

# Temporal grouping and direction of serial recall

Yang S. Liu

Department of Psychology and Psychiatry, University of Alberta, Edmonton, Alberta

Jeremy B. Caplan

Department of Psychology and Neuroscience and Mental Health Institute, University of Alberta, Edmonton, Alberta, Canada

## Abstract

When lists are presented with temporal pauses between groups of items, participants' response times reiterate those pauses. Accuracy is also increased, especially at particular serial positions. By comparing forward with backward serial recall, we tested whether the influence of temporal grouping is primarily a function of serial position or output position. Results favoured the latter, both when recall direction was known to participants prior to (Experiment 1) or only after (Experiment 2) studying each list. Alongside fits of variants of a temporal distinctiveness-based model, our findings suggest that the influence of temporal grouping is not just a consequence of grouping-information stored during the study phase. Rather, it depends critically on participants cueing with within-chunk position during recall, combined with response suppression.

*Keywords:* serial recall, order memory, memory models, grouping, response suppression

## Introduction

Memory for sequences is widely understood to be at the core of a broad range of human behaviour (e.g., Ebbinghaus, 1885/1913; Lashley, 1951). Accordingly, a major goal of memory research has been to determine how serial lists are stored and retrieved, with particular focus on the immediate serial-recall paradigm. In immediate serial recall, participants are presented with a list of items, such as consonants, for study, then asked to reproduce the list in order. This procedure resembles everyday tasks like remembering a phone number long enough to enter it into a smart phone or remembering a line of song lyrics long enough to repeat it in a pop concert. The effect of temporal grouping is prominent among the major empirical findings that have been used to constrain and guide the development of mathematical models of serial recall. Introduced by Ryan (1969a, 1969b), longer pauses are introduced between subgroups of items, while equating total list presentation time. For example, a list of nine consonants might be subdivided into three groups of three letters. Resulting plots of accuracy show a scalloping effect (e.g., Hitch, Burgess, Towse, & Culpin, 1996; Ng & Maybery, 2002; Ryan, 1969a, 1969b), with a relative recall advantage for the first and last item of each temporally defined group (e.g., items 1, 3, 4, 6, 7 and 9) compared to the respective middle-group items.

The dominant theoretical account of the advantage of grouped over control lists is that grouping affords an additional cue participants can use during recall. This idea has been incorporated in numerous models that rely on positional or ordinal cueing for serial recall (e.g., Brown, Preece, & Hulme, 2000; Brown, Neath, & Chater, 2007; Burgess & Hitch, 1999; Henson, 1998). Thus, in the well supported, Scale-Independent Memory, Perception and LEarning model (SIMPLE; Brown et al., 2007), for example, control lists are recalled by cueing memory with time (specifically, study–test recency of the desired serial position), whereas grouped lists are recalled by cueing with a weighted sum of time and position of the desired item within the current group (e.g., 1, 2 or 3, for lists composed of groups of three items).

Backward serial recall can provide additional constraints on model mechanisms, be-

cause it helps unravel effects of serial position versus response (output) position. However, we could find no published data on backward serial recall of temporally grouped lists. The closest precedent, to our knowledge, was by Anderson, Bothell, Lebiere, and Matessa (1998), who investigated both forward and backward recall of grouped and ungrouped lists. Latency measure showed long pauses across group boundaries, and the effect of grouping was similar between recall directions. Their lists were presented sequentially, but with groups denoted visually, both at study and test. In fact, serial position (of items within the whole list) were also cued spatially both during study and during recall. Thus, their outcome could have been due to explicit position-cueing at study and test, similar to what was found by Guitard, Saint-Aubin, Poirier, Miller, and Tolan (2019). Here we present data from two experiments comparing temporally grouped and ungrouped lists, between forward and backward serial-recall directions. No indication of serial position was visually present during either study or recall. In Experiment 1, participants had foreknowledge of recall direction, and thus had the opportunity to customize their approach to study depending on direction. In Experiment 2, we prevented this by cueing serial-recall direction of each list only after the list had been studied.

We sought to test the standard account, where serial recall is cued by a combination of serial position (or time) and within-group position). Because the additional dimension of discriminability is a function of serial position, we predicted the benefit due to temporal grouping would be a function of serial position, and not of output position. The results of our first experiment<sup>2</sup> directly contradicted this prediction.<sup>3</sup> Participants either studied control lists or temporally grouped lists of nine consonants, and were either asked to recall the lists in order of presentation (“forward” recall direction) or in reverse order (“backward” recall direction). Both grouped/control and forward/backward direction were manipulated factorially between subjects. To anticipate the results, the benefits of temporal grouping were a function of output position (and the benefit did not appear at each position), rather

---

<sup>2</sup>Experiment 1 was, in fact, designed with an entirely different goal in mind: to provide comparison data for a study of judgements of relative order of temporally grouped lists, reported by Liu (2015).

<sup>3</sup>Indeed, early attempts at fitting models to the aggregate data initially led us to conclude that this account was ruled out, the opposite of our ultimate conclusion here.

than serial position (initial presentation order). That is, backward-recall serial position effects were nearly mirror images of the forward-recall serial position effects. Experiment 2 was designed to test whether the effects would extend to participants who were post-cued for recall direction, to equate cognitive processes during study. The same basic outcome, the benefits of temporal grouping being a function of output position rather than serial position, was once again found. The question then became, whether the standard account of temporal grouping could produce such a pattern. Our intuition was that it could not, because the benefit of the additional retrieval cue (within-group position) should be a function of serial position, regardless of order of recall. After presenting data from the two experiments, we report fits of variants of SIMPLE: first, using only time as a retrieval cue, to check whether the temporal pattern present during study, combined with participants' tendency to reiterate this pattern in their interresponse times during recall, might already be sufficient to explain the data. We then compare with the two-dimensional model, adding within-group position as a retrieval cue. Finally, because backward serial recall is typically characterized by a recency-dominant serial position curve, rather than the primacy-dominant serial position curve found in the forward direction (e.g., Anderson et al., 1998; Bireta, Fry, Jalbert, Neath, & Surprenant, 2010; Farrand & Jones, 1996; Guérard, Saint-Aubin, Burns, & Chamberland, 2012; Guérard & Saint-Aubin, 2012; Haberlandt, Lawrence, Krohn, Bower, & Thomas, 2005; Li & Lewandowsky, 1993, 1995; S. A. Madigan, 1971; St. Clair Thompson & Allen, 2013; Thomas, Milner, & Haberlandt, 2003), we wondered if response suppression (participants' reluctance to repeat items in serial recall; Duncan & Lewandowsky, 2005) might be an influential factor that could affect model fit.

In both experiments, temporal grouping was manipulated between subjects, as is typically done (e.g., Ryan, 1969a, 1969b; Hitch et al., 1996; Ng & Maybery, 2002), because the presence of temporally grouped lists might induce participants to conceptualize ungrouped lists as grouped as well. Even so, some level of spontaneous grouping tends to be found even with ungrouped presentation procedures (e.g., Martin & Noreen, 1974; S. Madigan, 1980; Jou, 2011; Wickelgren, 1967; Ryan, 1969a). The control condition was ungrouped only in

the sense that presentation rate was not grouped, and the interpretation will rest upon the assumption that participants will organize lists into groups *less* in the control than grouped conditions. Recall direction was manipulated between subjects in Experiment 1. This had the advantage of allowing participants to optimize their performance on both backward and forward recall directions, but participants could have altered their study strategy between recall directions. To address this, recall-direction was manipulated within subjects in Experiment 2, by post-cueing direction. This ensured that participants could not study a list differently in anticipation of backward versus forward recall. Finally, fits of four variants of SIMPLE are presented, to directly test whether the standard account of temporal grouping might be able to accommodate our findings.

## Experiment 1

### Methods

**Participants.** A total of 168 undergraduate students from introductory psychology courses at the University of Alberta participated in exchange for partial course credit. The sample size was selected to be close to a related experiment, testing judgements of relative order of the lists with the same design, and with the same between-subjects design with four groups (Liu, 2015, chapter 5). Participants gave informed consent, had normal or corrected-to-normal vision and learned English before age 6. Participants were run in groups of about 10–15, with random assignment to testing conditions as follows. Forward/Grouped:  $N = 41$ ; Forward/Control:  $N = 42$ ; Backward/Grouped:  $N = 43$ ; Backward/Control:  $N = 42$ .

**Materials & Procedure.** To maintain continuity with previous studies (Chan, Ross, Earle, & Caplan, 2009; Liu, Chan, & Caplan, 2014), stimuli were 16 consonants (excluding S, W, X, and Z) from the English alphabet displayed in uppercase. Each list comprised 9 consonants drawn at random without replacement from the stimulus pool. Probability was equal for each consonant/serial-position combination. All participants were tested using a group of 15 computers (custom-built PCs) with identical hardware, identical

Samsung SyncMaster B2440 monitors and Logitech K200 keyboards, to minimize hardware precision variability in our between-subjects design (Plant & Turner, 2009).

The experiment was implemented with the Python Experiment-Programming Library (PyEPL; Geller, Schleifer, Sederberg, Jacobs, & Kahana, 2007). The procedure for the Control condition was identical to Liu et al.'s (2014) judgements of relative order experiment, but with list length 9 and a serial-recall test. Depending on Grouping, participants were either asked to type the list in forward or backward order: (a) Excerpt from “forward” instruction: “. . . you will be asked to type the list you just saw, starting from the first letter and ending with the last letter. In other words, type the list in forward order. . . .” (b) Excerpt from “backward” instruction: “. . . you will be asked to type the list you just saw, starting from the most recent letter and end with the first letter. In other words, type the list in backward order. . . .”. Participants were asked to press the SPACE key to skip over items they do not remember. Each trial began with a fixation asterisk, ‘\*’, in the center of the screen, followed by a consonant list presented sequentially in the center of the screen. Items were presented for 500 ms each. The Control condition had a constant inter-stimulus interval (ISI) of 350 ms, whereas the Grouped participants had an ISI of 950 ms between items 3 and 4, and between items 6 and 7, to create a longer temporal gap between-groups, and an ISI of 150 ms for all other transitions. The ISIs were selected to maintain a constant total presentation time of 7300-ms between Grouped and Control conditions (as done by, for example, Henson, 1998; Hitch et al., 1996). After a 2500-ms delay, participants were cued with an input line and a text reminder to type the list either in forward or backward direction. Participants could not backtrack to edit entered consonants, and they terminated recall by pressing the ENTER key. All letters of the English alphabet were accepted as input, and all typed letters stay on the computer screen until ENTER was pressed. After a 500-ms delay, participants could press a key to start the next trial. A total of 80 trials were presented in 5 blocks, where the first block contains 8 practice trials, followed by 4 blocks of 16 trials.

**Data analysis.** We analyzed our data with linear mixed effects (LME) models (Baayen, Davidson, & Bates, 2008; Bates, 2005). We adopted LME analysis because compared to ANOVA, LME can fit individual responses without the need for averaging of the data, and protects against Type II error due to increased power (Baayen et al., 2008; Baayen & Milin, 2010). LME analyses were conducted in R (Bates, 2005), using the LME4 (Bates & Sarkar, 2007), LanguageR (Baayen, 2007) and LMERConvenienceFunctions (Tremblay, 2013) libraries. The “lmer” function was used to fit the LME model. The “Anova” function from Companion to Applied Regression package (Fox & Weisberg, 2011) was used to conduct Wald chi-square test for the best fitting models.

