



Functional similarities and differences between the coding of positional information in
verbal and spatial short-term order memory

Mark J. Hurlstone

University of Western Australia and University of York

Graham J. Hitch

University of York

Word count: 8,784 (approximate count due to use of \LaTeX)

Author Note

Mark J. Hurlstone: School of Psychology, University of Western Australia, and Department of Psychology, University of York; Graham J. Hitch: Department of Psychology, University of York. This paper is based on part of the first author's doctoral dissertation completed at the University of York, England, which was supported by a research studentship from the Economic and Social Research Council of the United Kingdom. The first author is now based at the University of Western Australia. Correspondence concerning this article should be addressed to Mark Hurlstone, School of Psychology, University of Western Australia, Crawley, W.A. 6009, Australia. Email: mark.hurlstone@uwa.edu.au. URL: <http://mark-hurlstone.github.io>

Abstract

The presence of temporal grouping effects in verbal and spatial serial recall suggests that the representation of serial order in verbal and spatial short-term memory (STM) incorporates positional information. However, not all effects of grouping are created equal in the verbal and spatial domains. Although grouping a sequence of verbal items engenders an increase in between-group transpositions that maintain their within-group position, grouping a sequence of spatial items does not engender an increase in these so-called interposition errors. We test the hypothesis that this discrepancy reflects a difference in the way positional information is represented in verbal and spatial STM. An experiment is reported in which grouped verbal and spatial serial recall were compared when the presentation and recall methodology was equated in the two tasks. Consistent with previous studies, grouping engendered an increase in interpositions in the verbal, but not the spatial, serial recall task. We subsequently applied two positional models of serial recall—that differed in their assumptions regarding how position is coded in STM—to the data from the experiment and several previous serial recall experiments. The results indicate that in verbal STM, groups are coded for their position in a sequence and items are coded for their position in a group, whereas in spatial STM groups are also coded for their position in a sequence, but items are coded for their position in a sequence, rather than in a group. Findings support the notion that positional information in verbal and spatial STM is represented by modality-specific mechanisms.

Keywords: positional, serial order, short-term memory, spatial, verbal

Functional similarities and differences between the coding of positional information in verbal and spatial short-term order memory

—

The ability to maintain the sequential order of perceived events in short-term memory (STM) is crucial for various acts of higher level cognition. Accordingly, a central theme in research on STM has been to identify the nature of the mechanisms responsible for maintaining items in their serial order. An additional theme—influenced by the theoretical fractionation of verbal and spatial STM in multi-component models, such as the working memory model (Baddeley, 1986, 2000, 2007; Baddeley & Hitch, 1974)—has been to establish whether the mechanisms of serial order are the same in the verbal and spatial domains. Evidence of common ordering mechanisms has recently emerged from detailed analyses of the response times accompanying transposition errors in verbal and spatial serial recall (Farrell & Lewandowsky, 2004; Lewandowsky & Farrell, 2008; Hurlstone & Hitch, 2015a), which support the notion that serial order in verbal and spatial STM is coded via a representational mechanism embodying three principles, namely position marking (items are associated with some representation of their position in a sequence), a primacy gradient (items are encoded with gradually decreasing strength), and response suppression (items are suppressed or removed from memory once they have been retrieved). Additional evidence for the role of position marking in verbal and spatial STM has been provided by the observation of temporal grouping effects (see following)—a key source of evidence for the role of positional representations in STM—in the immediate serial recall of verbal (Henson, 1999; Ng & Maybery, 2002; Ryan, 1969a) and spatial sequences (Hurlstone & Hitch, 2015a; Parmentier, Andrés, Elford, & Jones, 2006; Parmentier, Maybery, & Jones, 2004).

Although the primacy gradient and response suppression components appear to be implemented by modality-specific mechanisms—e.g., the phonological loop and visuospatial sketchpad in the working memory model—it is currently unclear whether the position marking component is also mediated by modality-specific mechanisms, or instead by a domain-general system—e.g., the episodic buffer in the working memory model (Baddeley, 2000; Burgess &

Hitch, 2005)—that is shared between the verbal and spatial, and perhaps other, domains (Hurlstone & Hitch, 2015a). Support for the former view is provided by evidence of functional differences in the effects of temporal grouping on error patterns in verbal and spatial serial recall, which suggest that the representation of positional information in grouped sequences is subtly different in the two domains. The present article uses a combination of experimentation and computational modeling to elucidate the nature of this difference in positional representations.

Temporal grouping effects in serial recall

That serial order in verbal STM incorporates positional information is supported by the results of studies employing a temporal grouping manipulation. This manipulation involves contrasting immediate serial recall under two conditions: In the ungrouped condition, the to-be-remembered sequences are conveyed regularly, with a constant inter-item temporal interval, whereas in the grouped condition the to-be-remembered sequences are organized into sub-groups by inserting extended temporal pauses, typically after every few items. Temporal grouping has been shown to exert a number of systematic and widely replicated effects on the serial recall of verbal sequences. First, grouping enhances recall accuracy (Frankish, 1985, 1989; Henson, 1996, 1999; Hitch, Burgess, Towse, & Culpin, 1996; Maybery, Parmentier, & Jones, 2002; Ng & Maybery, 2005; Ryan, 1969a, 1969b) and modifies the shape of the accuracy serial position curve: Whereas the accuracy curve for ungrouped sequences exhibits effects of primacy and recency for the sequence overall, the accuracy curve for grouped sequences also exhibits mini primacy and recency effects within each group (Frankish, 1985, 1989; Hitch et al., 1996). Second, grouping changes the shape of the latency serial position curve (Anderson et al., 1998; Anderson & Matessa, 1997; Farrell, 2008; Farrell & Lewandowsky, 2004; Farrell, Wise, & Lelièvre, 2011; Maybery et al., 2002; Parmentier & Maybery, 2008): Whereas the latency curve for ungrouped sequences peaks at the first position—indicating that people leave a long pause before outputting the sequence—the latency curve for grouped sequences also exhibits smaller peaks at positions corresponding to the start of groups—indicating that people also leave a brief pause before outputting each group. Third,

and most importantly, grouping modifies the pattern of errors by reducing the number of transpositions overall, and between groups in particular. However, critically, one type of between-group transposition actually increases in grouped sequences: These *interpositions* (Henson, 1996) are transpositions between groups that preserve their within-group position (Henson, 1996, 1999; Ng & Maybery, 2002, 2005; Ryan, 1969a). For example, if a 9-item sequence is organized into three groups of three, interpositions are indicated by an increase in the probability of three- and six-apart transpositions.

Several computational models of serial order in STM have been developed in recent years (Botvinick & Plaut, 2006; Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1999; Farrell & Lewandowsky, 2002; Grossberg & Pearson, 2008; Henson, 1998; Lewandowsky & Farrell, 2008; Page & Norris, 1998). Temporal grouping effects have been interpreted in terms of a sub-class of these models, known as positional theories (Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1999; Farrell, 2012; Henson, 1998; Lewandowsky & Farrell, 2008). According to such theories, serial order is coded by associating items with some representation of their ordinal (Lewandowsky & Farrell, 2008), temporal (Brown et al., 2000; Burgess & Hitch, 1999), or relative (Henson, 1999) within-sequence position (viz. position marking). Serial recall is accomplished by reinstating these positional representations, which cue item retrieval. To model temporal grouping effects, it is assumed that positional information in grouped sequences is represented on two dimensions, with one dimension representing the positions of items in groups, and with the second dimension representing the positions of groups (Brown et al., 2000; Farrell, 2012; Henson, 1998; Lewandowsky & Farrell, 2008) or items (Burgess & Hitch, 1999) in the sequence overall. This multidimensional representation of serial order has been shown to be sufficient to account for the effects of temporal grouping just reviewed, including the occurrence of between-group interposition errors, which manifest because items in the same positions in different groups share overlapping within group positional codes, rendering them vulnerable to positional confusion. Accordingly, temporal grouping effects have been taken to confer evidence that serial order in verbal STM incorporates positional information.

Temporal grouping effects are not confined to verbal memoranda. They have also been documented employing nonverbal sequences containing visually (Hurlstone & Hitch, 2015a;

Parmentier et al., 2006) or auditorily (Parmentier, Maybery, & Jones, 2004) presented spatial locations. In each of these studies, grouping was shown to enhance recall performance; produce mini within-group primacy and recency effects; and cause longer recall latencies at group boundaries. That temporal grouping effects are witnessed with spatial material suggests that positional information also contributes to the representation of serial order in spatial STM. However, in a noteworthy departure from the results obtained with verbal material, in neither of these studies did grouping foster an increase in interpositions.

The absence of interpositions in grouped spatial sequences is puzzling given that, in all other respects, the effects of grouping were identical to those witnessed with verbal material. What is the source of this discrepancy between the effects of grouping in the verbal and spatial domains? The hypothesis to be explored here is that it reflects a fundamental difference in the way positional information is represented in grouped verbal and spatial sequences. Specifically, we speculate that in both domains, one dimension of order incorporates information about the positions of groups in the sequence. However, whereas in the verbal domain, the second dimension of order incorporates information about the positions of items in groups—a necessary prerequisite to account for interpositions—we propose that in the spatial domain this second dimension represents information about the positions of items in the sequence overall.

Present study

The aims of the current study were two-fold. The first aim was to replicate the previously documented similarities and differences between the effects of grouping in the verbal and spatial domains. Replication is required because although the effects of grouping in the verbal domain are highly robust—having been documented across numerous studies—the effects of grouping in the spatial domain rest on a more limited number of observations, which raises questions about their generality. Moreover, a direct comparison of grouping in the verbal and spatial domains is currently lacking, leaving open the possibility that the difference between the effects of grouping in the two domains in terms of interpositions may simply be attributable to variations in the study and recall protocols used in previous studies of grouped verbal and

spatial serial recall. This aim was addressed by performing a new experiment that directly compared grouped verbal and spatial serial recall using the same methodology. To foreshadow, the results of the experiment revealed an empirical pattern that was consistent with previous studies, thus ruling out the possibility that the difference between the effects of grouping in the two domains is attributable to methodological discrepancies between earlier studies.

