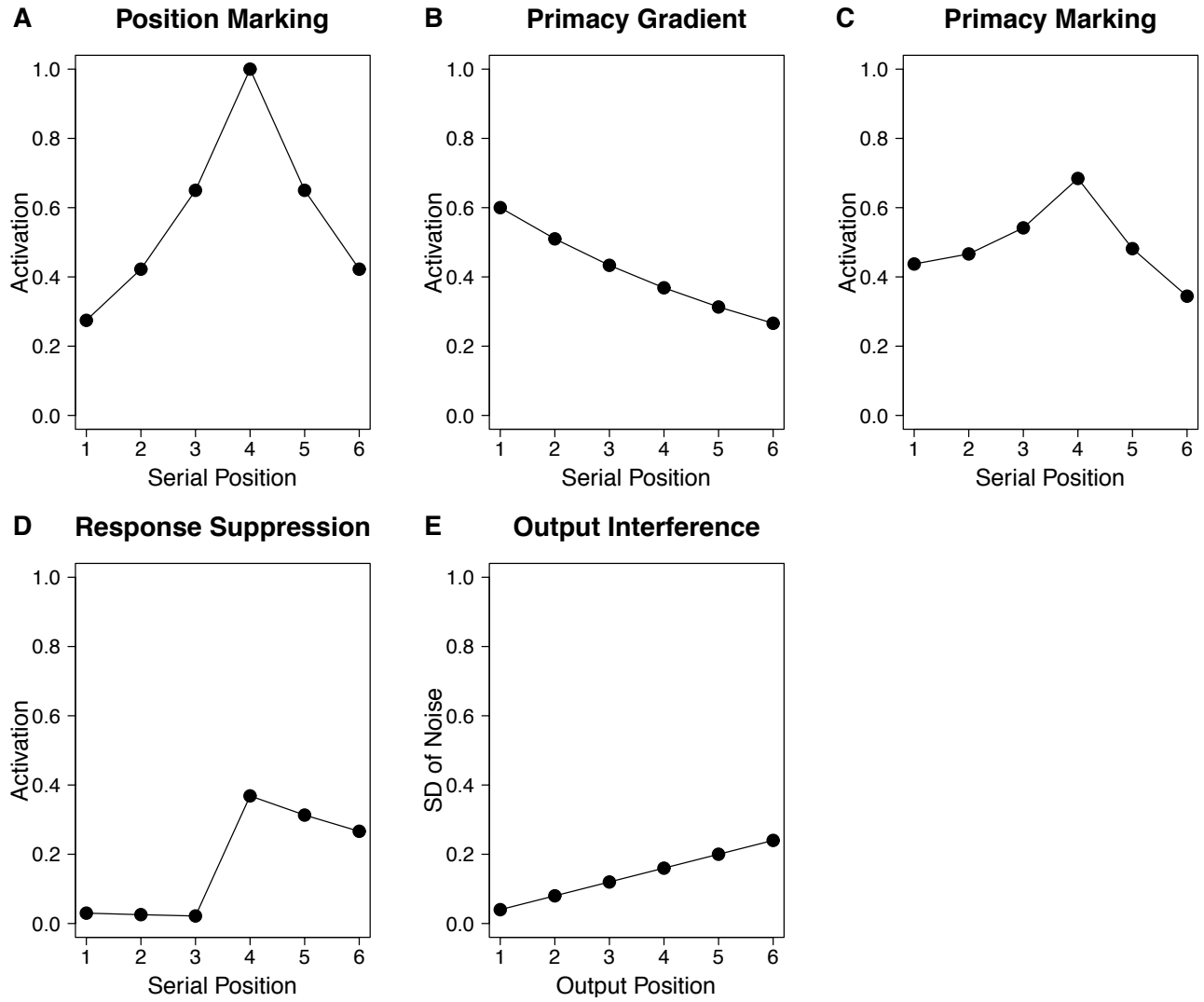


*Figure 2*. Example starting activations for the four representational principles—and combination of principles—for six-item sequences, based on the parameter settings employed for the initial simulations: (A) position marking (showing activations for the fourth output position), (B) a primacy gradient (showing activations for the first output position), (C) a primacy gradient and position marking (showing activations for the fourth output position), (D) response suppression (showing activations for a primacy gradient with suppression of the first three recalled items), (E) output interference (showing the increase in Gaussian noise applied to the starting activations across output positions).