We conducted accuracy analyses with Serial Position (order of an item during presentation) included as a factor and then again with Output Position (response number within a given recall sequence) instead. Serial Position and Output Position were treated as a continuous fixed-effects predictors. Categorical fixed-effects predictors were Grouping and Recall Direction (between-subjects). Subject was included as a random effect on intercept. The accuracy data were fitted with logistic regression as it is a binary variable (“correct” vs. “incorrect”). Because the latency measure was an inter-response time, the time difference between two successive responses, we analyzed latency data only with respect to Output Position. Inter-response times leading into errors were excluded from the analysis. Inter-response time was log-transformed to reduce skewness.

LME estimates random effects first, followed by fixed effects. In the results tables, the “Estimate” column reports the corresponding regression coefficients, along with their standard errors. For the purposes of reporting the LME results, Control condition, Forward direction were set as the reference levels for their corresponding variables. The best fits of LME models were obtained by conducting a series of iterative tests comparing progressively simpler models with more complex models using the Bayesian Information Criterion (BIC), as was done by Liu et al. (2014), using LMERConvenienceFunctions developed by Tremblay (2013).

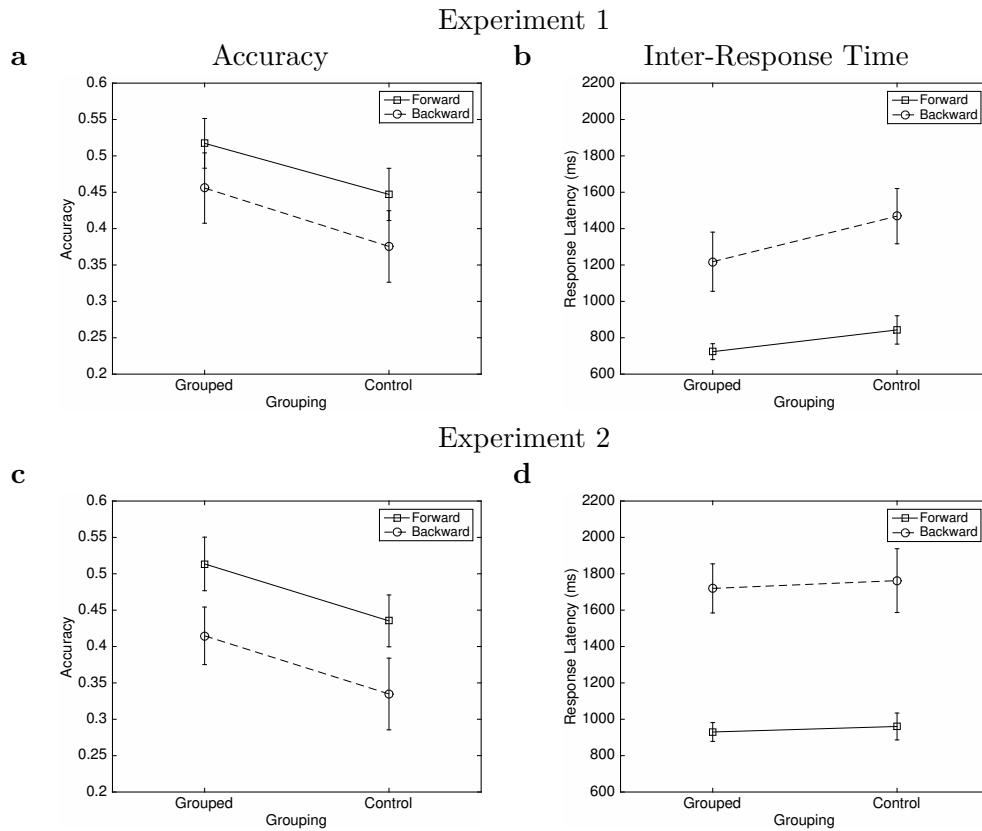
## Results

Nine participants were excluded because they recalled fewer than 2 items on average (making their data uninformative with respect to serial-order, scalloping effects, etc.), as follows: Forward/Grouped: 0; Forward/Control: 2; Backward/Grouped: 1; Backward/Control: 6. First we comment on the global effects of the between-subjects factors on accuracy and response time, and then drill down into serial-position and output-position effects.

**Global effects of grouping and recall direction on accuracy and inter-response time.** We first plot the effects of Grouping and Direction, collapsing across Input/Output Positions (Figure 1). For both accuracy and inter-response times of correct recalls, the basic pattern was the same: Backward recall was worse than forward recall, replicating numerous prior studies of backward recall. Also replicating prior temporal-grouping studies, Grouped lists produced higher accuracy (Figure 1a) and faster inter-response times (1b) than ungrouped, Control lists. The advantage of Grouped over Control was significant individually for Forward and Backward, both experiments, for accuracy. For inter-response time, the advantage of Grouped over Control was significant individually for Forward and Backward in Experiment 1, but for neither direction in Experiment 2. This pattern shows, for the first time, that grouping enhanced recall in the backward direction to a similar degree, overall, as in the forward direction. This may indicate that grouping influenced backward recall via a similar mechanism as forward recall, but this could be due to effects at study or during recall. To address this, we next break down accuracy and inter-response time by serial position and then by output position.

**Do grouping effects depend on input or output position?** Figure 2 plots serial-position curves for all four conditions. First, the grouped conditions show the characteristic scalloping pattern: mini U-shaped serial-position curves within groups (Ryan, 1969a). Unlike some previous results (e.g., Hitch et al., 1996), grouped lists do not exhibit an advantage at every serial position. This may be viewed as proof-of-principle that an account of temporal grouping that assumes position is encoded with greater resolu-





*Figure 1.* Mean accuracy (a, c) and inter-response time (b, d) as functions of group for Experiment 1 (a, b) and Experiment 2 (c, d). Error bars are 95% confidence intervals based on standard error of the mean. For both experiments, Grouped and Control were compared using Welch's  $t$ -test, as follows. Accuracy: Exp. 1 Forward:  $t(80.9) = 2.74$ ,  $p < 0.01$ , Backward:  $t(82.9) = 2.36$ ,  $p < 0.05$ ; Exp. 2 Forward:  $t(108) = 4.49$ ,  $p < 0.01$ , Backward:  $t(108) = 3.72$ ,  $p < 0.01$ . IRT, Exp. 1 Forward:  $t(64.4) = -2.59$ ,  $p < 0.05$ , Backward:  $t(82.7) = -2.18$ ,  $p < 0.05$ ; Exp. 2 Forward:  $t(97.3) = -1.53$ ,  $p = 0.13$ , Backward:  $t(107) = -.03$ ,  $p = 0.98$ .

tion due to an additional, within-group position code, may be insufficient. At the very least, a weighting function may be required to explain such effects, as in SIMPLE, where the within-group weight and the temporal-dimension weight must sum to 1 (Brown et al., 2007). Next, the overall serial-position curves for backward recall are nearly (although not precisely) mirror-images of the corresponding curves for forward recall, as previous studies have found. Moreover, the effects of grouping are approximate mirror-images as well. Thus, the visual impression is that the benefit of grouping is not a function of serial position, but rather, of output position. The difference between the grouped and control conditions, plotted as a function of output position, confirms this, as it is quite similar for backward as for forward recall (Figure 2c).

To check these effects, we fitted linear mixed effects models with either Serial Position (Table 1a) or Output Position (Table 1b), along with Grouping and Direction. Confirming visual inspection, results from the best fitting linear mixed effects models found, on the one hand, a Grouping  $\times$  Direction  $\times$  Serial Position interaction, and on the other hand, no Grouping  $\times$  Direction  $\times$  Output Position interaction. To select between these two best models, the BIC values show that the best-fitting model using Output Position absent a three-way interaction fit better than the best-fitting model with Serial Position (which includes a three-way interaction):  $\Delta \text{BIC} = 19$ , comfortably surpassing the conventional criterion of  $\Delta \text{BIC} > 2$ .

Turning next to inter-response time as a function of output position (Figure 3), all conditions showed the characteristic slow initiation time (Bireta et al., 2010, but see Thomas et al., 2003, for different results with backward recall). After the first recall, the Grouped participants had longer recall latencies across group boundaries, in exchange for shorter latencies within-group, compared to Control participants. This, too, held for both Forward and Backward recall directions. However, inter-response times were overall longer in the Backward than Forward direction. The scalloping that is considered evidence of subjective chunking based on temporally grouped presentation was evident in both Forward and Backward recall. In fact, it was even more pronounced in the Backward direction, and even

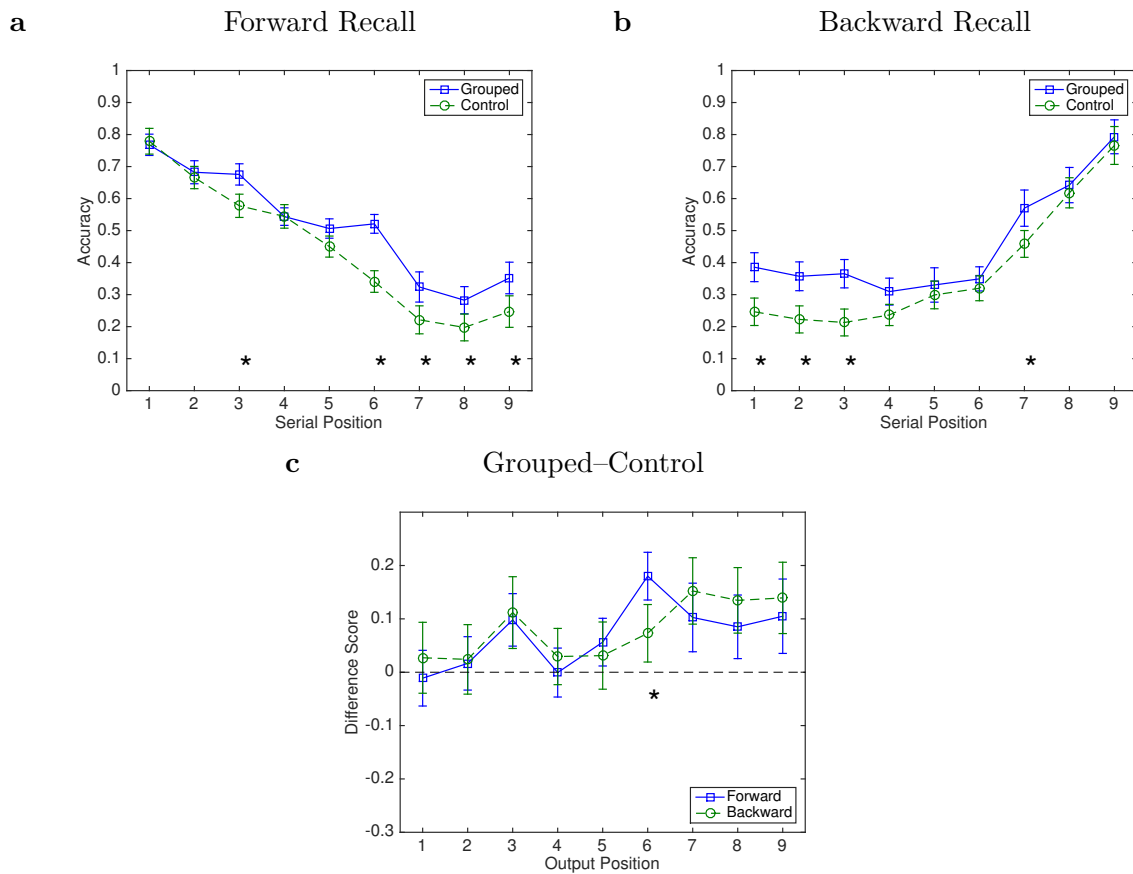


Figure 2. Recall accuracy as a function of serial position for forward recall (panel a) and backward recall (panel b). Accuracy difference of Grouped–Control as a function of output position for forward and backward recall (panel c). Error bars denote 95% confidence intervals based on standard error of the mean. \*  $p < 0.05$ , based on a two-tailed  $t$  test.

clearly evident in the Backward Control participants.