Having verified the robustness of the previously documented data patterns, the second aim was to establish the nature of the difference in positional representations in the two domains. This aim was addressed by applying two positional models of serial recall to the data from the experiment and several previous grouped serial recall experiments. The two models were functionally identical except for the putative representational differences hypothesized previously. The results of the modeling suggest that the representation of serial order in verbal STM incorporates information about the position of groups in a sequence and the position of items in a group, whereas the representation of serial order in spatial STM incorporates information about the position of groups in a sequence and the position of items in a sequence overall. The results support the notion that positional information in verbal and spatial STM is represented by modality-specific mechanisms rather than a domain-general system. Implications for theoretical accounts of working memory are considered.

Experiment

The experiment directly compared ungrouped and temporally grouped verbal and spatial serial recall in a mixed design, where the methodology was closely equated in the verbal and spatial tasks. Notably, the temporal presentation schedules for ungrouped and grouped sequences were identical in the two tasks and both employed a serial reconstruction recall procedure. Serial reconstruction is a variant of serial recall in which the to-be-remembered items are re-presented during the recall phase and participants must select the items in their original presentation order.

Method

Participants & apparatus. Thirty-six members of the campus community at the University of York took part in the experiment in exchange for course credit (in the case of

psychology students) or an honorarium of £3 (approximately \$4.50). The experiment was executed on a PC that presented all stimuli and collected and scored all responses.

Stimuli. The stimuli for the verbal task were sequences containing random orderings of the letters *F, H, J, L, N, Q, R, S, Y*. Each letter was presented visually in the central screen position in black point 18 Arial upper case font on a white background. The stimuli for the spatial task were sequences containing random orderings of nine visually presented spatial locations. The locations were nine gray squares (measuring 5cm × 5cm each) arranged haphazardly on a white background (Figure 1). The minimum and maximum distances between pairs of locations (measured from the center of each square) were 9 cm and 35 cm, respectively.

Design. The experiment employed a 2 (task: verbal vs. spatial) × 2 (grouping: ungrouped vs. grouped) × 9 (serial position: 1–9) mixed design: task was a between-participants factor, whereas grouping and serial position were within-participants factors. Half of the participants undertook the verbal task, whereas the other half undertook the spatial task. There were two blocks of 20 trials—one with ungrouped sequences and one with grouped sequences—which were each preceded by two practice trials. The ungrouped block was always administered first, since it is known that following experience with the serial recall task or exposure to a grouping strategy, participants will often subjectively group ostensibly ungrouped sequences (Farrell & Lelièvre, 2009; Henson, 1996; Madigan, 1980). Although this creates an order artifact, it was deemed necessary to ensure that the ungrouped condition provided a suitable baseline for examining the effects of objective grouping.

Procedure. Participants were tested individually in a quiet room in the presence of the experimenter. They initiated each trial by clicking on a “begin trial” button located in the central screen position using a mouse-driven pointer. A 2000-ms delay then ensued before presentation of the sequence during which a central fixation cross was displayed for the verbal task, and all locations were simultaneously visible for the spatial task. For ungrouped verbal sequences, each letter was presented singly for 500 ms, separated by a 250-ms blank inter-stimulus interval. For ungrouped spatial sequences, each location was highlighted yellow for 500 ms, followed by a 250-ms inter-stimulus interval during which all locations were gray. For grouped verbal and spatial sequences, the inter-stimulus intervals following the third and

sixth items were increased to 1000 ms to create the impression of three groups of three items.

Following the final item, there was a 1000-ms blank interval after which in the verbal task the letters were simultaneously presented in random positions within a 3×3 matrix and in the spatial task the locations reappeared in their previous spatial coordinates. Participants were required to click on the letters or locations in their presentation order using the mouse-driven pointer. Once an item was selected, its color changed temporarily to green for 50 ms to acknowledge that the computer had registered the response. Items could be selected on multiple occasions, meaning that repetition errors were possible. Once nine items had been selected, there was a 3000-ms inter-trial interval, which was followed by presentation of the “begin trial” button for the next trial. The entire experiment lasted approximately 30 minutes.

Results

The data were analyzed using a strict serial recall scoring procedure: An item was only scored as correct if its output serial position was the same as its input serial position. Although the analysis of transposition errors is of central interest, we begin by examining the accuracy and latency serial position curves to determine whether the grouping manipulation exerted the usual effects on these measures.

Accuracy serial position curves. Figure 2 shows the accuracy serial position curves for the verbal (panel one) and spatial (panel four) task. A 2 (task: verbal vs. spatial) \times 2 (grouping: ungrouped vs. grouped) \times 9 (serial position: 1–9) ANOVA revealed a significant main effect of grouping, $F(1, 34) = 13.09$, $p < .001$, with better recall of grouped than ungrouped sequences, and a significant main effect of serial position, $F(1, 34) = 13.09$, $p < .001$, but the main effect of task fell marginally short of significance, $F(1, 34) = 3.06$, $p = .09$. There was also a significant task \times serial position interaction, $F(8, 272) = 5.1$, $p < .001$, which arose because recall was markedly better in the verbal than the spatial task for all but the last three serial positions, and a significant grouping \times serial position interaction, $F(8, 272) = 3.07$, $p < .01$, which arose because grouping caused mini primacy and recency effects in each group.

Latency serial position curves. Panels two and five of Figure 2 show the mean recall latencies for correct responses as a function of serial position for the verbal and spatial task,

respectively. A 2 (task) \times 2 (grouping) \times 9 (serial position) ANOVA revealed a significant main effect of task, $F(1, 34) = 15.28$, $p < .01$, with longer recall latencies in the verbal than in the spatial task, no significant main effect of grouping, $F(1, 34) = 1.94$, $p = .17$, and a significant main effect of serial position, $F(8, 272) = 18.14$, $p < .01$, which arose because of a high amplitude peak at Serial Position 1 and smaller amplitude peaks at Serial Positions 4 and 7. There was also a significant task \times serial position two-way interaction, $F(8, 272) = 6.23$, $p < .01$, and a significant grouping \times serial position two-way interaction, $F(8, 272) = 6.84$, $p < .01$, which were both subsumed under a significant task \times grouping \times serial position three-way interaction, $F(8, 272) = 2.11$, $p < .05$. The three-way interaction arose because the recall times were longer for verbal than spatial sequences and grouping modified the shape of the latency serial position curve in the two tasks: The latency curves for ungrouped sequences peak at Serial Position 1, with the recall times over subsequent positions following an inverted U shaped trend; by contrast, the latency curves for grouped sequences exhibit peaks at positions straddling group boundaries (viz. Serial Positions 1, 4, and 7).

Transposition errors. Turning to the data of chief interest, panels three and six of Figure 2 show the transposition gradients, which plot the proportion of transpositions as a function of transposition distance for the verbal and spatial task, respectively. It can be seen from inspection of panel three of Figure 2 that for the verbal task, grouping increased the incidence of interpositions: The transposition gradient for ungrouped sequences decreases monotonically with increasing transposition distance, whereas the transposition gradient for grouped sequences is non-monotonic, exhibiting peaks for three- and six-apart transpositions, which necessarily represent interpositions. By contrast, it is apparent from inspection of panel six of Figure 2 that for the spatial task, grouping did not affect the incidence of interpositions: The transposition gradients for ungrouped and grouped sequences both decrease monotonically with transposition distance.

To scrutinize the error patterns more closely, transpositions were categorized as occurring within or between groups, with the latter errors being further sub-divided into interpositions and other between-group errors. The incidence of the different error-types in ungrouped and grouped sequences was compared separately for the verbal and spatial task via

paired comparisons performed on the log-odds transformed error proportions. For the verbal task, grouping did not affect the incidence of transpositions within groups (ungrouped $M = .43$; grouped $M = .45$), $t(17) = .51$, $p = .62$, but decreased the incidence of other between group errors (ungrouped $M = .4$; grouped $M = .27$), $t(17) = 5.58$, $p < .001$, and increased the incidence of interpositions (ungrouped $M = .17$; grouped $M = .29$), $t(17) = 5.093$, $p < .001$. For the spatial task, grouping increased the incidence of transpositions within groups (ungrouped $M = .55$; grouped $M = .71$), $t(17) = 6.23$, $p < .001$, and decreased the incidence of other transpositions between groups (ungrouped $M = .32$; grouped $M = .18$), $t(17) = 6.68$, $p < .001$. Critically, grouping did not increase the incidence of interpositions (ungrouped $M = .13$; grouped $M = .11$), $t(17) = 12.57$, $p = .52$.

Summary of results and simulations

The results of the experiment are straightforward and can be summarized as follows. Temporal grouping exerted a number of kindred effects on the recall of verbal and spatial sequences including an elevation in recall accuracy; mini primacy and recency effects; long output times prior to the production of the first item in each group; and a reduction in between group transpositions. Critically, however, although grouping increased the incidence of interpositions in the verbal task, it did not influence the incidence of these errors in the spatial task. These results suggest that it is unlikely that the failure to detect an increase in interpositions in grouped spatial sequences in previous studies (Hurlstone & Hitch, 2015a; Parmentier et al., 2006) is due to methodological differences between those studies and verbal studies of temporal grouping. This leaves open the possibility that this fundamental difference between the effects of grouping in the two domains might be attributable to variations in the way positional information is coded in verbal and spatial STM. We now report simulations of two positional models of serial recall that were designed to test this possibility.

Modeling of temporal grouping effects

To test our hypothesis that positional information is represented differently in the verbal and spatial domains, we contrasted the predictions of two versions of a positional model of

serial recall. There are several positional models that we could have used for this purpose (e.g., Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998; Lewandowsky & Farrell, 2008). Of these, we chose the start-end model (SEM) of Henson (1998), since it is a straightforward and representative member of the broader class of positional theories. As well as implementing position marking, the SEM also incorporates a primacy gradient and response suppression in common with most other models of serial recall.