Hitch, & Baddeley, 2014). Predictions were generated for each model using 50,000 simulation trials of six-item sequences.

### Extension to Grouped Sequences

In this section, we describe how the models implementing position marking can be extended to account for the recall of grouped sequences. Positional models of serial recall account for grouping effects by assuming that grouped sequences recruit two sets of position markers—one set that encodes the position of groups (Brown et al., 2000; Hartley et al., 2016; Henson, 1998; Lewandowsky & Farrell, 2008) or items (Burgess Hitch, 1999) in the sequence, and a second set

that encodes the position of items within groups. This multidimensional representation has been shown to be sufficient to account for the effects of grouping documented with verbal materials. In particular, the use of a set of position markers to represent the within-group position of items is crucial to explaining the between group interposition errors that are a hallmark feature of grouped verbal serial recall.

However, the failure to observe an increase in the frequency of interpositions in grouped visual serial recall (Experiment 3 in the main article; Hurlstone & Hitch, submitted) and grouped spatial serial recall (Hurlstone & Hitch, 2015; Experiments 1 & 3) raises the possibility that positional information might be represented differently in the visual and spatial domains. Specifically, Hurlstone and Hitch (2015) speculated that in the visual-spatial domain, position markers encoding the position of groups in sequence might be augmented by position markers encoding the position of items in the sequence overall, as opposed to within each group. This should in principle produce the usual effects of grouping on response probabilities, but without fostering an increase in the frequency of interpositions.

To provide a formal test of this hypothesis, we contrasted two approaches to extending the positional models to grouped sequences, one in which position markers encoding the position of groups in sequence were combined with position markers encoding the position of items within groups—viz. the standard approach to modeling grouping effects—and a second in which position markers encoding the position of groups in sequence were combined with position markers encoding the position of items in the sequence overall—viz. the revised approach proposed by Hurlstone and Hitch (2015).

In the following, we outline the equations used to calculate the starting activations for position marking for grouped sequences under the two different approaches to implementing grouped positional representations. When applying the PM, PM + RS, PM + OI, and PG + PM + RS models to grouped sequences, these equations were used in substitute for equation 2.

### Position of Group + Position Within Group

In the implementation of position marking combining information about the position of groups and the position of items within groups, starting activations were chosen that directly reflected the confusability of group positions in the sequence and item positions within groups:

$$a_j = (1 - \lambda)\Omega\theta^{(|g-l|)} + \lambda\Omega\theta^{(|i-r|)}, \tag{5}$$

Where $j$ indexes an item's input position, $g$ indexes its group's input position, $l$ represents the input position of the group of the target item to-be-recalled at the current output position, $i$ indexes the within-group input position of item $j$, and $r$ represents the within-group input position of the target item to-be-recalled at the current output position. To illustrate, suppose we wish to calculate the activation of the second item at the sixth output position in a six-item sequence grouped into threes. In this example, $g = 1$ (since item 2 belongs to the first group), $l = 2$ (since the target item 6 belongs to the second group), $i = 2$ (since item 2 appears in the second position in the first group), whilst $r = 3$ (since item 6 appears in the third position in the second group).

As in equation 2, the parameter $\theta$ governs the distinctiveness of the position markers ($0 < \theta \leq 1$), whilst $\Omega$ is a scaling parameter ($0 < \Omega \leq 1$). The first term in equation 5 generates gradients of activations representing the confusability of the positions of groups in the sequence, whereas the second term generates gradients of activations representing the confusability of the positions of items within groups. The parameter $\lambda$ weights the amount of attention allocated to the two positional dimensions ($0 < \lambda \leq 1$). When $\lambda = .5$, attention is directed equally to the two dimensions; when $\lambda < .5$, more attention is allocated to the group-position-in-sequence dimension of order; when $\lambda > .5$, more attention is allocated to the item-position-in-group representation of order. Example starting activations based on equation 5 for all output positions in a six-item sequence grouped into threes are shown in Figure 3.