**Within-list intrusions.** A favourite finding that has been used to support the standard, two-dimensional positional-coding account of temporal grouping is the increase in interpositions (an item recalled within the wrong group, but nonetheless, at the correct within-group position) and relative decrease in adjacent transpositions (an item exchanged with its neighbour) from control to grouped lists (Henson, 1998; Lee & Estes, 1981; Wickelgren, 1964). We also replicated the classic effects of grouping on within-list intrusion types, for forward serial recall (Figure 4 and Table 4). However, for backward recall, grouping did not produce an increase in the rate of interpositions (Figure 4), supported by a significant

Effect	Estimate (SE)	p
<b>a Analysis using Serial Position (BIC=127169)</b>		
Main effects		
Intercept	1.02 (0.10)*	<.001
Grouping	-0.054 (0.13)	0.678
Direction	-2.47 (0.06)*	<.001
Serial Position	-0.27 (0.01)*	<.001
Interactions		
Grouping×Direction	0.78 (0.08)*	<.001
Grouping×Serial Position	0.08 (0.01)*	<.001
Direction×Serial Position	0.40 (0.01)*	<.001
<b>Grouping × Direction × Serial Position</b>	-0.15 (0.01)*	<.001
<b>b Analysis using Output Position (BIC=127150)</b>		
Main effects		
Intercept	0.99 (0.10)*	<.001
Grouping	-0.005 (0.13)	0.971
Direction	-1.03 (0.04)*	<.001
Output Position	-0.26 (0.01)*	<.001
Interactions		
Grouping×Output Position	0.08 (0.01)*	<.001
Direction×Output Position	0.12 (0.01)*	<.001
$\Delta$ <b>BIC=19</b> , favouring the model with Output Position		

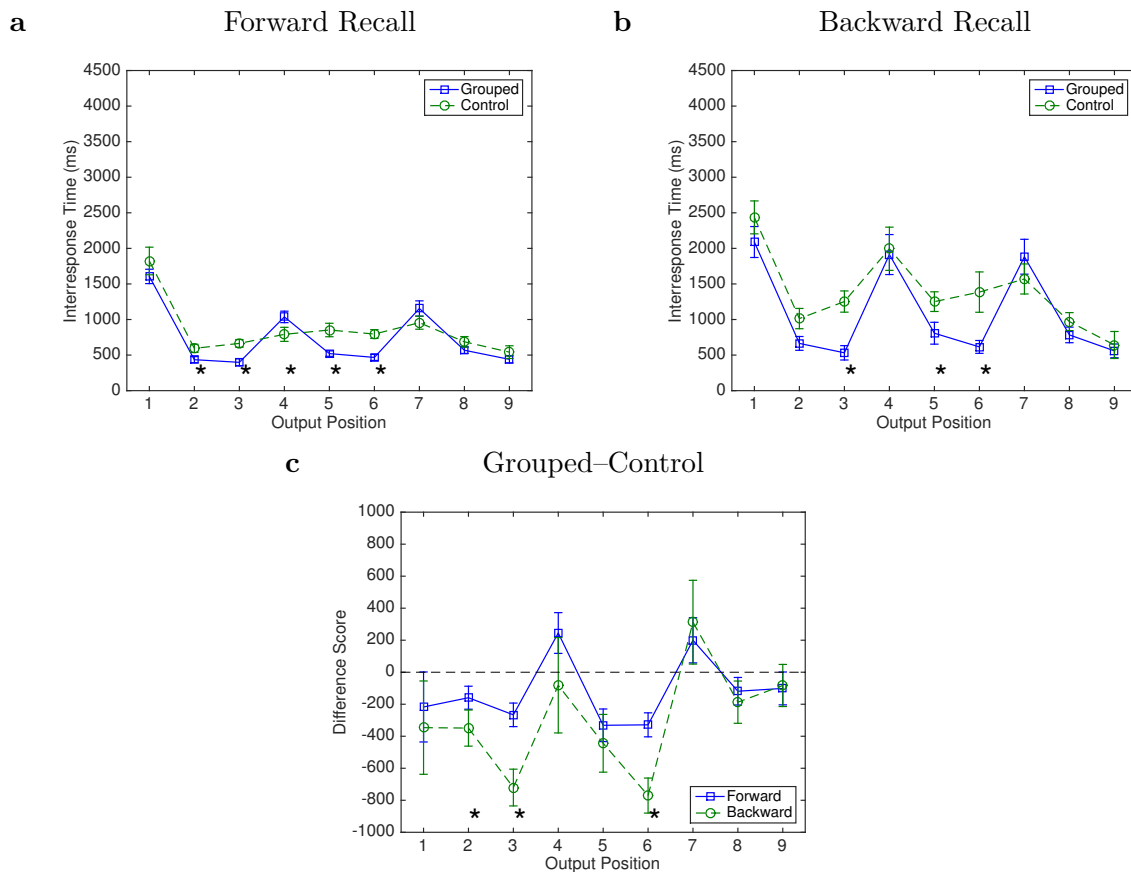
Table 1

The best-fitting LME model for recall accuracy (proportion correct) with Serial Position as a factor (a) or with Output Position as a factor (b) for Experiment 1. Estimate - the corresponding regression coefficient, along with its SE (standard error of the mean). BIC favours the model with Output Position; note that (a) includes a three-way interaction (highlighted in boldface) that is absent from (b). \* $p < 0.05$ .

Main effects	$\chi^2$	p
Intercept		
Grouping	16.8*	<.001
Direction	29.9*	<.001
Output Position	2890*	<.001
Interactions		
Grouping×Output Position	39.93*	<.001

Table 2

The best-fitting LME model for log inter-response time with Output Position as a factor for Experiment 1. The  $\chi^2$  column reports  $\chi^2$  from Wald test. \*  $p < 0.05$ .



*Figure 3.* Inter-response time as a function of output position for forward recall (panel a) and backward recall (panel b) for Experiment 1. Inter-response time difference of Grouped-Control as a function of output position for forward and backward recall (panel c). Error bars denote 95% confidence intervals based on standard error of the mean. Significant differences between Grouped and Control condition are denoted by  $*p < 0.05$ , based on a two-tailed  $t$  test.

Error Type  $\times$  Direction  $\times$  Grouping interaction in the best fitting LME model (Table 3).

Taking the standard argument further, if the advantage conferred by grouping is due to cueing with a two-dimensional position code, and the presence of interposition errors signifies the usage of within-group position as a cue (Wickelgren, 1964), then the more a participant commits interpositions (expressed as a proportion of all within-list intrusions), the greater their serial-recall accuracy should be. A positive correlation between interposition rate and accuracy is predicted for grouped conditions. The correlation was, indeed, positive and significant, but only for the Forward Control condition

Main effects	$\chi^2$	p
Intercept		
Grouping	0.25	0.615
Direction	383.22*	<.001
Error Type	70.37*	<.001
Interactions		
Error Type $\times$ Direction	3.94*	<.047
Error Type $\times$ Grouping	94.85*	<.001
Direction $\times$ Grouping	6.12*	0.013
Direction $\times$ Error Type $\times$ Grouping	26.45*	<.001

Table 3

The best-fitting LME model for error proportion scores (proportion of errors normalized by within-list errors) with Direction Grouping and Error types as factors, for Experiment 1. The  $\chi^2$  column reports  $\chi^2$  from Wald test. \* $p < 0.05$

	Forward		Backward	
	Grouped	Control	Grouped	Control
Experiment 1				
Adjacent Transpositions	32.11%	39.99%	12.06%	16.21%
Interpositions	25.75%	14.32%	6.25%	6.19%
Experiment 2				
Adjacent Transpositions	27.21%	38.68%	10.28%	17.01%
Interpositions	29.71%	14.74%	8.23%	6.79%

Table 4

Rates of interposition and adjacent errors (relative to all within-list intrusions) for each condition. Cell values are calculated by dividing the number of adjacent transposition and interposition errors by total number of errors per condition.

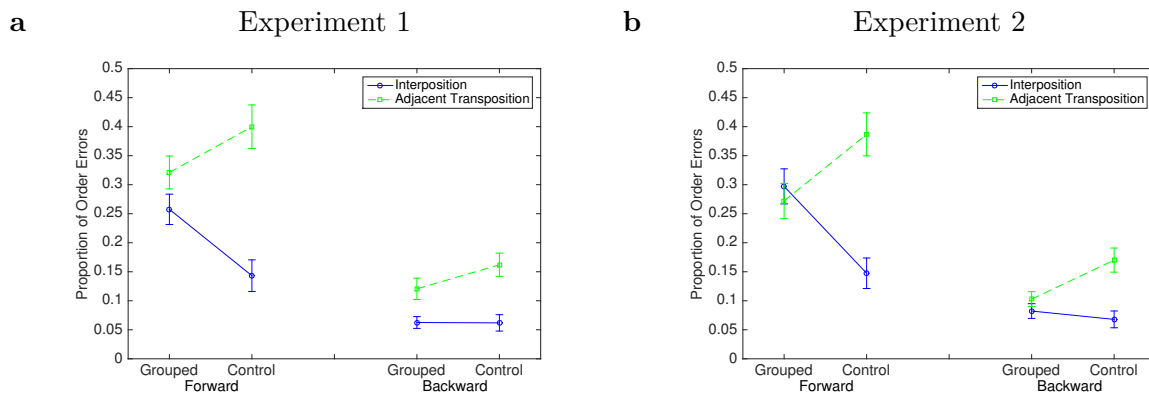


Figure 4. Proportion of adjacent errors and interposition errors for each condition. The error bars denote the 95% confidence interval based on standard error of the mean.

$r(38) = 0.33$ ,  $p = 0.039$ . For the Backward Control condition, the correlation was significant but negative,  $r(34) = -0.37$ ,  $p = 0.025$ , opposite the prediction. Moreover, for the Grouped conditions, the correlations were both quite small and non-significant, (Forward:  $r(39) = -0.063$ ,  $p = 0.70$ ; Backward:  $r(40) = -0.127$ ,  $p = 0.42$ ); thus, we found no simple relationship between interposition rates and the advantage due to grouping.

## Experiment 2

### Methods

**Participants.** A total of 110 undergraduate students from introductory psychology courses at the University of Alberta participated in exchange for partial course credit. Sample size was determined to be comparable to that of Experiment 1, but we anticipated greater sensitivity since one factor was within, rather than between subjects, thus requiring a smaller sample. Participants gave informed consent, had normal or corrected-to-normal vision and learned English before age 6. Participants were run in groups of about 10–15, with random assignment to testing groups.

**Materials & Procedures.** The procedure was identical to Experiment 1, except that participants were told sometimes they need to recall in forward order, and sometimes they need to recall in backward order, and they need to follow the instruction on the computer screen. The text reminder in Experiment 1 was replaced with the directionality cue (“Type the list in FORWARD order” or “Type the list in BACKWARD order”), after each list had been presented for study.

### Results

Participants recalling fewer than 2 items on average (making their data uninformative with respect to serial-order, scalloping effects, etc.) were excluded from the analysis: Grouped: 3; Control: 14. The number of included participants was significantly different between conditions,  $\chi^2(1)=40.33$   $p < .001$ .

As in Experiment 1, we first look at Grouping and Direction collapsing across Input/Output Positions for both accuracy and inter-response times of correct recalls (Figure 1). Direction is a within-subjects factor and Grouping is a between-subjects factor. The basic pattern of results was similar to Experiment 1, with main effects of Grouping and Direction.