In the SEM, a sequence is stored in STM as a set of unordered episodic tokens. These tokens contain information about the identity of the items conveyed and their position within the sequence. The positional information is represented by the joint action of a start marker whose strength is maximal upon presentation of the first item and then decreases exponentially across positions, and an end marker whose strength is weakest for the first item and then increases exponentially across positions. Together, the start and end markers provide an approximate two-dimensional representation of the position of each item relative to the start and end of a sequence (hence the model's namesake). When a sequence is grouped, the start and end markers vary over the positions of items within groups, and an additional set of start and end markers is recruited that encode the position of each group relative to the start and end of the sequence. In what follows, we contrasted this standard implementation of the SEM combining position-of-group and position-in-group representations of order (GP + PG model), with a representational scheme in which position was coded by combining position-of-group and position-in-sequence representations of order (GP + PS model). The two models were applied to the grouped data from the experiment and four previous verbal and spatial serial recall experiments.

Implementation

Encoding. In the GP + PG model, a sequence was encoded by associating items with a two-dimensional vector $\mathbf{p} = \{s_I, e_I\}$ coding the relative position of each item in a group and a second two-dimensional vector $\mathbf{g} = \{s_G, e_G\}$ coding the relative position of each group in the sequence. The values of the start and end markers coding the position of each item i within a group were determined as follows:

$$s_I(i) = S_I^{i-1}, \quad e_I(i) = E_I^{N_{I(g)}-i}, \quad (1)$$

Where S_I and E_I are free parameters ($0 < S_I \leq 1$; $0 < E_I \leq 1$) controlling the rate of change of the start and end item markers, respectively, and N_I is the number of items in group g . Similarly, the values of the start and end markers coding the position of groups in the sequence were given by:

$$s_G(g) = S_G^{g-1}, \quad e_G(g) = E_G^{N_{G-g}}, \quad (2)$$

Where S_G and E_G are free parameters ($0 < S_G \leq 1$; $0 < E_G \leq 1$) controlling the rate of change of the start and end group markers, respectively, and N_G is the number of groups in the sequence (for ungrouped sequences $N_G = 1$). In the GP + PS model, the encoding of a sequence proceeded as described above, except that the values of the item markers specified by equation 1 varied over the positions of items in the sequence overall ($i = 1 \dots 9$), rather than within each group ($i = 1 \dots 3$), and were thus insensitive to group information. The item markers for this model were therefore constructed as if the sequence was an ungrouped sequence.

Retrieval. Recalling a sequence involved reinstating the position markers for each position to probe for a response. For each position, the overlap between the position marker being used to cue recall and the positional information stored in each of the episodic tokens was computed in parallel and used to determine the strength with which items competed for recall in response to the cue. For the item markers, the overlap between the item marker for the current position j , $\mathbf{p}(j)$ and item l , $\mathbf{p}(l)$ was determined by:

$$o_I(j, l) = \sqrt{\mathbf{p}(j) \cdot \mathbf{p}(l)} \exp \left\{ -\sqrt{\sum_k [p_k(j) - p_k(l)]^2} \right\}, \quad (3)$$

Where \cdot is the inner product between vectors and k indexes the two elements of vector \mathbf{p} . Similarly, the overlap between the group marker for the current position $\mathbf{g}(j)$ and group $\mathbf{g}(l)$ was given by:

$$o_G(j, l) = \sqrt{\mathbf{g}(j) \cdot \mathbf{g}(l)} \exp \left\{ -\sqrt{\sum_k [g_k(j) - g_k(l)]^2} \right\}, \quad (4)$$

Finally, the overall strength with which each item competed for selection in response to the item and group position markers for a given position was determined by:

$$c(j, l) = o_I(j, l)o_G(j, l)[1 - r(l)] + N(0, G_c), \quad (5)$$

Where $r(l)$ represents the suppression of item l once recalled—which was set to a constant value of .95—and N represents random noise drawn from a Gaussian distribution with a mean of 0 and standard deviation given by the free parameter G_c ($0 < G_c < 1$).¹

Quantitative fits of models to present experiment

The models were fit to the accuracy serial position curves and transposition gradients of individual participants for the grouped verbal and spatial sequence conditions of the experiment using maximum likelihood parameter estimation (assuming normally distributed data with constant variance). The parameters of each model were varied systematically to find the combination of parameter values that maximized the log-likelihood:

$$\ln L = \frac{-n}{2} \ln \left(\frac{RSS}{n} \right) \quad (6)$$

Where \ln is the natural logarithm, RSS is the Residual Sum of Squares, and n is the number of observations entering into the log-likelihood calculation (17 points in total; 9 for accuracy serial position curve and 8 for transposition gradient). The log-likelihood was converted to a negative value and minimized using the SIMPLEX algorithm of Nelder and Mead (1965). Each parameter vector explored by the search algorithm involved 1000 model

¹We do not simulate the dynamics of recall since like most other models of serial recall, the SEM does not incorporate a mechanism for generating response time predictions. Although such a mechanism can easily be introduced by augmenting the recall process in the SEM with a set of competitive decision accumulators (e.g., Usher & McClelland, 2001), additional assumptions about the hierarchical nature of retrieval of positional information would also need to be incorporated to account for the long recall times at the beginning of the sequence and at group boundaries (Farrell, 2012). In the interest of parsimony, we chose not to model these phenomenon, since the latency data are only reported here to verify that grouping exerted its usual effects on the dynamics of recall.

simulation trials of 9-item sequences grouped into threes. To increase the likelihood of finding the global minimum of the goodness-of-fit function, the search process was conducted multiple times using 16 different starting parameter combinations.

Table 1 contains the parameter estimates for the fits of the models to the grouped conditions of the present and previous experiments (see next section), whereas Table 2 contains the corresponding log-likelihood estimates ($\ln L$). Note that since both models contain the same number of free parameters, model comparison can proceed based on the raw log-likelihoods without adjusting for the degrees of freedom of the models (e.g., using AIC or BIC, see e.g., Lewandowsky & Farrell, 2011). Table 2 also contains the pairwise differences in log-likelihood between each model and the best fitting model ($\Delta \ln L$), and the percentage of participants for which each model provided the best fit (Σ).

Considering first the fits of the models to the verbal data, it is apparent from inspection of Table 2 that the average log-likelihood estimate was larger for the GP + PG model than for the GP + PS model, with the former model offering the best fit for 11 out of 18 participants (61%). The reason for this better fit can be seen in panels one and three of Figure 3, which show, respectively, the serial position curves and transposition gradients predicted by the models (note that for visual comparison, Figure 3 also contains fits of the SEM to the ungrouped conditions of the experiment for which the GP + PG and GP + PS models are formally identical). It is apparent from inspection of the figure that the GP + PG model does an adequate job of reproducing the within-group primacy and recency of the serial position curve and the non-monotonicity of the transposition gradient (although the model over-predicts the magnitude of three-apart transpositions and under-predicts the magnitude of six-apart transpositions). By contrast, the GP + PS model predicted a serial position curve and transposition gradient that bear a greater resemblance to the data for ungrouped, than for grouped, sequences—the serial curve and transposition gradient are not scalloped in appearance, but instead decrease monotonically with serial position and transposition distance, respectively.

Turning to the fits of the models to the spatial data, Table 2 shows that the average log-likelihood estimate was larger for the GP + PS model than for the GP + PG model, with the

former model offering the best fit for 14 out of 18 participants (78%). Nevertheless, Figure 3 shows that both models do an adequate and comparable job of accounting for the serial position curve (panel two) and transposition gradient (panel four). Notably, both predict the observed within-group primacy and recency of the accuracy serial curve and the negative monotonicity of the transposition gradient. It merits comment, however, that whilst the aggregate transposition gradient predicted by the GP + PG model does not contain a peak in three- or six-apart transpositions, inspection of the gradients generated for individual participants revealed that the model predicted a peak in three-apart transpositions for 6 out of 18 participants (33%), when such a peak was not witnessed in the empirical data of those (or any) participants.

Quantitative fits of models to previous experiments

To help adjudicate between the models, we next apply them to data from the grouped conditions of four previous representative serial recall experiments. The details of those experiments are provided in Table 3. They included experiments by Ryan (1969; Experiment 1) and Henson (1996; Experiment 2), which provide representative results for grouped verbal serial recall, and experiments by Hurlstone and Hitch (2015a; Experiment 3) and Parmentier et al. (2006; Experiment 4), which provide representative results for grouped spatial serial recall. The experiments all involved serial recall of 9-item sequences grouped in a 3-3-3 pattern. The model fitting procedure was the same as described previously. For the experiment of Hurlstone and Hitch (2015a), the models were fit to the data of individual participants, whereas for the other experiments the models were applied to the aggregate results, since individual participant data were not available.

Figure 4 shows the serial position curves and transposition gradients for the grouping conditions of the previous serial recall experiments (panels one and four, respectively) and the corresponding fits to those data of the GP + PG model (panels two and five, respectively) and the GP + PS model (panels three and six, respectively). Commencing with the fits of the models to the verbal data, Table 2 shows that for the data of Ryan (1969) and Henson (1996), the log-likelihood estimates were markedly larger for the GP + PG model than for the GP + PS

model. For both data sets, the GP + PG model accurately reproduced the within-group primacy and recency of the accuracy serial position curves, the decrease in transpositions with transposition distance, and the upturn in three-apart transpositions (Figure 4, panels two and five). The GP + PS model, by contrast, was unable to reproduce the scalloped appearance of the accuracy serial position curves for either data set, nor the peak in three-apart transpositions in the Henson data (Figure 4, panels three and six). Surprisingly, however, the model did predict the increase in three-apart transpositions for the data of Ryan, indicating that the GP + PS model can produce interpositions under some circumstances, despite not incorporating any explicit information about the positions of items within groups. A closer inspection of the model predictions indicated that these interpositions were the result of whole groups exchanging position in the sequence, and were attributable to the low parameter settings of the start and end group markers ($S_G = 0.05$; $E_G = 0.20$), which rendered the representations of groups highly confusable and prone to error.