### Position of Group + Position Within Sequence

In the implementation of position marking combining information about the position of groups and the position of items within sequence, starting activations were chosen that directly reflected the confusability of group and item positions within sequence:

$$a_j = (1 - \lambda)\Omega\theta^{(|g-l|)} + \lambda\Omega\theta^{(|j-p|)}, \tag{6}$$

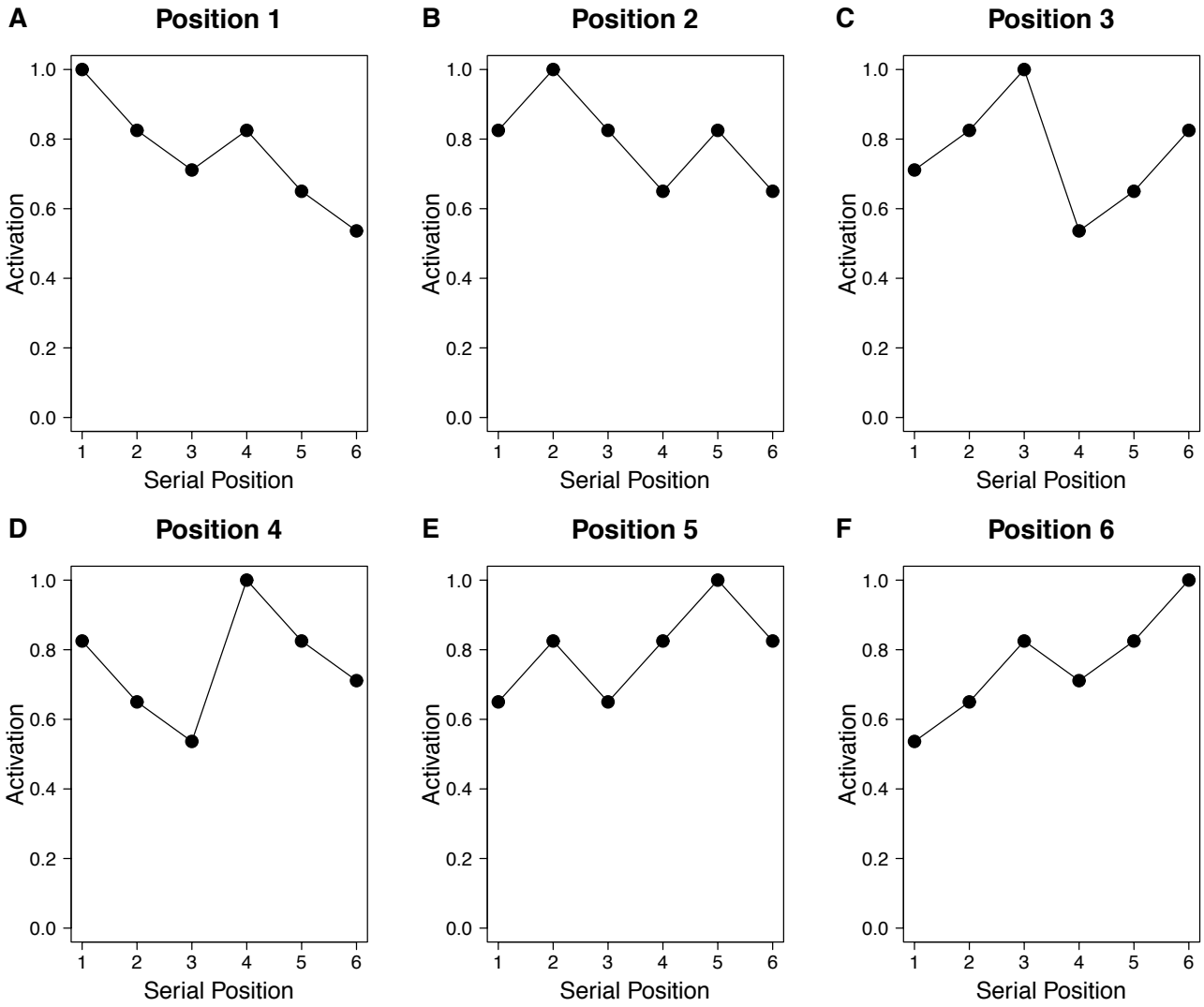


*Figure 3*. Example starting activations for the position-of-group and position-within-group implementation of position marking for a six-item sequence grouped into threes. Activations were generated using the following parameter values: $\Omega = 1$; $\theta = .65$; $\lambda = .5$.