Figure 5 plots serial-position curves for all four conditions. The Grouped participants showed the characteristic scalloping pattern. Once again, the overall serial-position curves for backward recall are nearly (although not precisely) mirror-images of the corresponding curves for forward recall and the effects of grouping are approximate mirror-images as well. This visual impression that the benefit of grouping is a function of output position is further confirmed by the plot of Grouped–Control accuracy as a function of output position (Figure 5c), which is quite similar for backward as for forward recall .

As in Experiment 1, we fitted linear mixed effects models with either Serial Position or Output Position (Table 5), along with Grouping and Direction. The best fitting linear mixed effects models using Serial Position as a factor found a Grouping  $\times$  Direction  $\times$  Serial Position three-way-interaction. Using Output Position instead, the best fitting model did not include Grouping  $\times$  Direction  $\times$  Output Position. As in Experiment 1, the BIC values show that the best-fitting model using Output Position (absent a three-way interaction fit better than the best-fitting model with Serial Position (which includes a three-way interaction):  $\Delta\text{BIC} = 21$ .

Inter-response time as a function of output position looks similar to Experiment 1 (Figure 6). All groups show the characteristic slow initiation time. After the first recall, the Grouped participants had longer recall latencies across group boundaries than Control participants, in exchange for shorter latencies within groups for both forward and backward recall directions. As in Experiment 1, inter-response times were longer overall in the backward than forward direction. The scalloping that is considered evidence of subjective chunking was evident in both forward and backward directions, and it was, again, more pronounced in the backward direction, and in the backward Control condition. It may be



Effect	Estimate (SE)	
<b>a Analysis using Serial Position (BIC=75195)</b>		
Main effects		
Intercept	0.78 (0.10)*	<.001
Grouping	0.20 (0.14)	0.163
Direction	-2.70 (0.057)*	<.001
Serial Position	-0.28 (0.0070)*	<.001
Interactions		
Grouping×Direction	0.89 (0.076)*	<.001
Grouping×Serial Position	0.086 (0.0095)*	<.001
Direction×Serial Position	0.45 (0.010)*	<.001
<b>Grouping × Direction × Serial Position</b>	<b>-0.18 (0.013)*</b>	<b>&lt;.001</b>
<b>b Analysis using Output Position (BIC=75174)</b>		
Main effects		
Intercept	0.79 (0.10)*	<.001
Grouping	0.19 (0.14)	0.164
Direction	-1.07 (0.036)*	<.001
Output Position	-0.28 (0.0060)*	<.001
Interactions		
Grouping×Output Position	0.088 (0.0067)*	<.001
Direction×Output Position	0.12 (0.0066)*	<.001

$\Delta BIC=21$ , favouring the model with Output Position

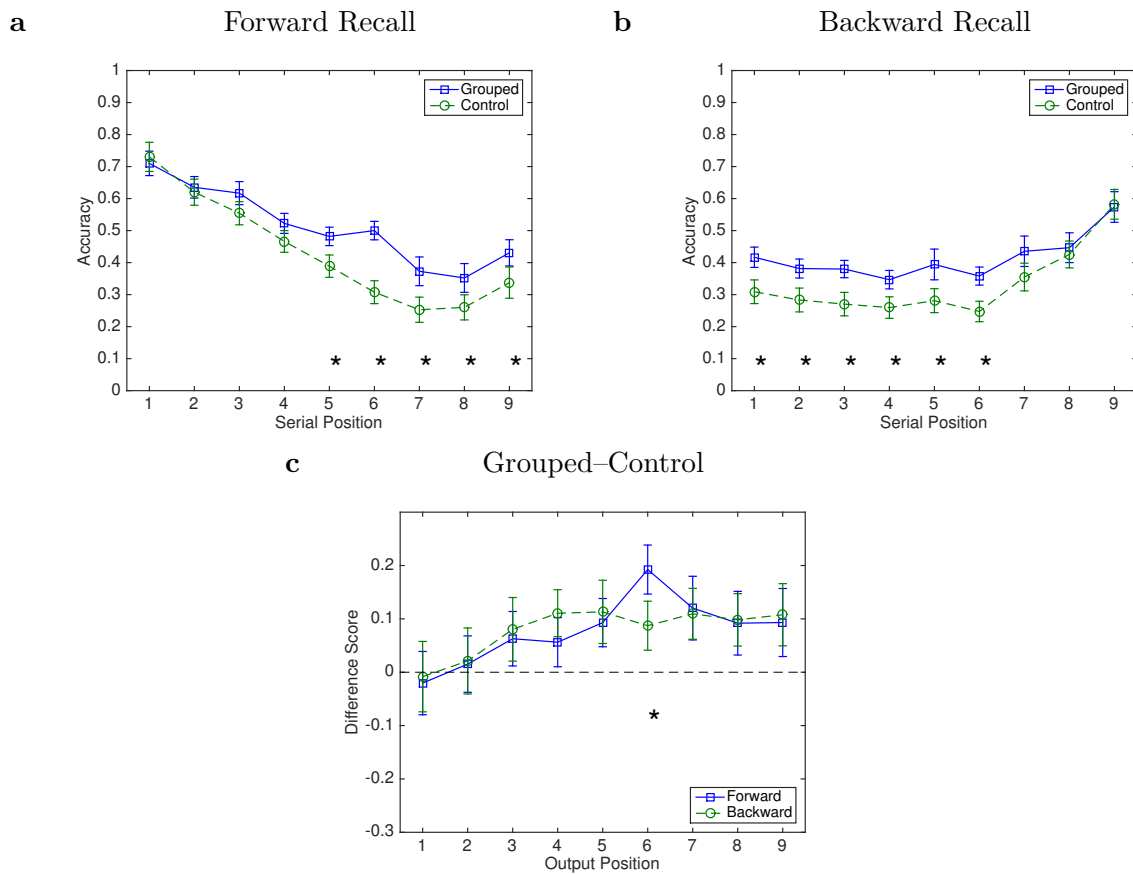
Table 5

The best-fitting LME model for recall accuracy (proportion correct) with Serial Position as a factor (a) or with Output Position as a factor (b) for Experiment 2. Estimate - the corresponding regression coefficient, along with its SE (standard error of the mean). BIC favours the model with Output Position; note that (a) includes a three-way interaction (highlighted in boldface) that is absent from (b). \* $p < 0.05$ .

Main effects	$\chi^2$	p
Intercept		
Grouping	0.55	0.458
Direction	895.31*	<.001
Output Position	3641.12*	<.001
Interactions		
Group×Direction	8.51	0.004
Direction×Output Position	18.46*	<.001
Grouping×Output Position	0.27*	0.604
Grouping×Direction×Output Position	19.86*	<.001

Table 6

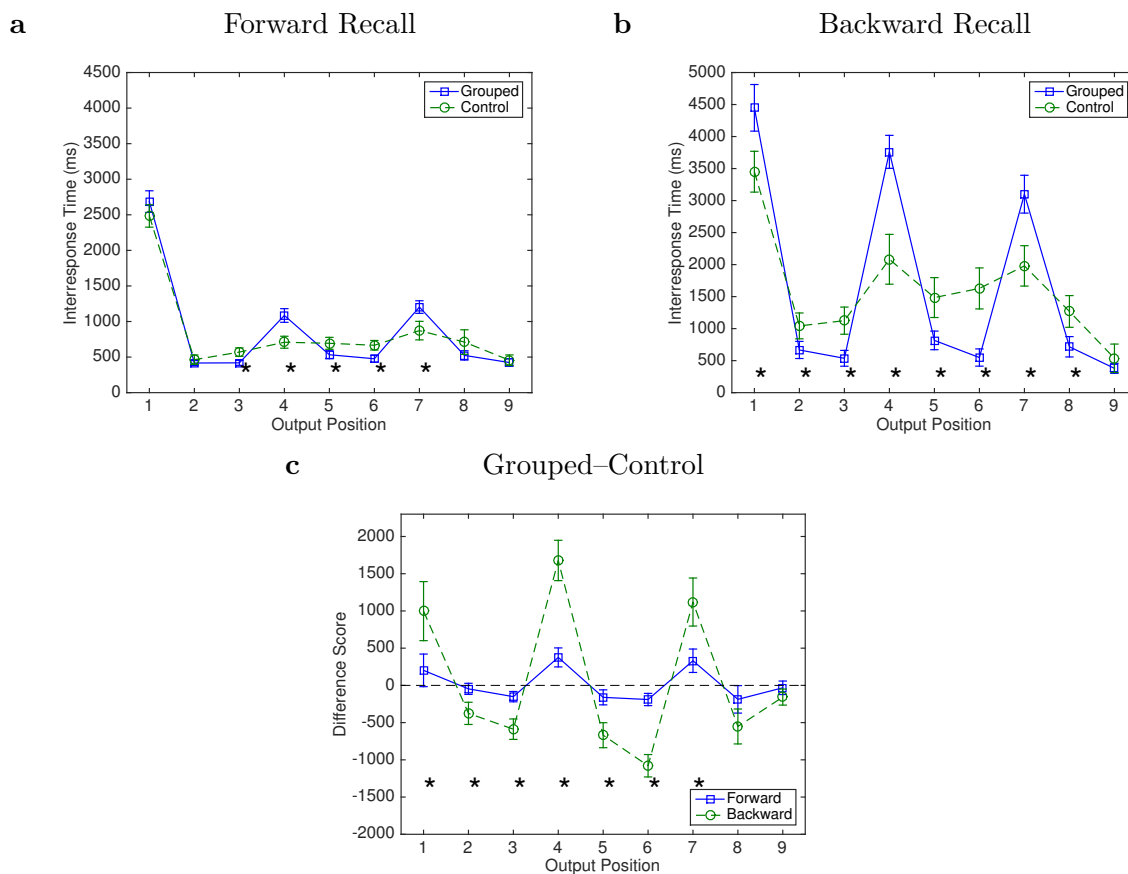
The best-fitting LME model for log inter-response time with Output Position as a factor for Experiment 2. The  $\chi^2$  column reports  $\chi^2$  from Wald test. \*  $p < 0.05$ .



*Figure 5.* Recall accuracy as a function of serial position for forward recall (panel a) and backward recall (panel b) for Experiment 2. Accuracy difference of Grouped-Control as a function of output position for forward and backward recall (panel c). Error bars denote 95% confidence intervals based on standard error of the mean. \*  $p < 0.05$ , based on a two-tailed  $t$  test.

worth noting that the inter-response times were slower than in Experiment 1 (Figure 1), perhaps due to a form of task-switching cost.

**Recall-direction mismatches.** We were concerned that manipulating recall direction within subjects might have confused participants, making it difficult for them to perfectly comply, and sometimes recalling a list in the wrong direction. To check the potential impact of such recall direction mismatches, we implemented a conservative recall directionality measure: the sum of differences between the serial positions of all adjacent recalls for a list. A positive value was considered evidence of an attempt to recall forward,



*Figure 6.* Inter-response time as a function of output position for forward recall (panel a) and backward recall (panel b) for Experiment 2. Inter-response time difference of Grouped–Control as a function of output position for forward and backward recall (panel c). Error bars denote 95% confidence intervals based on standard error of the mean. Significant differences between Grouped and Control condition are denoted by \*  $p < 0.05$ , based on a two-tailed  $t$  test.

and a negative value, an attempt to recall backward. Across all lists, 18.08% of backward recall trials were suspected possible wrong directions (recalled forward) and 8.08% of forward recall lists were suspected possible wrong directions (recalled backward). When we replotted the serial-position and output-position curves with those (conservatively) suspected wrong-direction lists removed, this did not change the major features of the plots (Figure S1 and S2 in supplementary materials).