Turning now to the fits of the models to the previous spatial serial recall experiments, Table 2 shows that for the data of Hurlstone and Hitch (2015a), the average log-likelihood estimate for the GP + PS model was once again larger than for the GP + PG model, with the former model providing the best fit for 21 out of 26 participants (81%). For the data of Parmentier et al. (2006), the log-likelihood estimate for the GP + PS model was also larger than for the GP + PG model, although only marginally so. For both data sets, the better fit of the former model arose because it did a slightly better job of reproducing the last few positions of the serial position curve. However, the transposition gradients predicted by the models were virtually indistinguishable, with both models reproducing the monotonic decrease in the frequency of transpositions with transposition distance. It merits comment, however, that for the data of Hurlstone and Hitch, the GP + PG model generated transposition gradients with identifiable peaks for three-apart transpositions for 3 out of 26 participants (12%) when no such peaks were observed for the transposition gradients of those (or any) participants.

In brief, for grouped verbal serial recall, the GP + PG model provided a better fit to both of the previous data sets, critically capturing the increase in interpositions, whereas the GP + PS model was only able to generate interpositions for the Ryan data set and these arose because of

whole groups exchanging positions, which is at variance with the empirical data showing that interpositions occur in isolation (see Henson, 1996). By contrast, for grouped spatial serial recall, the GP + PS model provided the better fit to the two previous data sets, although the qualitative predictions of the two models were nevertheless very similar, and notably the GP + PG model was able to reproduce the monotonic decrease in the frequency of transpositions with transposition distance observed empirically. Thus, it appears that the GP + PG model is sufficiently flexible that it can accommodate both the presence of interpositions in the verbal data, and their absence in the spatial data. We explore the issue of model flexibility in the next section.

Robustness of predictions

To probe model behavior more deeply, we next subjected the models to a parameter sensitivity analysis in order to determine the robustness of their predictions regarding interpositions. For each model, we varied the values of the start and end markers for items S_I and E_I , and groups S_G and E_G from .05 to .95 in steps of 0.1 and factorially combined these values to create a grid containing 10,000 (10^4) parameter setting combinations to-be-explored by simulation (the degree of response suppression r and the amount of Gaussian perturbation G_c were fixed to constant values of .95 and .15, respectively). We conducted two sets of simulations: One with ungrouped sequences and one with grouped sequences. For each parameter setting combination, model predictions were generated for 1000 simulation trials of ungrouped and 3-3-3 grouped 9-item sequences. The dependent measure of interest was the change in the frequency of interpositions in grouped, relative to ungrouped, sequences. This was calculated by subtracting the proportion of interpositions predicted for ungrouped sequences for a particular parameter setting combination from the proportion predicted for grouped sequences under the same parameter setting combination.

Figure 5 summarizes the results of this analysis for the GP + PG model (left panel) and the GP + PS model (right panel), which are displayed as density histograms. To facilitate interpretation, the solid vertical line in each panel represents a situation where the frequency of interpositions is the same in grouped and ungrouped sequences. Bins to the left of this

criterion line represent observations where the incidence of interpositions was smaller in grouped than in ungrouped sequences; conversely, bins to the right of the criterion line contain observations where the incidence of interpositions was greater in grouped than ungrouped sequences. The broken vertical lines in each panel represent the values obtained in six conditions, namely the verbal and spatial conditions of the present Experiment, the verbal data of Ryan (1969) and Henson (1996), and the spatial data of Hurlstone and Hitch (2015a) and Parmentier et al. (2006). It is apparent from inspection of the figure that the observed values for the verbal data fall to the right of the criterion line, indicating that grouping engendered an increase in interpositions, whereas the spatial data fall just to the left of the criterion line, indicating that grouping engendered a very slight reduction in interpositions.

It is also apparent from inspection of the figure that the two models generate very different distributions. Most of the distribution of the GP + PG model lies to the right of the criterion line, indicating that the model predicts an increase in interpositions in grouped sequences as its main theoretical prediction. However, there is also a small portion of the model's parameter space in which it predicts a slight decrease in interpositions. Indeed, the distribution of the GP + PG model is sufficiently wide that it straddles the observed values for both the verbal and the spatial data. By contrast, most of the distribution of the GP + PS model lies to the left of the criterion line, indicating that this model predicts a decrease in interpositions in grouped sequences as its main theoretical prediction, although there is a smaller portion of its parameter space within which it predicts a very slight increase in interpositions. The distribution for this model is much narrower and straddles the observed values for the spatial data, but falls outside the range of observed values for the verbal data.²

In brief, the results of the parameter sensitivity analysis confirm that, despite sharing the

²We also examined the prevalence of interpositions in grouped sequences by comparing the frequency of three-apart interpositions with the frequency of two-apart transpositions. This analysis revealed that the proportion of interpositions was greater for 94% of the parameter setting combinations simulated for the GP + PG model, compared to only 2% for the GP + PS model. This analysis confirms that interpositions are a representative feature of the GP + PG model, but not the GP + PS model, which only very rarely generates these errors.

same number of model parameters, the functional form of the GP + PG model—viz. how those parameters are combined in model equations 1 and 2—renders it more flexible than the GP + PS model. This conclusion is buttressed by an additional set of analyses reported in the Appendix that employed a complementary model selection procedure known as landscaping (Navarro, Pitt, & Myung, 2004), which provides an assessment of how well two models can fit each others data. Those analyses indicate that the GP + PG model mimics the behavior of the GP + PS model better than the latter model is able to mimic the behavior of the former.

General discussion

We began by reporting an experiment that directly compared temporal grouping effects in the verbal and spatial domains. The results revealed that grouping exerted several kindred effects on the recall of verbal and spatial sequences including an improvement in recall accuracy; mini within-group primacy and recency effects; elevated response times at group boundaries; and a decrease in between-group transpositions. However, whereas grouping a verbal sequence engendered an increase in interpositions—consistent with previous studies of grouping in the verbal domain (Henson, 1996, 1999; Ng & Maybery, 2002, 2005; Ryan, 1969a)—grouping a spatial sequence did not engender an increase in these positional errors—consistent with recent studies of grouping in the spatial domain (Hurlstone & Hitch, 2015a; Parmentier et al., 2006). This discrepancy cannot be explained by recourse to differences in the presentation or recall procedures, since these were equated in the verbal and spatial tasks. Instead, the results of the modeling of the present and previous experiments suggest that positional information is coded differently in the two domains—in the verbal domain, groups are coded for their position in a sequence and items are coded for their position in a group; whereas in the spatial domain, groups are also coded for their position in a sequence, but items are coded for their position in a sequence, rather than in a group.

Model selection

The conclusion that the results are best understood by recourse to differences in the way positional information is coded in the verbal and spatial domains rests on the observation that

a model combining position-of-group and position-in-sequence information consistently provided a better fit to the spatial data than a model combining position-of-group and position-in-group information. The former model also consistently predicted the absence of an increase in interpositions in grouped sequences across variations of its parameter settings, whereas the latter model consistently predicted an increase in these errors. We note also that similar findings were reported in model comparisons reported by Hurlstone and Hitch (2015b) using a lateral inhibition framework that permitted the generation of response probability and recall latency predictions, showing that the results presented here are in no way tied to the SEM framework used for the modeling. However, since the model integrating position-of-group and position-in-group information was able to reproduce the qualitative pattern of results in both domains (whereas the model integrating position-of-group and position-in-sequence information was not)—including critically, the absence of interpositions in grouped spatial sequences—one might conjecture that this model should nevertheless be preferred on grounds of parsimony, since it enables the data from both domains to be explained in terms of a common model and set of representational assumptions.

We caution against this interpretation, however, since given the results of our sensitivity analysis it goes against the grain of a fundamental principle that lies at the core of model selection—namely that models should be chosen that are sufficiently flexible that they can explain the range of data patterns witnessed empirically, but not so flexible that they predict data patterns that do not resemble human behavior (Pitt, Kim, Navarro, & Myung, 2006; Pitt, Myung, Montenegro, & Pooley, 2009; Pitt, Myung, & Zhang, 2002). Put another way, a desirable model of the cognitive process it was designed to explain should predict the empirical pattern typically observed across a wide range of its parameter settings. This shows that the empirical pattern is representative of the model's behavior and thus follows from its core representational assumptions. The model integrating position-of-group and position-in-group information meets this criterion when examined with reference to the verbal data, since it predicts an elevation in interpositions in grouped sequences across most of its parameter space, which is representative of the empirical pattern witnessed in verbal serial recall. However, the model does not meet this criterion when examined against the spatial data, because this data pattern

has not been observed in any of the four experiments that have examined grouping effects in spatial serial recall (the present experiment; Experiments 1 and 3 of Hurlstone & Hitch, 2015a; and Experiment 4 of Parmentier et al., 2006). The models core prediction is thus unrepresentative of the typical empirical pattern observed in this context. By contrast, the model integrating position-of-group and position-in-sequence information predicts this empirical pattern almost universally.

Theoretical implications

We next consider the theoretical implications of our results. To frame our discussion, we begin with a reprisal of the conclusions drawn in a previous study by Hurlstone and Hitch (2015a). In that study, we presented evidence for the use of positional information in spatial STM based on temporal grouping effects in spatial serial recall. However, we also provided further evidence for the use of position marking based on an empirical and modeling analysis of the latencies of transposition errors in spatial serial recall. This chronometric analysis also yielded evidence for two additional principles for representing serial order in spatial STM—viz. a primacy gradient and response suppression. These results replicated previous work by Farrell and Lewandowsky (Farrell & Lewandowsky, 2004; Lewandowsky & Farrell, 2008) who examined the dynamics of transposition errors in the serial recall of verbal sequences and also found empirical support for a representational mechanism embodying these three principles (see Hurlstone, Hitch, & Baddeley, 2014 for a review of additional sources of evidence for the operation of these three representational principles).