Where $p$ represents the output position, and $g$, $l$, and $j$ are as before. To illustrate, using the earlier example of calculating the activation of the second item at the sixth output position in a six-item sequence grouped into threes, $g = 1$ (since item 2 belongs to the first group), $l = 2$ (since the target item 6 belongs to the second group), $j = 2$ (since this is the item whose activation is being calculated), and $p = 6$ (since this is the current output position). The first term in equation 6 is the same as in equation 5 and generates gradients of activations representing the confusability of the positions of groups in the sequence. The second term generates gradients of activations representing the confusability of the positions of items in the sequence, and is identical to equation 2 used to generate starting activations for position marking for ungrouped sequences. As in equation 5, the parameter $\lambda$ weights the amount of attention allocated to the two positional dimensions. Figure 4 shows example starting activations based on equation 6.

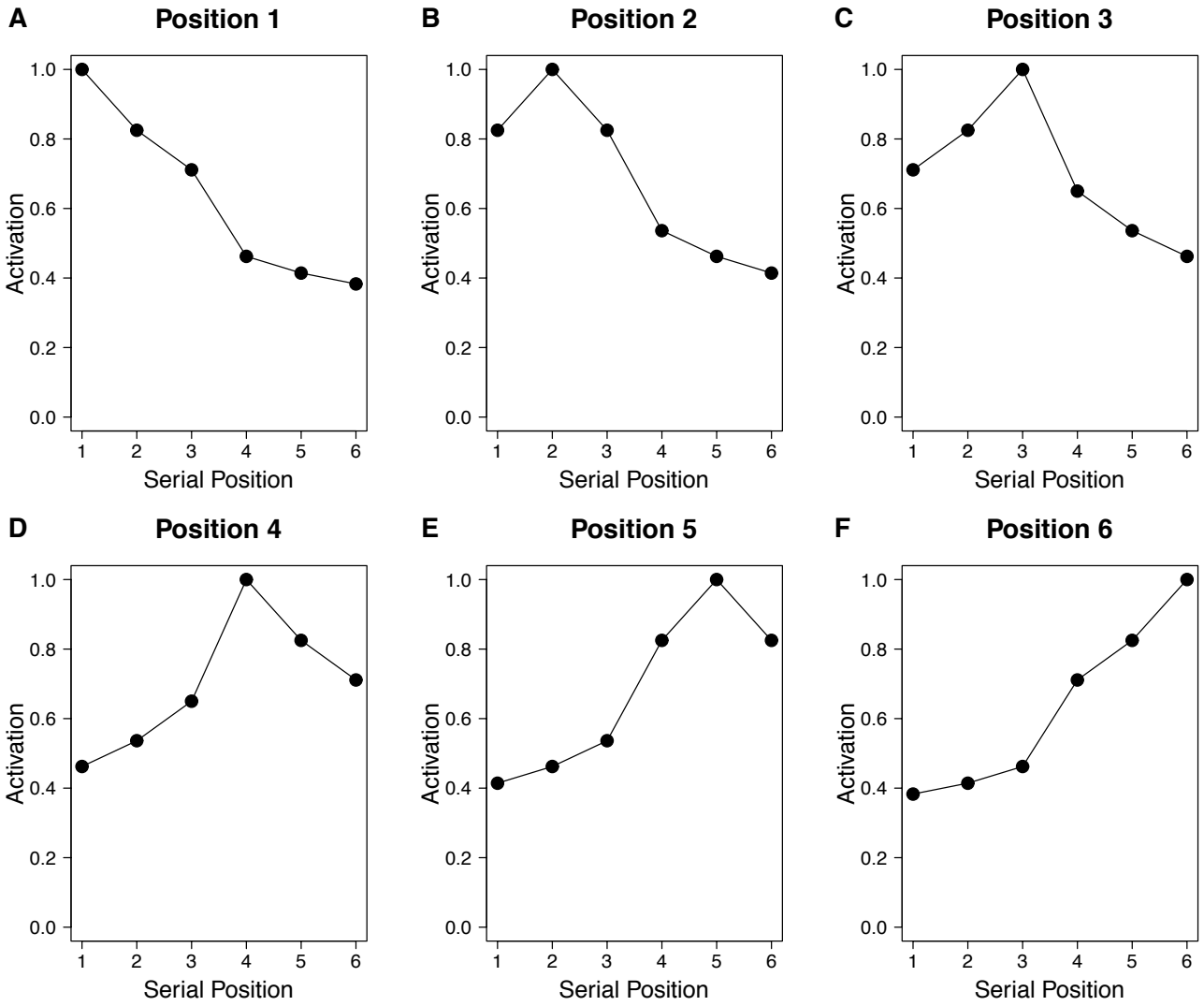


*Figure 4*. Example starting activations for the position-of-group and position-within-sequence implementation of position marking for a six-item sequence grouped into threes. Activations were generated using the following parameter values: $\Omega = 1$; $\theta = .65$; $\lambda = .5$.