**Within-list intrusions.** We conducted within-list intrusion analysis following the same methods in Experiment 1, and replicated those findings. For forward serial recall

Main effects	$\chi^2$	
Intercept		
Grouping	0.0014	0.970
Direction	382.65*	<.001
Error Type	89.06*	<.001
Interactions		
Error Type $\times$ Direction	6.90*	0.009
Error Type $\times$ Grouping	90.97*	<.001
Direction $\times$ Grouping	12.64*	<.001
Direction $\times$ Error Type $\times$ Grouping	25.31*	<.001

Table 7

*The best-fitting LME model for error proportion scores (proportion of errors normalized by within-list errors) with Direction Grouping and Error types as factors, for Experiment 2. The  $\chi^2$  column reports  $\chi^2$  from Wald test. \* $p < 0.05$*

(Figure 4b and Table 4b), grouping reduced adjacent transposition errors and increased interposition errors, but for backward recall, grouping did not produce an increase in the rate of interpositions (Figure 4b), supported by a significant Error Type  $\times$  Direction  $\times$  Grouping interaction in the best fitting LME model (Table 7). In line with Experiment 1, the correlation between proportion of interpositions and recall accuracy was either non-significant or even (nominally) negative for the Grouped conditions (Forward:  $r(50) = -0.076$ ,  $p = 0.59$ ; Backward:  $r(50) = -0.58$ ,  $p < 0.001$ ). The correlation was positive, although small and non-significant for the Forward Control condition ( $r(39) = 0.20$ ,  $p = 0.20$ ), was significant but negatively correlated for the Backward Control condition ( $r(39) = -0.50$ ,  $p < 0.001$ ). As with Experiment 1, these results are hard to reconcile with the hypothesis that committing interpositions is related to the mechanism by which temporal grouping enhances serial recall accuracy.

### Discussion of Experiments

In sum, inter-response time data suggest participants did appear to subjectively chunk temporally grouped lists into three groups of three items, in the backward as well as forward direction. Grouping improved accuracy for backward as well as forward recall. However, apparently inconsistent with the standard positional-coding account, the benefit to accuracy was not a function of serial position, but rather, of output position. Interpositions were ob-

served, but they were not more prevalent in the temporally grouped condition for backward recall. Interpositions were not reliably associated with better accuracy in Grouped conditions, so their relationship to the effects of temporal grouping on accuracy and inter-response times is not clear. Finally, these key characteristics were found both when recall-direction was known to participants prior to studying a list (Experiment 1) and when recall-direction was not instructed until after each list had been studied (Experiment 2). This challenges the idea that in Experiment 1, participants in the backward recall-direction were simply reverse-coding list-position or within-group position. It also provides a clue that the mechanism by which temporal grouping enhances serial-recall accuracy may be located at the recall, rather than the study phase.

## Models

The results of both experiments suggested the advantage due to temporal grouping depends on output position, not serial position. At first blush, the empirical pattern seems at odds with current positional-coding accounts of temporal grouping effects. However, positional-coding models do have ways in which phenomena could respond to output position. For example, in SIMPLE, as recall proceeds, study-test intervals for all items lengthen. Because SIMPLE acts on the logarithm of the study-test interval,<sup>4</sup> item-discriminability worsens at later output positions. Second, response suppression (the tendency of participants to avoid repeating items during recall; Duncan & Lewandowsky, 2005) enables accuracy of one recall to influence probability of later recalls. We first checked whether the current formulation of SIMPLE already possesses the ability to produce output-order effects of temporal grouping, by fitting SIMPLE to the data from Experiment 1. There may be good intuitive reasons to expect this model would be insufficient. However, we felt

---

<sup>4</sup>As discussed by Brown et al. (2007), findings of absent isolation effects in forward serial recall (e.g., Lewandowsky & Brown, 2005; Lewandowsky, Brown, Wright, & Nimmo, 2006) lead to the idea that participants virtually ignore the temporal dimension during serial recall, in the same paragraph (p. 562), they nonetheless explain that they retain their focus on the temporal dimension. Indeed, several of the serial-recall phenomena they fit seem to rely upon (temporal) recency and relative time differences, not just ordinal position. We adopted their assumptions, having adapted their MATLAB model code, and acknowledge that Brown and colleagues' formulation of SIMPLE, as well as our own, could be fairly challenged or modified on this basis.

it important to give this model a fair chance to fit the data for several reasons. First, the principle of parsimony entails that one should not jump to a more complicated account when a simpler account is sufficient (for known data). Second, even in this basic version of SIMPLE, recall probability is determined by temporal discriminability. Thus, the temporal pattern of presentation might be sufficient to produce the kind of advantage due to temporal grouping we observed. Third, in addition to temporally patterned presentation, participants' own responses were temporally patterned. We found it intriguing to consider the possibility that temporally patterned responses might have some separate cause (e.g., participants feeling compelled to mimic the rhythm of presentation), but could then have a spillover effect, causing temporal discriminability to evolve over the course of recall, in a characteristic way that could, itself, be the cause of the effects of temporal grouping on probability of recall.

Next, we added within-group position cueing to the model to test whether Brown et al.'s (2007) two-dimensional cueing account could be sufficient to explain both the forward and backward serial recall data. Finally, we incorporated response suppression into both the two-dimensional and one-dimensional models, to test whether response suppression could be critical to understanding the precise pattern of effects of temporal grouping on recall probability.

**Modelling methods.** The model was adapted from the MATLAB code used to simulate temporal grouping effects by Brown et al. (2007). SIMPLE starts with the assumption that memory is driven by discriminability of presentation times of items (Brown et al., 2007), relative to time of test. Thus, psychological time,  $M_i = \log(\text{test} - \text{study time})$ . Similarity,  $\eta_{i,j}$ , between the psychological times of two items,  $i$  and  $j$ , is:

$$\eta_{i,j} = e^{-c|M_i - M_j|}, \quad (1)$$

where  $c$  parameterizes the level of confusibility of presentation times and  $|\cdot|$  denotes absolute value (magnitude). To accommodate temporal grouping in the second model variant, we

follow Brown et al. (2007) and assume a second dimension of discriminability, due to position within-group. Each item thus has one within-group position value; as a function of serial position, these values,  $P_i$ , are 3, 2, 1, 3, 2, 1, 3, 2, 1 (inverted just as time is relative to time of test). Again, following Brown et al. (2007), we assume that confusibility is a function of the L1 (taxi-cab metric) distance between the two-dimensional (time, within-group position) values:

$$\eta_{i,j} = e^{-c(w_t|M_i-M_j|+w_p|P_i-P_j|)}, \quad (2)$$

where  $w_t$  and  $w_p$  are weighting parameters to vary the relative reliance on time versus within-group position, respectively, and  $w_t + w_p = 1$ . Because spontaneous grouping is often observed even when lists are presented ungrouped (e.g., Martin & Noreen, 1974; S. Madigan, 1980; Jou, 2011),  $w_p$  is allowed to vary between grouped and control lists, with the expectation that  $w_p$  will not necessarily be zero for Control participants, but will at least be greater for Grouped than Control conditions.

During recall, the model probes successively in an attempt to retrieve the first item, then the second, etc. Thus, for forward recall, the model probes with the time (recency) of serial position 1, then with the *updated* recency of serial position 2, proceeding through serial position  $L$ . Following Bireta et al. (2010), backward recall starts by probing with the recency of serial position  $L$ , then with the *updated* serial position  $L - 1$ , down through serial position 1. The probability of recalling item  $j$  in position  $i$  is based on the confusibility between  $i$  and  $j$ . First, the discriminability of item  $j$ , given position  $i$  as the cue, is:

$$D(j|i) = \frac{\eta_{i,j}}{\sum_{k=1}^L \eta_{i,k}}, \quad (3)$$

These  $D$  values are passed through a sigmoid function to produce recall probabilities:

$$P(j|i) = \frac{1}{1 + \exp[-s(D(j|i) - \theta)]}, \quad (4)$$

where  $s$  and  $\theta$  determine the slope and threshold of the sigmoid, respectively. In case the sum of probabilities for a given recall position exceed 1, which is logically impossible, the probabilities are renormalized by that sum to fix the total recall probability over all list-items to 1. We thus tacitly assume that if probabilities do not sum to 1, the remaining instances are omissions (but might also be thought of as extra-experimental intrusions).

SIMPLE was developed to explain accuracy data, but not yet to address latency data. However, because the model includes a temporal dimension, timing of both list presentation and responses can critically influence model output. Brown et al. (2007) did not have inter-response time values for the grouping data they fit (Hitch et al., 1996); in lieu of such data, they assumed that response times increased exponentially,  $n^{1.5}$  s for response  $n$ . Because of the primary importance of inter-response times in determining temporal discriminability in SIMPLE, we fit SIMPLE using the mean inter-response times observed in our data set. These inter-response times were not exactly those plotted in Figures 3 and 6, because those plot inter-response times for correct recalls only; for model-fitting, we used the total time from study to time of recall, regardless of accuracy of the prior responses. List length and presentation times were the same in the model as in the experiment.

***Response Suppression.*** Finally, although SIMPLE, like other positional-coding models, assumes independent-cueing, there is one more way in which recall can change systematically over the course of the recall sequence. Because participants rarely repeat an item during serial recall (e.g., Duncan & Lewandowsky, 2005; Henson, 1998; Vousden & Brown, 1998), many models have incorporated a mechanism for *response suppression*, making it impossible or unlikely to recall an item that has already been produced as a response (e.g., Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998; Lewandowsky & Murdock, 1989). Thus, as recall progresses, the size of the response set will diminish. This suggests that if accuracy is increased early in recall, those correct recalls will, helpfully, reduce the response set available for subsequent recalls, making it easier to produce accurate responses later in the sequence. Thus, an early benefit will not only persist, but could accumulate over the course of recalls. We added response suppression to the model as



follows.<sup>5</sup> 1) For each recall, compute the probability of each list item not having yet been recalled. That is the product of  $1 - p$  up to the current recall operation. 2) In the denominator of the discriminability equation: for each list item within the summation, multiply by the probability that this item was not yet recalled. 3) Calculate probabilities as usual (via the sigmoid function). 4) Multiply the final probabilities by the probability that each item has not yet been recalled. Note that this also avoids the need to check for probabilities exceeding 1.

**Model fitting.** SIMPLE was fit to each participant’s data individually, with the SIMPLEX algorithm (Nelder & Mead, 1965), using an adaptation of MATLAB’s `fminsearch` function and root-mean-squared-deviation (RMSD) as the measure of fitness. The ranges searched were designed to encompass and extend the ranges of parameters reported by Brown et al. (2007), Table 1:  $c = [0.45, 30]$ ,  $\theta = [0, 1]$ ,  $s = [1, 100]$  and the within-group position weights were searched over the allowable range  $[0, 0.99]$ . For a given model, we fit the model to each participant 100 times and report the best-fitting model across those 100 runs, summarized across participants. Because BIC, log-likelihood and RMSD are all monotonically related to one another, when the number of free parameters is held constant, optimizing for RMSD produces the same result as optimizing for log-likelihood or BIC. BIC values were, in turn, computed via an estimation of log-likelihood, following Burnham and Anderson (2004). Bear in mind that our purpose here is not to find the “correct” model of immediate serial recall and temporal grouping effects. Rather, the model fits will indicate whether our findings do or do not pose a challenge to a model that incorporates that standard positional-cueing account of temporal grouping. The parameter searches give the model the opportunity to provide its best quantitative fit, but the primary way we evaluate the model is by checking whether it produces the key qualitative effects found in the data; most notably: an increase in the benefit of temporal grouping over the course of the recall sequence for both forward and backward recall directions.