Hurlstone and Hitch (2015a) asked how these three principles for representing serial order map onto the different components of the working memory model of Baddeley and Hitch (Baddeley, 1986, 2000, 2007; Baddeley & Hitch, 1974). They presented arguments and evidence in favor of the view that the primacy gradient and response suppression are implemented in a modality-specific manner within the working memory slave systems; that is, the phonological loop and visuospatial sketchpad each possess their own mechanisms for generating a primacy gradient and implementing response suppression. However, in common with other serial recall theorists (Burgess & Hitch, 1999; Henson 1998; Smyth, Hay, Hitch, & Horton, 2005; Tremblay,

Guérard, Parmentier, Nicholls, & Jones, 2006), they noted the possibility that positional information in the verbal and nonverbal domains might be encoded via a common domain-general mechanism. For example, Burgess and Hitch (1999) postulated that the positional context signal in their network model of the phonological loop might also be responsible for coding the position of nonverbal items. In later work (Burgess & Hitch, 2005), the same authors postulated that the locus of their context signal within the working memory framework might be the episodic buffer (Baddeley, 2000)—a component of the working memory model responsible for integrating information from the working memory slave systems and long-term memory—and that the buffer might serve as a common positional coding mechanism for items maintained in the phonological loop and the visuospatial sketchpad.

The results of the present study invite a reappraisal of this shared system account of positional coding. Specifically, our results seem to undermine this view because the differences in error patterns in grouped verbal and spatial sequences appear to mandate distinct positional coding mechanisms, with subtly different representational characteristics. On this alternative, distinct system view, functionally similar but nevertheless distinct mechanisms would be responsible for representing the positions of items maintained in the phonological loop and visuospatial sketchpad. Combined with our earlier conclusions that the phonological loop and visuospatial sketchpad each possess their own mechanisms for generating a primacy gradient and suppressing recalled items (Hurlstone & Hitch, 2015a), this points to a working memory model in which serial order is represented in an entirely modality-specific manner, without the domain-general support of the episodic buffer (although see e.g., Meiser & Klauer, 1999 for some evidence that the central executive component might provide such support during the encoding of verbal and spatial sequences). The present study therefore advances our previous work by placing further constraints on the locus of the different principles for representing serial order within the framework of the working memory model.

Why is positional information in grouped verbal and spatial sequences represented differently? Why is verbal STM sensitive to the within-group position of items, when spatial STM is not? Our preferred answer to this question is that the coding of positional information

in verbal STM is parasitic upon a mechanism for serial ordering in language more generally, one whose operation is constrained by the linguistic properties of speech. Evidence for such a domain general (linguistic) ordering mechanism comes from the well established link between verbal STM and vocabulary acquisition (Baddeley, Gathercole, & Papagno, 1998), as well as similarities between error patterns in speech production and verbal serial recall (Page, Madge, Cumming, & Norris, 2007). One linguistic constraint that the ordering mechanism may exploit is syllable structure, which plays an important role in vocabulary acquisition. Children are remarkably adept, for example, at the task of nonword repetition (Gathercole, 2006)—which involves repeating back a novel sequence of phonemes—and it has been suggested that their success at accomplishing this task arises due to their ability to exploit knowledge of syllable structure, which places constraints on the possible orderings of phonemes that can occur (Hartley, 2002; Hartley & Houghton, 1996).

Although the presence of syllable structure aids the mental representation of speech, one adverse side effect is that it induces a particular pattern of speech errors, where phonemes from one syllable migrate to the same position in a different syllable (Ellis, 1980; Treiman & Danis, 1988). These errors are reminiscent of the interposition errors witnessed in grouped verbal sequences. Current models of phonological word form learning and speech production (Hartley, 2002; Hartley & Houghton, 1995; Vousden, Brown, & Harley, 2000) explain these errors by assuming that phonemes in a word become associated with a positional context signal, which includes a component that tracks within-syllable position. This is akin to the use of within-group positional codes in models of verbal serial recall (e.g. Brown et al., 2000; Burgess & Hitch, 1999) and it means that phonemes in the same position in different syllables become associated with similar states of the context signal. Thus, when an error occurs it will tend to involve the movement of one phoneme to the same position in another syllable.

Based on the foregoing considerations, it is possible that the mechanisms in models of STM, and language perception and production are one and the same thing, and that the component of the context signal that tracks within-syllable position in language perception also tracks within-group position in immediate serial recall. According to this account, the reliance of verbal STM on within-group positional codes is the result of an ordering mechanism

that has been shaped and constrained by the bottom-up properties of language. If one accepts this interpretation then the only reason to expect a role for kindred positional representations in the spatial domain would be if domain-specific experience had imposed similar constraints on the ordering mechanism in spatial STM. This seems unlikely, however, since there is no organisational unit in the nonverbal modality that is analogous to the syllable in language, and therefore no reason to expect that within-group positional codes should be utilised in the spatial domain.

Conclusions

We have shown that the nature of positional representations in STM produced by grouping a sequence of verbal or spatial items is different. In verbal STM, groups are coded for their position in a sequence and items are coded for their position in a group; whereas in spatial STM, groups are also coded for their position in a sequence, but items are coded for their position in a sequence, rather than in a group. This finding is incompatible with the hypothesis that positional information in verbal and spatial STM is represented via a domain-general positional coding mechanism, and instead supports the competing view that such information is represented via separate modality-specific mechanisms.

References

- Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, 38, 341-380. doi:10.1006/jmla.1997.2553.
- Anderson, J. R., & Matessa, M. P. (1997). A production system theory of serial memory. *Psychological Review*, 104, 728-748. doi:10.1037//0033-295X.104.4.728
- Baddeley, A. D. (1986). *Working Memory*. Oxford: Clarendon Press.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417-423. doi:10.1016/S1364-6613(00)01538-2
- Baddeley, A. D. (2007). *Working memory, thought and action*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198528012.001.0001

- Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105, 158-173. doi:10.1037//0033-295X.105.1.158
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47-89). New York: Academic Press.
- Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review*, 113, 201-233. doi:10.1037/0033-295X.113.2.201
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, 114, 539-576. doi:10.1037/0033-295X.114.3.539
- 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
- Burgess, N., & Hitch, G. (2005). Computational models of working memory: Putting long-term memory into context. *Trends in Cognitive Sciences*, 9, 535-541. doi:10.1016/j.tics.2005.09.011
- Ellis, A. W. (1980). Errors in speech and short-term memory: the effects of phonemic similarity and syllable position. *Journal of Verbal Learning and Verbal Behavior*, 19, 624-634. doi:10.1016/S0022-5371(80)90672-6
- Farrell, S. (2008). Multiple roles for time in short-term memory: Evidence from serial recall of order and timing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 128-145. doi:10.1037/0278-7393.34.1.128
- Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review*, 119, 223-271. doi:10.1037/a0027371
- Farrell, S., & Lelièvre, A. (2009). End anchoring in short-term order memory. *Journal of Memory and Language*, 60, 209-227. doi:10.1016/j.jml.2008.09.004
- Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review*, 9, 59-79. doi:10.3758/BF03196257
- Farrell, S., Wise, V., & Lelièvre, A. (2011). Relations between timing, position and grouping in short-term memory. *Memory Cognition*, 39, 573-587. doi:10.3758/s13421-010-0053-0

- Frankish, C. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language*, 24, 200-209. doi:10.1016/0749-596X(85)90024-5
- Frankish, C. (1989). Perceptual organization and precategorical acoustic storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 469-479. doi:10.1037//0278-7393.15.3.469
- Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, 27, 513-543. doi:10.1017.S0142716406060383
- Grossberg, S., & Pearson, L. R. (2008). Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: Towards a unified theory of how the cerebral cortex works. *Psychological Review*, 115, 677-732. doi:10.1037/a0012618
- Hartley, T. (2002). Syllabic Phase: A bottom-up representation of the structure of speech. In J. Bullinaria & W. Lowe (Eds.), *7th Neural Computation and Psychology Workshop*. Singapore: World Scientific,.
- Hartley, T., & Houghton, G. (1996). A linguistically restrained model of short-term memory for nonwords. *Journal of Memory and Language*, 35, 1-31. doi:10.1006/jmla.1996.0001
- Henson, R. N. A. (1996). *Short-term memory for serial order*. Unpublished doctoral thesis. Cambridge University, Cambridge, U.K.
- 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
- Henson, R. N. A. (1999). Positional information in short-term memory: Relative or absolute? *Memory & Cognition*, 27, 915-927. doi:10.3758/BF03198544
- Hitch, G. J., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal grouping effects in immediate recall: A working memory analysis. *Quarterly Journal of Experimental Psychology*, 49, 116-139. doi:10.1080/027249896392829
- Hurlstone, M. J., & Hitch, G. J. (2015a). 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/a0038223
- Hurlstone, M. J., & Hitch, G. J. (2015b). Supplementary material for: How is the serial order of a spatial sequence coded? Insights from transposition latencies. Unpublished manuscript

(downloadable from: <http://mark-hurlstone.github.io>).