### Incorporating a Primacy Gradient

When position marking was combined with a primacy gradient—viz. the PG + PM + RS model—equations 5 and 6 were augmented as follows:

$$a_j = (1 - \omega)(\text{Eq. 5} \mid \text{Eq. 6}) + \omega\phi\rho^{(j-1)}, \qquad (7)$$

Where Eq. 5 | Eq. 6 implements equation 5 or 6, $\phi\rho^{(j-1)}$ corresponds to equation 3 used to compute the primacy gradient over the input position of items, and $\omega$ is the same weighting parameter used in equation 4 to determine the attentional weight assigned to the primacy gradient and position markers.

### Description of Parameter Fitting Procedure

For each of the four target data sets, the models were fit to the response time distributions for each output position—with the effects of output position subtracted, as per the experiments and initial qualitative model predictions—using a maximum likelihood method for quantiles. For each output position, the response times associated with items recalled from different input posi-

tions on the input sequence were sorted in ascending order by each participant and the reaction times for the .5 quantiles that divided the data into two bins were calculated for each individual. These quantile reaction times were then averaged over participants to obtain group response time distributions.[1] The to-be-fitted data for each output position therefore took the form of a $2 \times sl$ (where $sl$ represents the sequence length) matrix of bins, where the rows represent categories defined by quantile-averaged statistics (one category corresponding to scores at or below the .5 quantile average; the second corresponding to scores above the .5 quantile average) and the columns represent input positions.[2] The values in each bin represent the number of recall latencies for a given category and input position. We opted to fit quantile-averaged group data—rather than individual participant data—to compensate for the fact that participants contributed few or no response times in bins corresponding to large output-input position displacements. However, we note that quantile-averaging preserves information about the shape of the individual response time distributions (Ratcliff, 1979) and yields parameter estimates that are comparable to the average parameter estimates obtained by fitting individual participants (Ratcliff, Thaper, & McKoon, 2001; Thaper, Ratcliff, & McKoon, 2003).

The models were used to obtain predicted frequencies in each bin for each output position. The discrepancy between observed and predicted frequencies was evaluated using a maximum likelihood criterion. Specifically, for a given model and set of parameter values, the likelihood of the observed response frequencies for a given output position $p$ was determined using the multinomial log-likelihood function:

$$\ln L_{pos}(p) = \sum_j \sum_i N_{ij} \ln(\pi_{ij}), \tag{8}$$

Where $N_{ij}$ is the frequency of observations in the bin for the $i$th category at the $j$th input position, $\pi_{ij}$ is the corresponding probability in the bin predicted by the model, and ln is the natural logarithm. The number of bins varied according to the sequence length of the data set being fitted. For four-item sequences, there were 8 bins per output position (32 in total); for five-item sequences there were 10 bins per output position (50 in total); for six-item sequences there were 12 bins per output position (72 in total). A total log-likelihood was then calculated by summing the individual log-likelihoods for each output position:

$$\ln L = \sum_{p=1}^{P} \ln L_{pos}(p), \tag{9}$$

Where $\ln L$ is the joint multinomial log-likelihood, which was converted to a negative value. Model parameters were varied systematically using the SIMPLEX function minimization algorithm (Nelder & Mead, 1965) until the smallest possible value of this objective function was obtained. Each parameter vector explored by the minimization algorithm involved 5,000 model simulation trials of the sequence length and grouping condition being simulated.

The parameters that were free to vary for the PM model were the weighting ($\Omega$) and distinctiveness ($\theta$) of the position markers. These free parameters were augmented by the amount of response suppression ($\tau$) in the PM + RS model and the amount of output interference ($\delta$) in the PM + OI model. The PG + RS model took as its free parameters the starting point ($\phi$) and steepness ($\rho$) of the primacy gradient, and the degree

[1]This procedure of forming group response time distributions is known as Vincentizing (or Vincent averaging). There are two different versions of this procedure (Andrews & Heathcote, 2001; Jian, Rouder, & Speckman, 2004). The first produces a group response time distribution by averaging over the quantiles—the values beneath which given proportions of the distribution occur—of individual response time distributions. The second produces a group response time distribution by averaging over the Vincentiles—the average of values between pairs of quantiles—of individual response time distributions. The approach adopted here is based on the first version of this procedure.