### Modelling results.

---

<sup>5</sup>To our knowledge, published versions of SIMPLE have never incorporated response suppression, although Brown et al. (2007) commented that they had done so.

***SIMPLE with a single dimension.*** At its heart, SIMPLE explains accuracy in serial recall as primarily dependent upon discriminability of study–test intervals. We noted that the pattern of study–test intervals differs for grouped lists, and moreover, study times (inter-response times) are temporally patterned. This led us to propose that the absolutely simplest plausible explanation would be that the effects of temporal grouping on accuracy are a direct consequence of the temporal patterning of study and recall times. We thus first fit a 1-D version of SIMPLE, where the only dimension was (study–test) time.

Plotted in Figure 7, although the quantitative fit is better for backward than forward recall, it is a bit worse for Grouped than Control data. More concerning, the model underpredicts the amount of scalloping of the Grouped data, particularly in the forward direction, and underestimates primacy while overestimating recency accuracy in the forward direction. Finally, the model underpredicts the difference between Grouped and Control accuracy late in recall.

***SIMPLE with a second dimension coding within-group order.*** Because the 1-D model failed to produce scalloping of accuracy, we next added within-group position as a cue, following Brown et al. (2007). Plotted in Figure 8, it is clear that this model fit quantitatively better for the grouped conditions (and interestingly, nearly no change in BIC values for the Control conditions). This model produced more scalloping, particularly in the Grouped conditions. The model also succeeded fairly well in capturing the increase in advantage for grouping with increasing output position. Where the model fell short was, again, in underestimating early-list accuracy and overestimating late-list accuracy for the forward recall direction. In addition, the model produced nearly-null effects of grouping at the initial two response positions, but not quite as equivalent accuracy levels as in the data.

***Response Suppression.*** Next, we added response suppression to the 2-D model (see Modelling Methods). Note that our implementation adds no free parameters to the model. That is because we assume response suppression is complete, and does not relax over the course of the recall period, as shown by Duncan and Lewandowsky (2005).

As is evident in Figure 9, response suppression, combined with 2-D cueing, produced

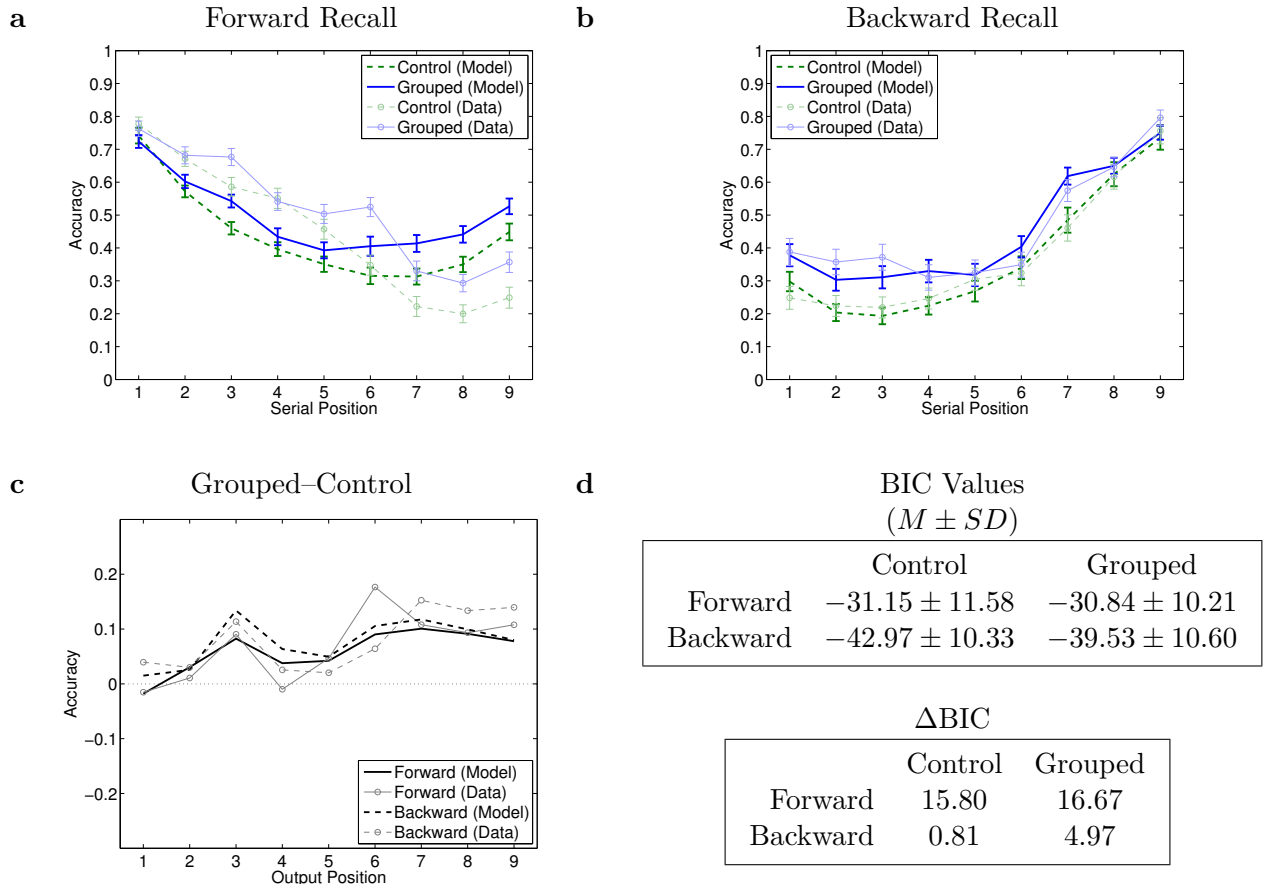


Figure 7. Best-fitting parameter sets for the One-dimensional model, fit to individual participants. Recall accuracy as a function of serial position for forward recall (a) and backward recall (b) and for the Grouped-Control difference (c). Error bars plot standard error of the mean across participants. (d) BIC values (upper table) and mean  $\Delta$ BIC compared to the best model (lower table). In panels a-c, the data from Experiment 1 are plotted in lighter shades.

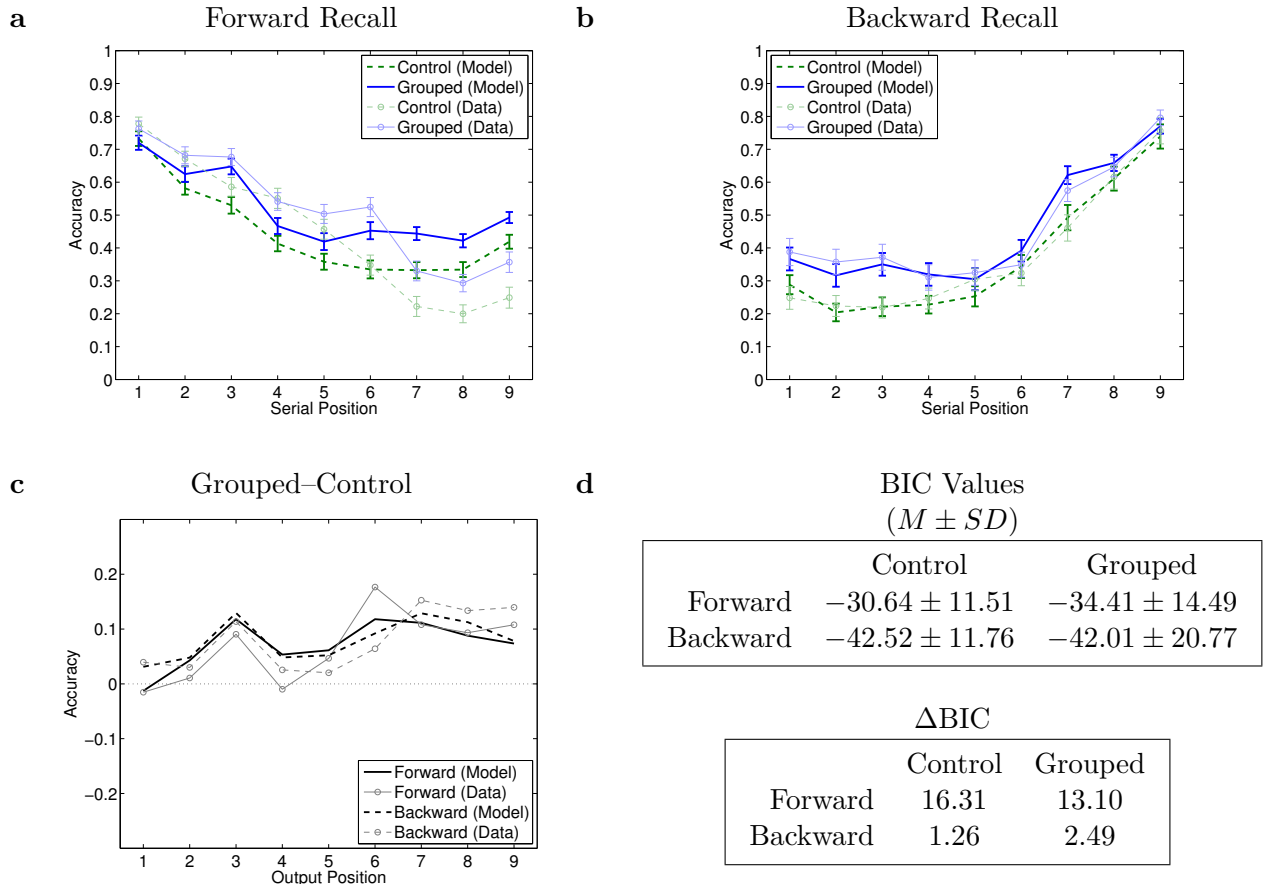


Figure 8. Best-fitting parameter sets for the Two-dimensional model, fit to individual participants. Recall accuracy as a function of serial position for forward recall (a) and backward recall (b) and for the Grouped-Control difference (c). Error bars plot standard error of the mean across participants. (d) BIC values (upper table) and mean  $\Delta$ BIC compared to the best model (lower table). In panels a-c, the data from Experiment 1 are plotted in lighter shades.

better BIC values than the previous two models in all conditions, and with a mean  $\Delta$  BIC  $> 2$  in all conditions except Backward Control. Panels a–c show that the model behaviour approached the data quite well, capturing all the noteworthy qualitative features of the data, including near-null effects of grouping at output positions 1, 2, 4 and 5, and an overall trend of increasing benefit due to grouping with increasing output position. In sum, this model produced the lowest BIC values and the best qualitative fits.

To follow up on the excellent fit of this model, we next examine the distributions of best-fitting parameter values, which are plotted in Figure 10. First,  $c$ , which controls distinctiveness along the temporal dimension, fit to higher values for Grouped than Control for Forward recall participants, but the reverse for the Backward recall participants. On the whole, the model estimated more temporal distinctiveness for Backward than Forward-recall participants. However, this parameter also traded off with the sigmoid parameters, so there may be some dependencies amongst these three parameters. Interestingly, the model fit higher values of the within-chunk cueing parameter,  $w_p$ , for Grouped than Control participants, and this was the case for both the Forward and Backward recall data. Moreover, in each direction, there was nearly stochastic dominance; at nearly every percentile the value of  $w_p$  was greater for the grouped participant than the corresponding control participant. This is consistent with the idea that within-chunk cueing explains the difference between Grouped and Control data.