- 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
- 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.
- Madigan, S. A. (1980). The serial position curve in immediate serial recall. *Bulletin of the Psychonomic Society*, 15, 335-338. doi:10.3758/BF03334550
- Maybery, M., Parmentier, F. B. R., & Jones, D. M. (2002). Grouping of list items reflected in the timing of recall: Implications for models of serial verbal memory. *Journal of Memory and Language*, 47, 360-385. doi:10.1016/S0749-596X(02)00014-1
- Navarro, D. J., Pitt, M. A., & Myung, J. I. (2004). Assessing the distinguishability of models and the informativeness of data. *Cognitive Psychology*, 49, 47-84. doi:10.1016/j.cogpsych.2003.11.001
- Nelder, J. A., Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7, 308-313. doi:10.1093/comjnl/7.4.308
- Ng, L. H., & Maybery, M. T. (2002). Grouping in verbal short-term memory: Is position coded temporally? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 55A, 391-424. doi:10.1080/02724980143000343
- Ng, L. H., & Maybery, M. T. (2005). Grouping in short-term memory: Do oscillators code the positions of items? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 175-181. doi:10.1037/0278-7393.31.1.175
- Page, M. P. A., Madge, A., Cumming, N., & Norris, D. G. (2007). Speech errors and the phonological similarity effect in short-term memory: Evidence suggesting a common locus. *Journal of Memory and Language*, 56, 49-64. doi:10.1016/j.jml.2006.09.002
- Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105, 761-781. doi:10.1037//0033-295X.105.4.761-781

- Parmentier, F. B. R., Andrés, P., Elford, G., & Jones, D. M. (2006). Organization of visuo-spatial serial memory: Interaction of temporal order with spatial and temporal grouping. *Psychological Research*, 70, 200-217. doi:10.1007/s00426-004-0212-7
- Parmentier, F. B. R., & Maybery, M. T. (2008). Equivalent effects of grouping by time, voice and location on response timing in verbal serial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1349-1355. doi:10.1037/a0013258
- Parmentier, F. B. R., Maybery, M., & Jones, D. M. (2004). Temporal grouping in auditory spatial serial memory. *Psychonomic Bulletin & Review*, 11, 501-507. doi:10.3758/BF03196602
- Pitt, M. A., Kim, W., Navarro, D. J., & Myung, J. I. (2006). Global model analysis by parameter space partitioning. *Psychological Review*, 113, 57-83. doi:10.1037/0033-295X.113.1.57
- Pitt, M. A., Myung, J. I., Montenegro, M., & Pooley, J. (2009). Measuring model flexibility with parameter space partitioning: An introduction and application example. *Cognitive Science*, 32, 1285-1303. doi:10.1080/03640210802477534
- Pitt, M. A., Myung, J. I., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, 109, 472-491. doi:10.1037/0033-295X.109.3.472
- Ryan, J. (1969a). Grouping and short-term memory: Different means and patterns of grouping. *Quarterly Journal of Experimental Psychology*, 21, 137-147. doi:10.1080/14640746908400206
- Ryan, J. (1969b). Temporal grouping, rehearsal and short-term memory. *Quarterly Journal of Experimental Psychology*, 21, 148-155. doi:10.1080/14640746908400207
- Towse, J., & Houston-Price, C. T. (2001). Combining representations in working memory: A brief report. *British Journal of Developmental Psychology*, 19, 319-324. doi:10.1348/026151001166100
- Treiman R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology, Learning Memory and Cognition*, 14, 145-152.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108, 550-592. doi:10.1037/0033-295X.108.3.550

Vousden, J. I., Brown, G. D. A., Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology*, 41, 101-175.
doi:10.1006/cogp.2000.0739

Table 1

Parameter estimates for the fits of the models to the present and previous experiments.

Domain	Data	Model									
		GP + PG					GP + PS				
		S_I	E_I	S_G	E_G	G_c	S_I	E_I	S_G	E_G	G_c
Verbal	Experiment	0.88	0.53	0.41	0.84	0.23	0.84	0.68	0.74	0.93	0.12
	R69E1	0.98	0.37	0.52	0.98	0.98	0.91	1.00	0.05	0.20	0.01
	H96E2	0.80	0.48	0.31	1.00	0.28	0.89	0.92	1.00	0.99	0.09
Spatial	Experiment	0.94	0.90	0.46	0.77	0.18	0.90	0.72	0.55	0.86	0.16
	HH15E3	0.93	0.93	0.47	0.82	0.20	0.91	0.81	0.61	0.83	0.17
	PAEJ06E4	0.92	0.94	0.57	0.97	0.31	0.96	0.76	0.06	1.00	0.22

Note— S_I = start marker strength for items; E_I = end marker strength for items; S_G = start marker strength for groups; E_G = end marker strength for groups; G_c = Gaussian perturbation; Experiment = present Experiment; R69E1 = Ryan (1969) Experiment 1; H96E2 = Henson (1996) Experiment 2; HH15E3 = Hurlstone & Hitch (2015a) Experiment 3; PAEJE406 = Parmentier et al. (2006) Experiment 4.

Table 2

Goodness-of-fit quantities for the fits of the models to the grouped verbal and spatial conditions of the present and previous experiments.

Domain	Data	Fitting	Model					
			GP + PG			GP + PS		
			$\ln L$	$\Delta \ln L$	Σ	$\ln L$	$\Delta \ln L$	Σ
Verbal	Experiment	Individual	44.11	1.59	61%	41.25	4.44	39%
	R69E1	Aggregate	55.49	0	-	50.46	5.03	-
	H96E2	Aggregate	47.84	0	-	40.57	7.27	-
Spatial	Experiment	Individual	46.14	3.04	22%	48.70	0.48	78%
	HH15E3	Individual	50.28	4.87	19%	54.88	0.27	81%
	PAEJ06E4	Aggregate	54.54	2.12	-	56.66	0	-

Note—Fitting = whether the model was fit to aggregate or individual participant data; $\ln L$ = log maximum likelihood (with the best model's $\ln L$ in bold face); $\Delta \ln L$ = difference in $\ln L$ with respect to the best fitting model; Σ = percentage of participants for which model provided best fit; Experiment = present Experiment; R69E1 = Ryan (1969) Experiment 1; H96E2 = Henson (1996) Experiment 2; HH15E3 = Hurlstone & Hitch (2015a) Experiment 3; PAEJ06E4 = Parmentier et al. (2006) Experiment 4.

Table 3

Details of the previous experiments fitted by the models.

Data	Domain	Stimuli	Presentation	Response	List	Grouping
			Modality	Modality	Length	Pattern
R69E1	Verbal	Digits	Auditory	Written	9	3-3-3
H96E2	Verbal	Digits	Visual	Written	9	3-3-3
HH15E3	Spatial	Squares	Visual	Reconstruction	9	3-3-3
PAEJ06E4	Spatial	Dots	Visual	Reconstruction	9	3-3-3

Note—R69E1 = Ryan (1969) Experiment 1; H96E2 = Henson (1996) Experiment 2; HH15E3 = Hurlstone & Hitch (2015a) Experiment 3; PAEJ06 = Parmentier et al. (2006) Experiment 4.

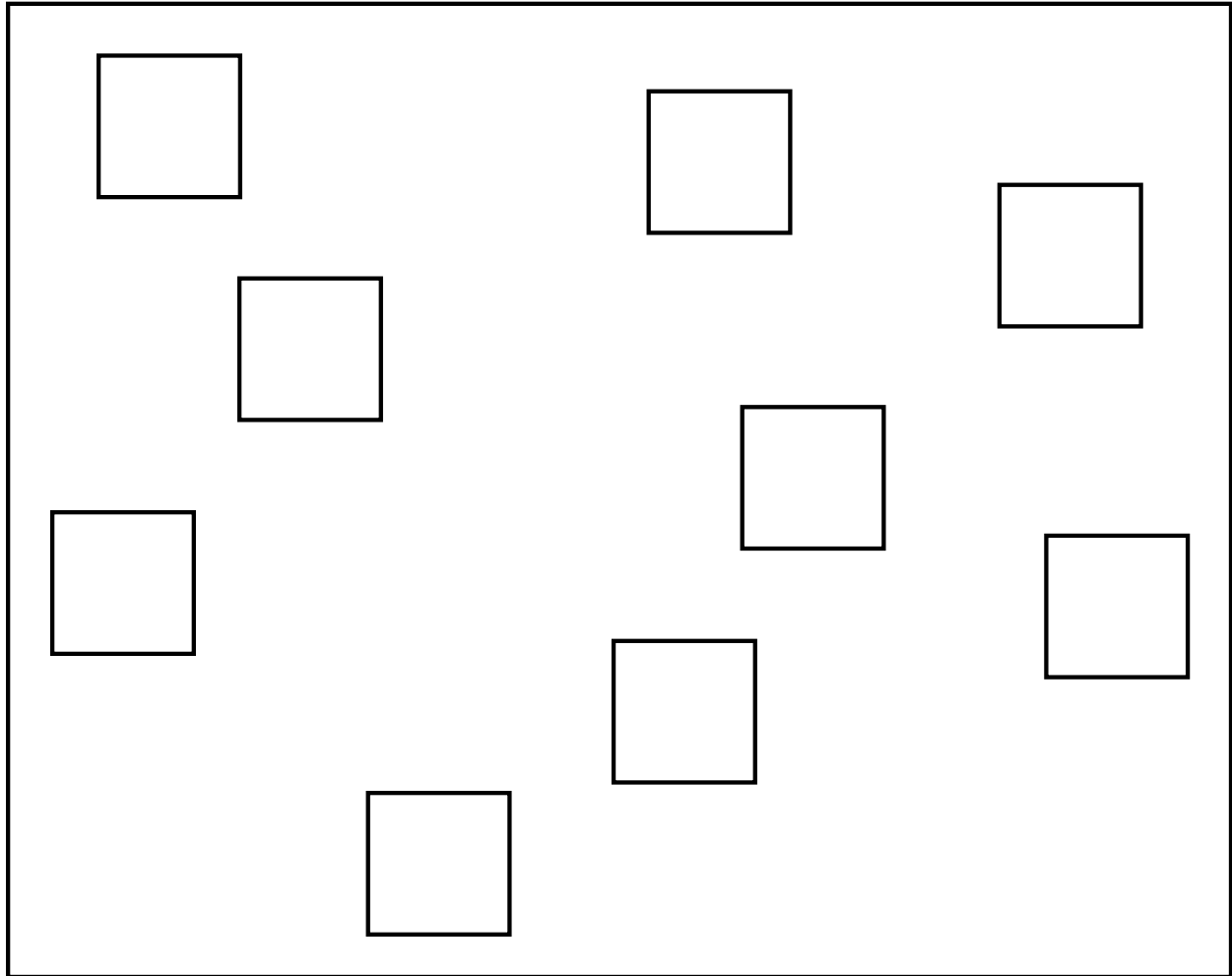


Figure 1. Schematic of the placement of the nine squares from which items were randomly selected for sequential presentation in the spatial serial recall task.