[2]We have also applied the model fitting and evaluation procedure described here to a larger number of bins using categories defined either by the .25, .5, and .75 quantile averages, or the .1, .3, .5, .7, and .9 quantile averages (both of which are common choices in the evidence accumulation model fitting literature). The relative fits of the models—examined using the goodness-of-fit criteria specified below—using these three- and five-quantile summaries was comparable to our single-quantile summary, except that the discrepancy between the observed and predicted accuracy serial position curves was too large in the former instances. Accordingly, we report the fits of the models using the single-quantile summary, which provided the best approximation of the accuracy and latency data of the three quantile summary approaches.

of response suppression ($\tau$). Finally, the free parameters for the PG + PM + RS model were the weighting ($\Omega$) and distinctiveness ($\theta$) of the position markers, the starting point ($\phi$) and steepness ($\rho$) of the primacy gradient, and the degree of response suppression ($\tau$). The weighting of the position markers and primacy gradient ($\omega$) was frozen to a value of .5.[3] For the simulations of grouped sequences, the weighting of the two sets of position markers ($\lambda$) in the two versions of the PM, PM + RS, PM + OI, and PG + PM + RS models was also set to a fixed value of .5.[4] In addition to the above-mentioned parameters, the iteration-to-ms scaling parameter $S$ was included as a free parameter in all models. In summary, the number of free parameters was three for the PM model, four for the PM + RS, PM + OI, and PG + RS models, and six for the PG + PM + RS model.

The models were initially fit to the data according to the procedure described above, which yielded for each data set and each model, a set of best fitting parameter values and an associated maximum log-likelihood estimate. However, as the models differ in their number of free parameters, it is necessary to augment this goodness-of-fit metric with a penalty term that punishes excessive model complexity. Accordingly, in order to provide a measure of the descriptive accuracy of the models that takes into consideration differences in their degree of complexity, the log-likelihood estimates were converted into Akaike and Bayesian information criterion scores (AIC, Akaike, 1973; BIC, Schwartz, 1978, respectively). The AIC was calculated as:

$$\mathrm{AIC}_i = -2 \ln L_i + 2 V_i, \tag{10}$$

Where $V$ is the number of free parameters involved in maximizing $\ln L$ and $i$ indexes the model for which AIC is being calculated (smaller values of AIC indicate a better fit). As can be seen from equation 10, the AIC rewards a model for its goodness-of-fit via its maximized log-likelihood and punishes it as a function of its number of free parameters. Similarly, the BIC was calculated as:

$$\mathrm{BIC}_i = -2 \ln L_i + V_i \ln(n), \tag{11}$$

Like the AIC, the BIC rewards a model for its goodness-of-fit via its maximized log-likelihood but punishes it as a function of the number of free parameters weighted by the number of observations $n$ entering into the log-likelihood calculation (smaller values of BIC indicate a better fit). Accordingly, the BIC offers a more stringent correction for model complexity. As the BIC was calculated on quantile-averaged group data, the value of $n$ in the penalty term was set equal to the average number of observations per participant in the data set.

To aid interpretation, the raw AIC and BIC scores were converted into so-called IC weights (Burnham & Anderson, 2002; Lewandowsky & Farrell, 2011; Wagenmakers & Farrell, 2004), which express the degree of support for each model on a continuous measure of evidence. The IC weight for model $i$ was calculated by:

$$w\mathrm{IC}_i = \frac{\exp(-0.5\,\Delta\mathrm{IC}_i)}{\sum_{k=1}^{K} \exp(-0.5\,\Delta\mathrm{IC}_k)}, \tag{12}$$

Where $\Delta\mathrm{IC}_i$ is the difference in IC between model $i$ relative to the best model, and each $\Delta\mathrm{IC}_k$ is the difference in IC between a specific model $k$ in the candidate set $K$ and the best model. These IC weights—normalized to sum to 1—represent the probability that each model is

[3] We opted to freeze $\omega$ because there is some redundancy in this parameter, given the other parameters varied in the fitting. Specifically, the free parameters $\phi$ and $\Omega$ will determine the relative weight given to the primacy gradient and position markers, thus removing the need to also systematically vary $\omega$. An alternative approach—adopted previously (Hurlstone & Hitch, 2015)—would be to fix the latter parameters to values of 1—rendering them inactive—and incorporate $\omega$ as a free parameter to govern the relative influence of the primacy gradient and position markers. Varying all three parameters simultaneously improves the quantitative fit of the PG + PM + RS model, but we have not pursued this option here since it is not necessary to bring out the qualitative effects in the data.