Next, we asked whether the rate of interposition intrusions (a misplaced list item that is, nonetheless, in the “correct” position within the wrong group) might indeed reflect the action of within-chunk cueing, as was suggested by Henson (1998). We correlated the value of  $w_p$  with the proportion of errors that were interpositions, across participants within each of the four groups. The only significant correlation was for the Backward Control participants (Forward Control:  $r(40) = -0.18$ ,  $p = 0.27$ , Forward Grouped:  $r(39) = -0.15$ ,  $p = 0.35$ ; Backward Control:  $r(40) = 0.60$ ,  $p < 0.001$ ; Backward Grouped:  $r(41) = -0.04$ ,  $p = 0.78$ ). To account for a participant’s overall tendency to produce intrusion responses, we also correlated  $w_p$  with the ratio of interpositions to adjacent transpositions,

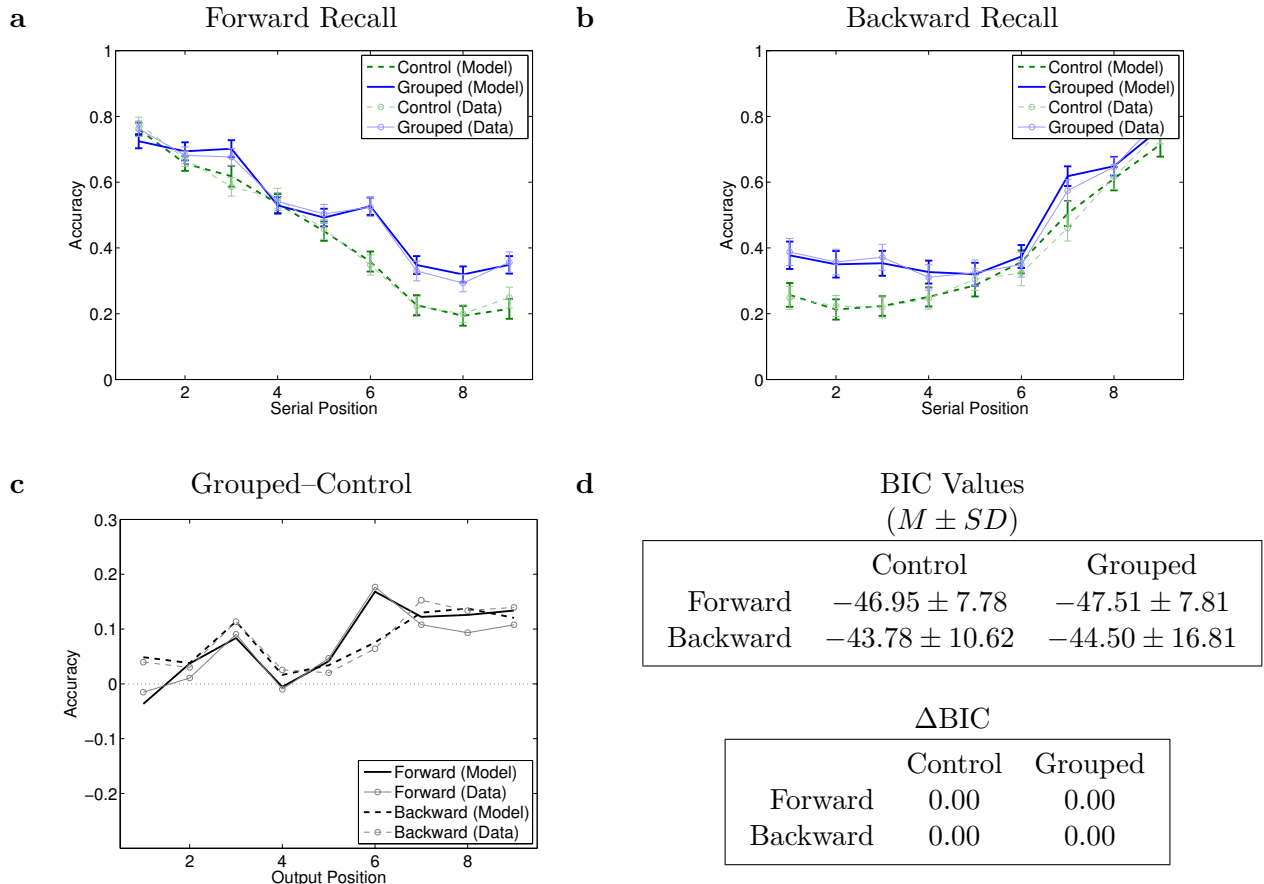


Figure 9. Best-fitting parameter sets for the Two-dimensional with Response Suppression model, fit to individual participants. Recall accuracy as a function of serial position for forward recall (a) and backward recall (b) and for the Grouped-Control difference (c). Error bars plot standard error of the mean across participants. (d) BIC values (upper table) and mean  $\Delta$ BIC compared to the best model (lower table). In panels a-c, the data from Experiment 1 are plotted in lighter shades.

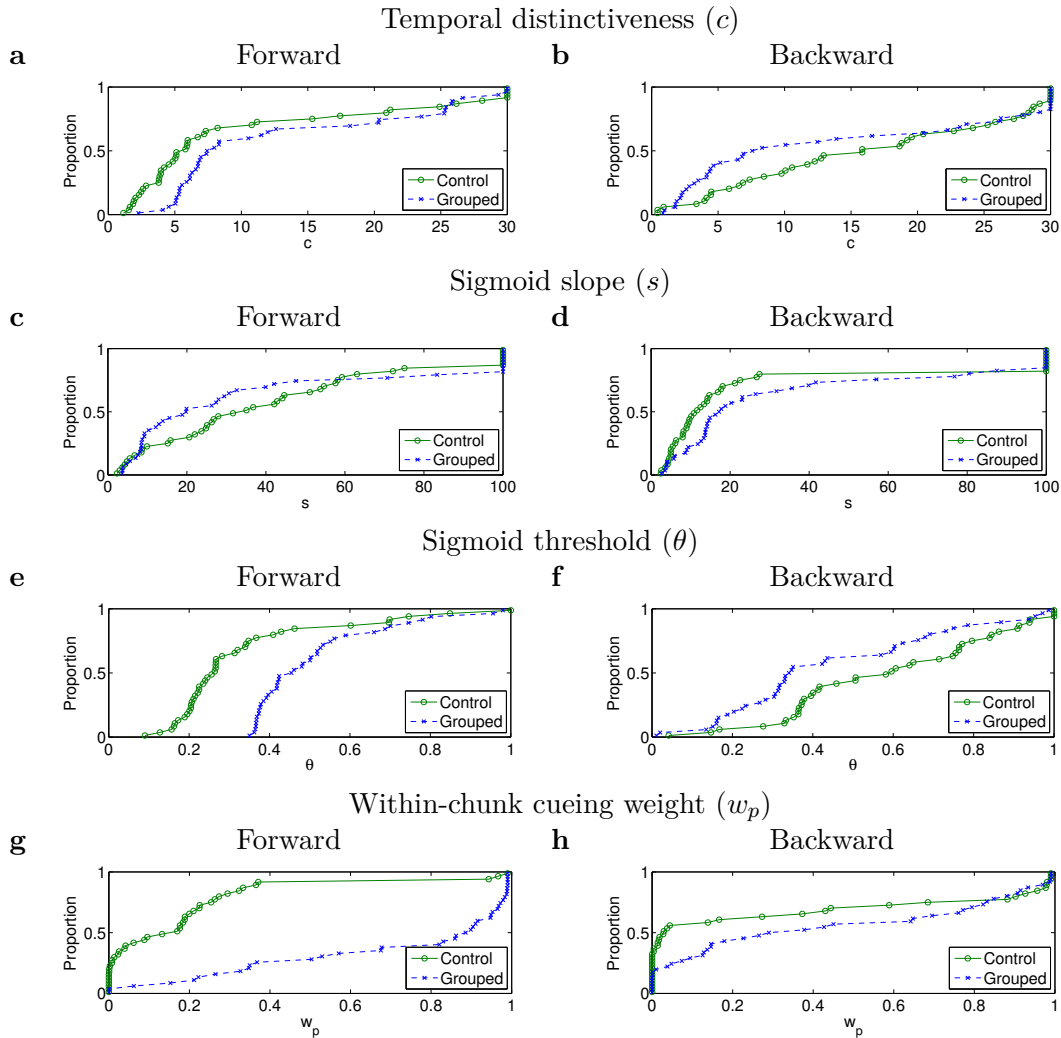


Figure 10. Cumulative distribution functions of best-fitting parameters for the 2-D model with response suppression.

but the pattern was unchanged ( $r(40) = -0.13$ ,  $p = 0.40$ ;  $r(39) = -0.14$ ,  $p = 0.37$ ,  $r(40) = 0.450$ ,  $p = 0.003$  and  $r(41) = 0.012$ ,  $p = 0.940$ , respectively). Although this does not rule out the idea that interpositions result from cueing with within-chunk position, there is not a clear one-to-one relationship between the relative weighting of within-chunk to temporal cueing and the tendency to make interposition errors.

**Response Suppression and the 1-D model.** Finally, we wanted to check whether response suppression, alone, might be largely responsible for the strong fit to the data—namely, with the within-group cueing mechanism removed. When we added

response suppression to the one-dimensional model (Figure 11), BIC values for the Control conditions were only slightly worse, but BIC values were far worse for the Grouped conditions. Qualitatively, the shapes of the output-position effects were fairly similar to the data, deviating notably by too much of an advantage for Grouped participants at output position 3 (serial position 7) and too little at output position 7 (serial position 3). For the Forward conditions, the fit was quite good for Control participants, but completely missed the scalloping effect for Forward, Grouped participants, even with an inverted U shape for the first group (positions 1–3). Thus, even with the help of response suppression to produce effects dominated by output order, relying on temporal patterning during presentation and during recall could not adequately fit the Grouped data.

### General Discussion

Our novel condition, backward serial recall of temporally grouped lists, produced the characteristic scalloping effect, in both accuracy and inter-response time measures, extending this type of finding to backward recall. Importantly, the backward recall condition revealed that the benefit of grouping was driven by output-order rather than input-order. Foreknowledge (in Experiment 1, compared to Experiment 2) did not change this characteristic of the result (resembling the effect of word length reported by Surprenant et al., 2011 but differing from the effect of spatial manual tapping reporting by Guitard et al., 2019). This provided a clue that the most likely locus of the advantage of temporal grouping is during recall rather than during study.