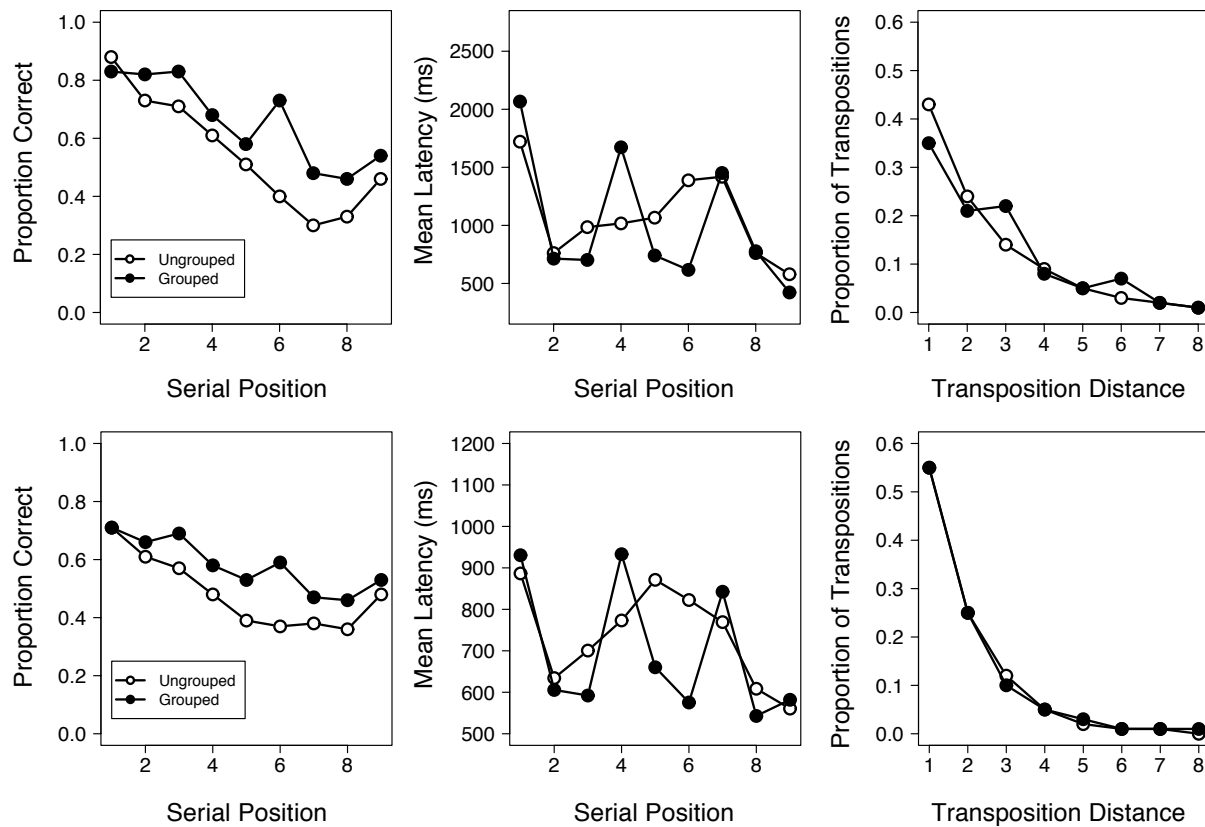


Figure 2. Accuracy serial position curves (left panels), latency serial position curves (middle panels), and transposition gradients (right panels) for the experiment. The upper panels show the data for the verbal task, whilst the lower panels show the data for the spatial task.

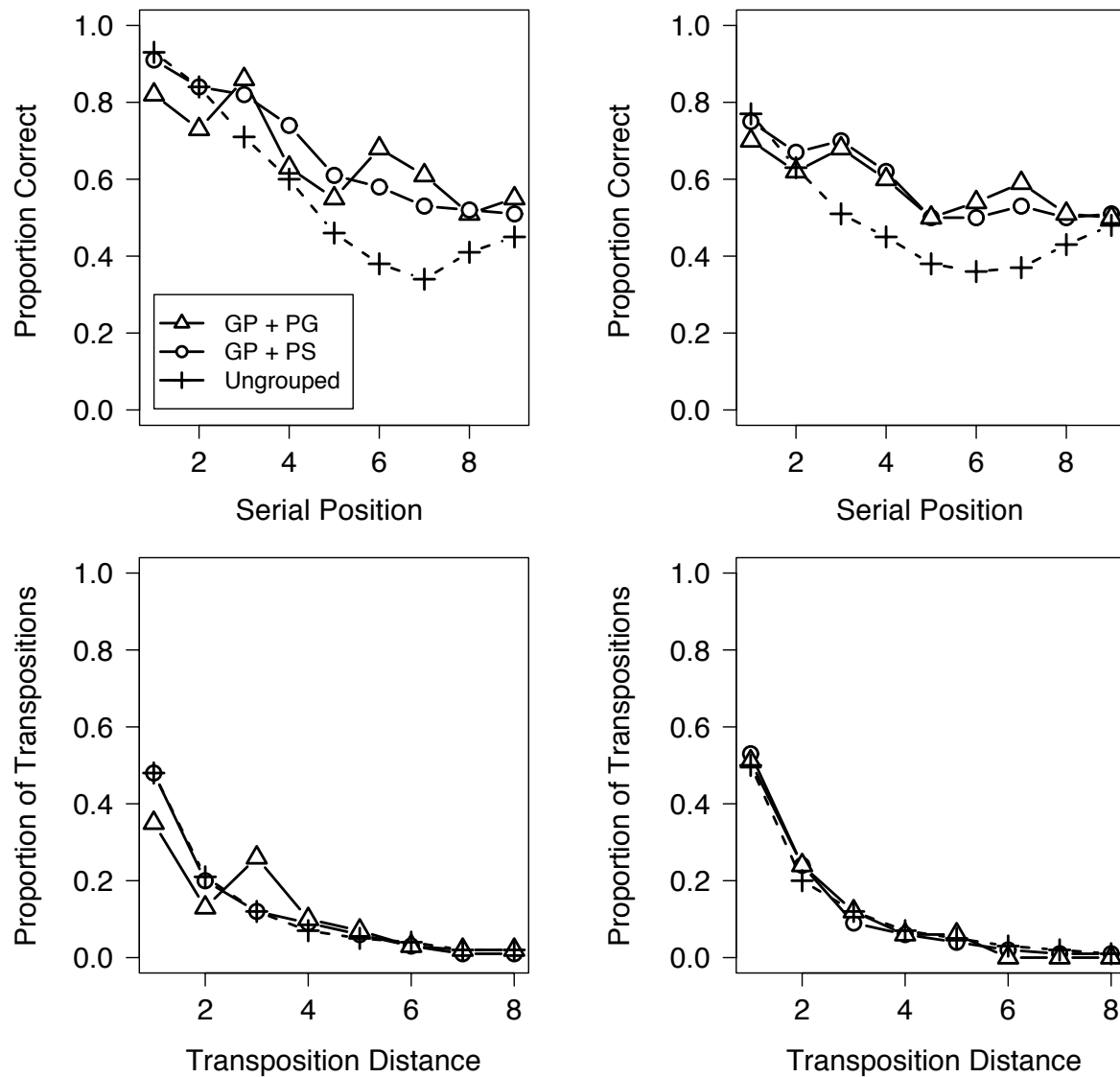


Figure 3. Serial position curves (upper panels) and transposition gradients (lower panels) predicted by the models for the grouped conditions of the Experiment. The left panels show the predictions for the verbal task, whereas the right panels show the predictions for the spatial task. GP + PG = position-of-group + position-in-group; GP + PS = position-of-group + position-in-sequence.