[4] We fixed the setting of $\lambda$ because the main focus of our model comparisons for grouped sequences was to establish whether the data is better understood in terms of a model combining position-of-group with position-within-group markers, or a model combining position-of-group with position-within-sequence markers. The problem with allowing $\lambda$ to vary is that it opens up the possibility that the models may preferentially engage only one of the two sets of position markers, which spoils this comparison (we consider the impact of varying $\lambda$ in the General Discussion of the main article). With $\lambda$ fixed, the magnitude of grouping effects in the simulations is governed by the setting of the free parameter $\theta$.

the best given the data and the competitor models under comparison. The support for a model is considered equivocal if its IC weight does not exceed $1/N$—where $N$ is the number of models under comparison. Thus, with five models, the support for a particular model is considered equivocal if its IC weight does not exceed 0.2.

We also conducted model comparisons using likelihood ratio tests (Lamberts, 2004), which compared the improvement in fit of a general model (viz. the PG + PM + RS model) with restricted versions of that model in which one of its components was removed (viz. PM + RS and PG + RS models). The likelihood ratio statistic is calculated by:

$$\chi^2 = -2[\ln L(\text{restricted}) - \ln L(\text{general})], \qquad (13)$$

Where $\ln L(\text{restricted})$ is the log-likelihood of the restricted model and $\ln L(\text{general})$ is the log-likelihood of the general model. The likelihood ratio statistic provides a measure of the reliability of the difference in goodness-of-fit between the restricted and general model. If $\chi^2$ exceeds the critical value corresponding to the 95th percentile of the chi-squared distribution—with degrees of freedom determined by the number of free parameters in the restricted model that were removed from the general model—then the null hypothesis that the restricted model is the best model can be rejected.

## References


Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19,* 716-723. doi:10.1109/TAC.1974.1100705

Andrews, S., & Heathcote, A. (2001). Distinguishing common and task-specific processes in word identification: A matter of some moment? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 514-544. doi: org/10.1037/0278-7393.27.2.514

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

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

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multi-modal inference: A practical information-theoretic approach.* New York, NY: Springer-Verlag. doi:10.1007/b97636

Farrell, S., & Lewandowsky, S. (2004). Modeling transposition latencies: Constraints for theories of serial order memory. *Journal of Memory and Language, 51,* 115-135. doi:10.1016/j.jml.2004.03.007

Hartley, T., Hurlstone, M. J., & Hitch, G. J. (2016). Effects of rhythm on memory for spoken sequences: A model and tests of its stimulus-driven mechanism. *Cognitive Psychology, 87,* 135-178. doi:10.1016/j.cogpsych.2016.05.001

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

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin, 140,* 339-373. doi:10.1037/a0034221

Hurlstone, M. J., & Hitch, G. J. (2015). How is the serial order of a spatial sequence represented? Insights from transposition latencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 295-324. doi:10.1037/a0038223295

Hurlstone, M. J., & Hitch, G. J. (2017). How is the serial order of a visual sequence represented? Insights from transposition latencies. *Manuscript submitted for publication.*

Jiang, Y., Rouder, J. N., & Speckman, P. L. (2004). A note on the sampling properties of the Vincentizing (quantile averaging) procedure. *Journal of Mathematical Psychology, 48,* 186-195. doi:10.3758/BF03196589

Lamberts, K. (2004). Mathematical Modelling of Cognition. In K. Lamberts and R. Goldstone (Eds.), *Handbook of Cognition* (pp. 407-421). London: Sage.

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. *The Psychology of Learning and Motivation, 49,* 1-48. doi:10.1016/S0079-7421(08)00001-7

Lewandowsky, S., Farrell, S. (2011). Computational modeling in cognition: Principles and practice. Thousand Oaks, CA: Sage. doi:10.4135/9781483349428

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal, 7,* 308-313. doi:10.1093/comjnl/7.4.308

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin, 86,* 446-461. doi:10.1037/0033-2909.86.3.446

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging, 16,* 323-341. doi:10.1037/f0882-7974.16.2323

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6,* 461-464. doi:10.2307/2958889

Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging, 18,* 415-429. doi:10.1037/0882-7974.18.3.415

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11,* 192-196. doi:10.3758/BF03206482