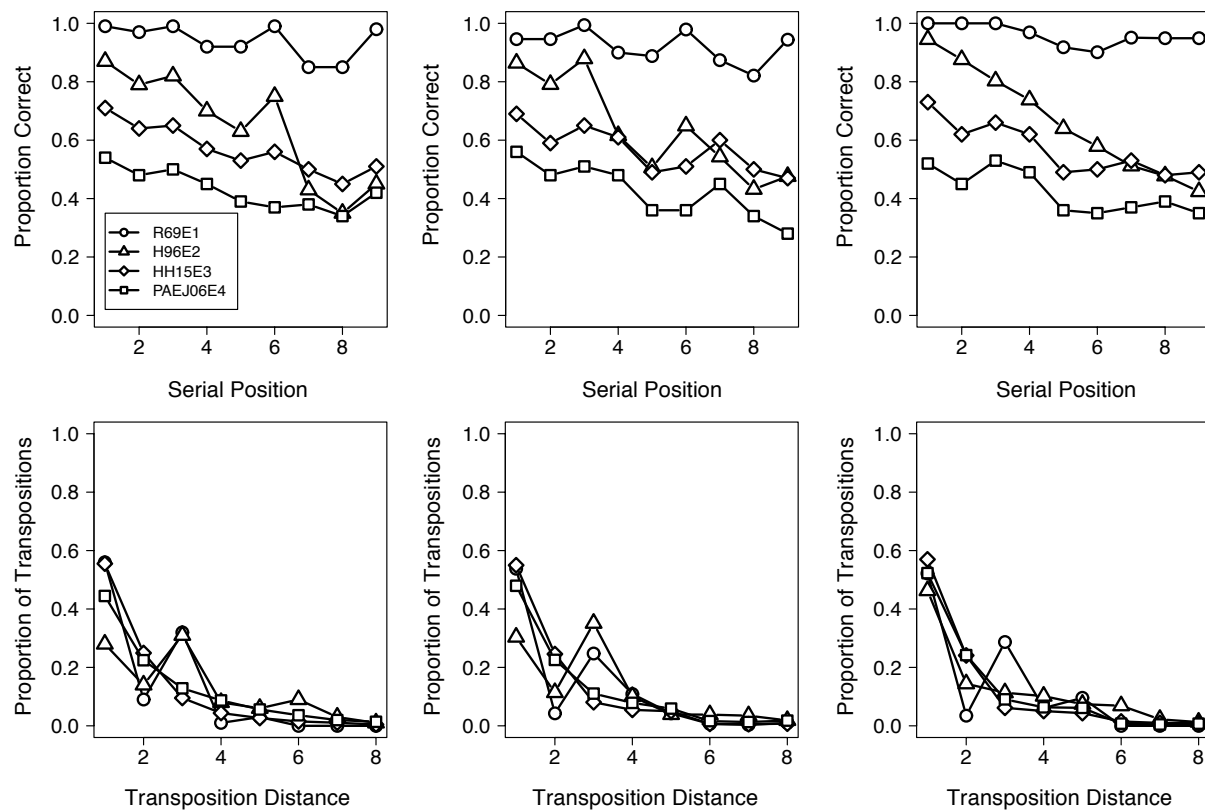


Figure 4. Serial position curves (upper panels) and transposition gradients (lower panels) for the previous serial recall experiments (left panels) and fits of the GP + PG model (middle panels) and GP + PS model (right panels) to those data. R69E1 = Ryan (1969) Experiment 1; H96E2 = Henson (1996) Experiment 2; HH15E3 = Hurlstone & Hitch (2015a) Experiment 3; PAEJ06 = Parmentier et al. (2006) Experiment 4.

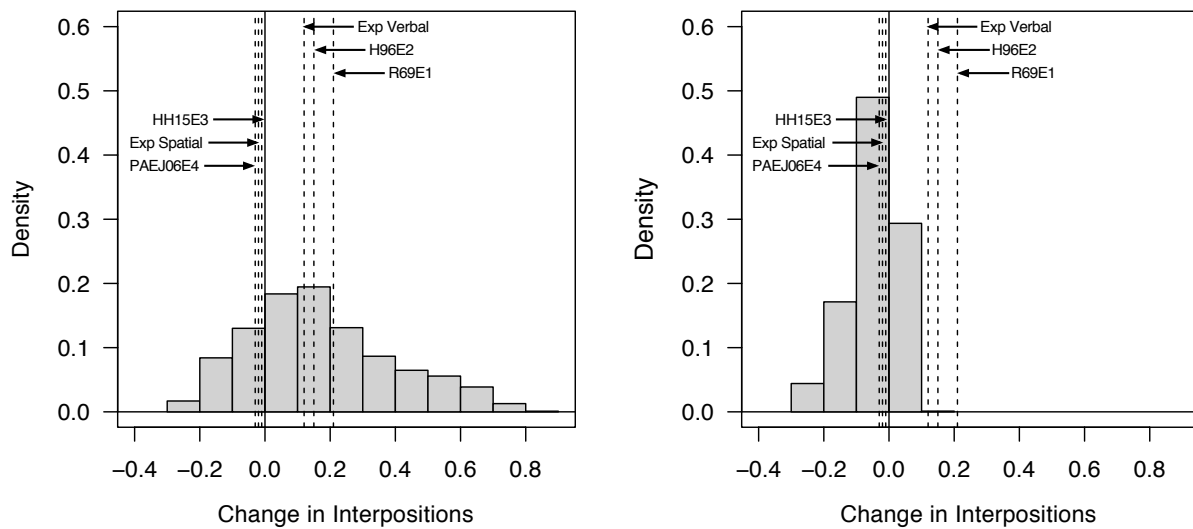


Figure 5. Density histograms showing the change in interpositions between ungrouped and grouped sequences across the parameter space of the GP + PG (left panel) and GP + PS (right panel) models. The solid vertical line in each panel corresponds to a change of 0—viz. a case where the incidence of interpositions is the same in grouped and ungrouped sequences. Bins to the left of the vertical line represent observations where the incidence of interpositions was smaller in grouped than in ungrouped sequences, whereas bins to the right of the vertical line contain observations where the incidence of interpositions was greater in grouped than in ungrouped sequences. The broken vertical lines in each panel correspond to the observed values from six experiments. Exp Verbal = verbal condition of present Experiment; H96E2 = Henson (1996) Experiment 2; R69E1 = Ryan (1969) Experiment 1; Exp Spatial = spatial condition of present Experiment; HH15E3 = Hurlstone & Hitch (2015a) Experiment 3; PAEJE4 = Parmentier et al. (2006) Experiment 4.

Appendix

Landscaping Analysis

To verify that the GP + PG model is more flexible than the GP + PS model—as indicated by the results of our parameter sensitivity analysis—we contrasted the models using another model comparison method known as *landscaping* (Navaro et al., 2004). This technique allows one to determine how well two models fit each others data; in other words, how well each model is able to mimic the behavior of the other. Like all landscaping analyses, our analysis involved three steps. In the first step, we generated a large number of data sets (viz. 1000) for each model. These were generated by sampling model parameter values for S_I , E_I , S_G , and E_G (r and G_c were set to constant values of .95 and .15, respectively, as per the parameter sensitivity analysis) from uniform distributions and then generating model predictions for accuracy serial position curves and transposition gradients for 3-3-3 grouped 9-item sequences using the resulting parameter values. In the second step, we added noise to the model predictions. In the third step, we fitted each model to its own data set, as well as its counterparts, using the same procedure used to fit the models to the experiments.

The results of the analysis are shown graphically as landscape plots in Figure A1. Each plot shows the fits of the data generating model on the x axis against the fits of the competing model on the y axis. Points lying on the diagonal line represent instances where both models provide an equally good fit of the data. Points below the diagonal line represent instances where the data generating model fit its own data best, whereas points above the diagonal line represent instances where the competing model provided a better fit than the data generating model. It can be seen from inspection of the figure that when the GP + PS model is the data generating model (left panel), the data points fall approximately symmetrically around the criterion line. Thus, whilst there are several instances where the GP + PS model provides the best fit to its own data (points below the criterion line), there are a comparable number of instances where the GP + PG model provides a better fit than its data generating counterpart (points above the criterion line). By contrast, when the GP + PG model is the data generating model (right panel), the data points do not fall symmetrically around the criterion line. Instead,

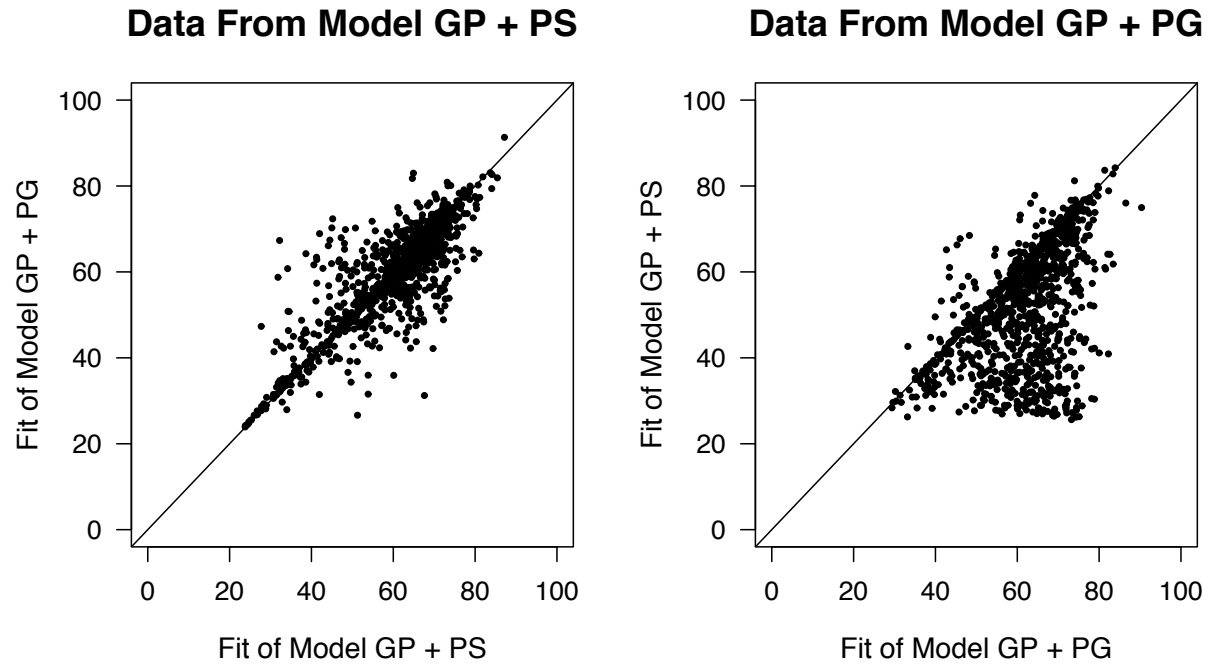


Figure A1. Landscape plots for data generated by the GP + PS model (left panel) and the GP + PG model (right panel). Each plot shows the fits of the data generating model on the x axis against the fits of the competing model on the y axis. Points lying on the diagonal line represent instances where both models provide an equally good fit of the data. Points below the diagonal line represent instances where the data generating model fit its own data best, whereas points above the diagonal line represent instances where the competing model provided a better fit than the data generating model.

the majority of the data points fall below the criterion line—indicating that the GP + PG model most often provided the best fit to its own data—and only a few data points fall above the criterion line—indicating that the GP + PS model rarely provides a better fit than the data generating model.

In brief, the results of the landscaping analysis confirm that, despite sharing the same number of model parameters, the GP + PG model is more flexible than the GP + PS model, and this enhanced flexibility enables it to mimic the behavior of the latter model.