The overall dominance of output position on recall accuracy and near-mirror symmetry of forward and backward serial recall seemed, at first, to challenge positional-coding models, and in particular, the two-dimensional positional cueing account of temporal grouping effects. Indeed, our first models, based on SIMPLE, were severely inadequate accounts of the data, particularly of the Grouped conditions. When we added response suppression, the model with both temporal and within-chunk-position cueing fit the data quite well, both quantitatively and qualitatively, and without adding any additional free parameter. Thus,



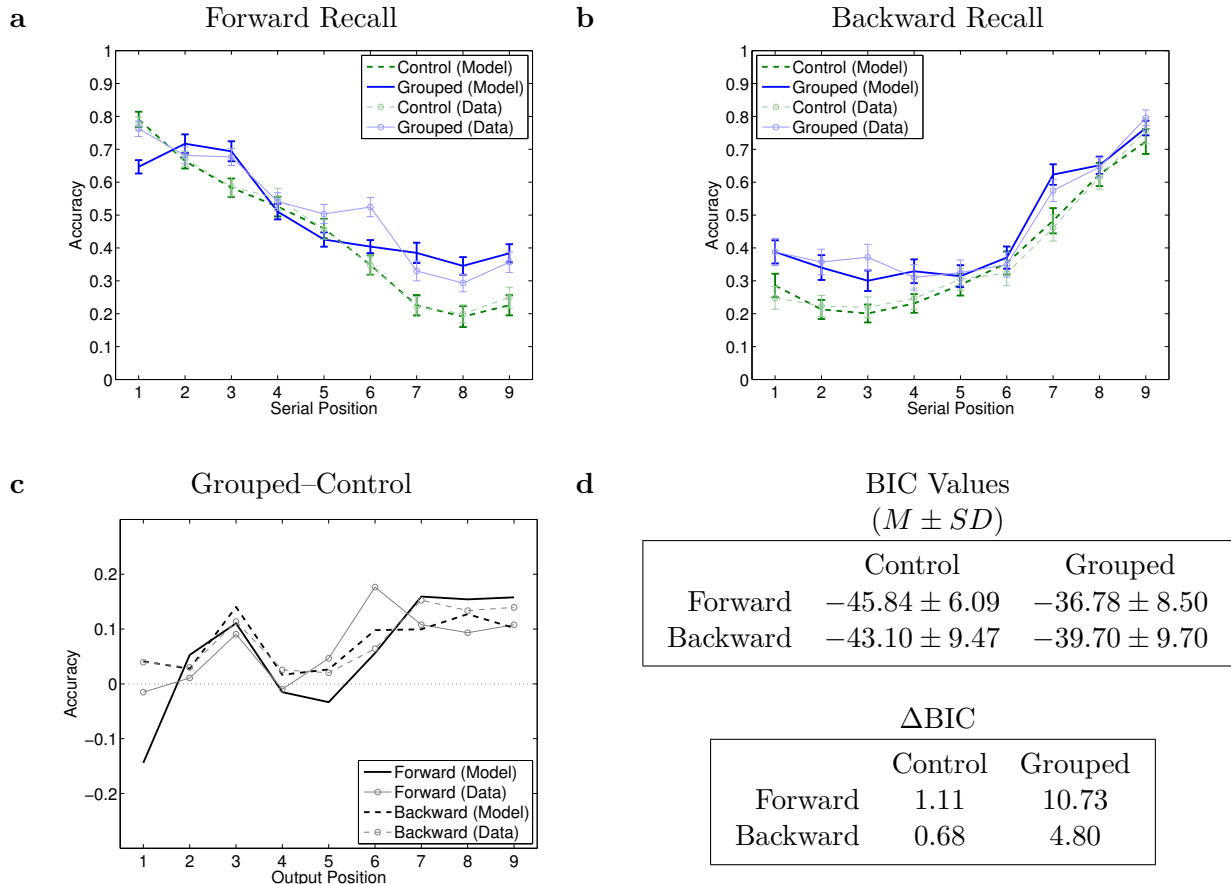


Figure 11. Best-fitting parameter sets for the One-dimensional with Response Suppression model, fit to individual participants. Recall accuracy as a function of serial position for forward recall (a) and backward recall (b) and for the Grouped-Control difference (c). Error bars plot standard error of the mean across participants. (d) BIC values (upper table) and mean  $\Delta$ BIC compared to the best model (lower table). In panels a-c, the data from Experiment 1 are plotted in lighter shades.

the dominance of output position appears compatible with the idea of temporal and ordinal cueing, as long as response suppression is included. It would be interesting to investigate whether response suppression might play a similar role, enabling benefits to accumulate over the course of recall in other paradigms.

The presence of interpositions has been viewed as evidence of cueing with within-group position. Although we cannot rule this out (a risk here is that if there is a fairly restricted range across participants, the correlations will be weak even if the parameter is strongly related to the phenomenon), interposition rates in our data appear not to be directly related to within-group cueing (the  $w_p$  parameter in the model), nor to accuracy overall.

Although the fit of the two-dimensional version of simple with response suppression is remarkably close to the data (Figure 9), the success of this model variant does not directly rule out other model accounts. Although it is beyond the scope of this manuscript, it would be interesting to fit other positional/ordinal cueing models to the serial/output-position effects presented here (e.g., Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998), and to determine whether response suppression is a critical ingredient for those models as well. Associative chaining models were argued to be ruled out for immediate serial recall of relatively short lists (e.g., Henson, 1998), but many of the early arguments against associative chaining have been challenged (e.g., Caplan, 2015; Caplan, Madan, & Bedwell, 2015; Farrell, 2006; Hulme, Stuart, Brown, & Morin, 2003; Kahana, Mollison, & Addis, 2010; Serra & Nairne, 2000; Solway, Murdock, & Kahana, 2012), suggesting that a re-consideration of the evidence against chaining models may be in order. Such a chaining model would need to be updated, to include symmetric associations and remote associations, for example. An associative chaining account of our findings might be plausible, although this would entail a large amount of work, as no update to associative chaining models has been tested against a comprehensive set of benchmark findings since Lewandowsky and Murdock (1989). One well known challenge to an associative chaining model would be to produce within-list intrusion phenomena that suggest positional-cueing; namely, interpositions and protrusions,

which are intrusions from prior lists that are recalled in the same position as their original presentation position, more often than expected by chance (Fischer-Baum & McCloskey, 2015; Henson, 1998, 1999; Osth & Dennis, 2015). While these phenomena should not be overlooked, it is not yet clear how they should be accommodated in any model. First, protrusions (and prior-list intrusions) are rare, as these studies have shown. Second, the greater rate of interpositions for grouped versus control lists replicated for forward but not for backward serial recall, even in Experiment 2, suggesting that even a positional cue might fail to capture the interaction of interposition rate with recall direction without modification. Finally, as has been reasoned, hybrid models should be considered (Caplan, 2015), incorporating both associative and positional or ordinal cueing. It could be the case that the benefit of temporal grouping to serial recall accuracy derives from one mechanism (e.g., associative cueing) while protrusions and the elevated interposition rate in the forward, but not backward, direction, are due to the contribution of positional/ordinal cueing. Hierarchical models, wherein chunks are stored distinctly from isolated list items, might be able to accommodate our findings. One such model, by Farrell (2012), may require some modification; currently, it includes the assumption that chunks are retrieved in forward order. This would seem contradicted by the striking symmetry of our effects of temporal grouping on forward and backward recall directions, but this remains to be tested directly.

There is an abundance of research that has suggested participants can re-code lists into smaller units, often called “chunks,” where a chunk is assumed to be retrieved in an all-or-none step (e.g., Anderson & Matessa, 1997; Anderson et al., 1998; Farrell, 2012; Johnson, 1969, 1972; Lee & Estes, 1981; Martin & Noreen, 1974; Miller, 1956; Murdock, 1995, 1997). Our data do not speak directly to this possibility. We note that the experimental paradigms that have led to this conception of chunking are quite different, and are typically transfer designs, either based on pre-experimental knowledge or two-list transfer paradigms (e.g., Chekaf, Cowan, & Mathy, 2016; Johnson, 1969; Thalmann, Souza, & Oberauer, 2019) rather than temporal grouping. The possibility that temporal grouping effects might be related to those transfer effects is intriguing, but would also be best investigated directly in

future studies.

In recent years, there has been an accelerating interest in comparing backward serial recall with forward serial recall. Results seem split as to whether a given experimental manipulation is consistent between recall directions (e.g., Bireta et al., 2010; Guérard et al., 2012; Guérard & Saint-Aubin, 2012; Guitard et al., 2019; Manning & Pacifici, 1983) or dissociates recall directions (e.g., Baker, Tehan, & Tehan, 2012; Olivia, Saint-Aubin, Guérard, & Pâquet, 2017; Guitard et al., 2019; Li & Lewandowsky, 1993, 1995; S. A. Madigan, 1971; Norris, Hall, & Gathercole, 2019; Richardson, 2007; Ritchie et al., 2015; Surprenant et al., 2011; Thomas et al., 2003). Our temporal-grouping manipulation would add to the former, suggesting a common mechanism, across recall direction, for the effect of temporal grouping on recall. However, this does not change the fact that backward serial recall differs from forward serial recall in other ways. Evidence from self-reported strategy usage, validated with accuracy and response-time serial-position effects, suggests that participants may use a mix of strategies to perform backward, as well as forward recall (Norris et al., 2019). Although our model fits found a single model to be sufficient to fit individual participants as a group, some fit better than others; thus, although overall, temporal grouping effects were well explained by a combination of explicit within-group position-cueing along with repetition suppression, it is possible that this could be modulated by particular strategies. A complete model would need to accommodate both commonalities and dissociations between recall directions.

In sum, both experiments revealed the effects of temporal grouping on immediate serial recall to be primarily functions of output position and not serial position. This presents a novel constraint on mathematical models of serial recall. However, with the addition of the well supported process of response suppression, a temporal-distinctiveness model, incorporating within-chunk position cueing, was able to fit the empirical pattern quite closely. In conclusion, the characteristic scalloping pattern observed in immediate serial recall of temporally grouped items appears to be driven by a combination of within-group cueing and the accumulation of benefits to accuracy due to response suppression.

### Acknowledgement

Portions of this work were reported in a dissertation presented by Yang S. Liu to the Department of Psychology, University of Alberta, 2015. Supported in part by the Natural Sciences and Research Council of Canada. The authors thank Madhawa Alahakoon for assistance conducting the second experiment.

### Open Practices Statement

The data and materials for all experiments are available at [https://osf.io/evmct/?view\\_only=be14d124209442f8882e476f812965b6](https://osf.io/evmct/?view_only=be14d124209442f8882e476f812965b6), and none of the experiments was preregistered.

### 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.
- Anderson, J. R., & Matessa, M. (1997). A production system theory of serial memory. *Psychological Review*, *104*(4), 728-748.
- Baayen, R. H. (2007). *LanguageR (R package on cran version 1.1)*. <http://cran.r-project.org/web/packages/languageR/index.html>.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12-28.
- Baker, R., Tehan, G., & Tehan, H. (2012). Word length and age influences on forward and backward immediate serial recall. *Memory & Cognition*, *40*(1), 40-51.
- Bates, D. M. (2005). Fitting linear mixed models in R. *R News*, *5*, 27-30.
- Bates, D. M., & Sarkar, D. (2007). *lme4: Linear mixed-effects models using s4 classes (version 0.999375-39)* [Computer software and manual]. <http://cran.r-project.org/web/packages/lme4/>.
- Bireta, T. J., Fry, S. E., Jalbert, A., Neath, I., & Surprenant, A. M. (2010). Backward recall and benchmark effects of working memory. *Memory & Cognition*, *38*(3), 279-291.

- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539-576.
- Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*(1), 127-181.
- Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*(3), 551-581.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research*, *33*(2), 261-304.
- Caplan, J. B. (2015). Order-memory and association-memory. *Canadian Journal of Experimental Psychology*, *69*(3), 221-232.
- Caplan, J. B., Madan, C. R., & Bedwell, D. J. (2015). Item-properties may influence item-item associations in serial recall. *Psychonomic Bulletin & Review*, *22*(2), 483-489.
- Chan, M., Ross, B., Earle, G., & Caplan, J. B. (2009). Precise instructions determine participants' memory search strategy in judgments of relative order in short lists. *Psychonomic Bulletin & Review*, *16*(5), 945-951.
- Chekaf, M., Cowan, N., & Mathy, F. (2016). Chunk formation in immediate memory and how it relates to data compression. *Cognition*, *155*, 96-107.
- Duncan, M., & Lewandowsky, S. (2005). The time course of response suppression: no evidence for a gradual release from inhibition. *Memory*, *13*(3/4), 236-246.
- Ebbinghaus, H. (1885/1913). *Memory: A contribution to experimental psychology*. New York: Teachers College, Columbia University.
- Farrand, P., & Jones, D. (1996). Direction of report in spatial and verbal serial short-term memory. *Quarterly Journal of Experimental Psychology*, *49A*(1), 140-158.
- Farrell, S. (2006). Mixed-list phonological similarity effects in delayed serial recall. *Journal of Memory and Language*, *55*, 587-600.
- Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review*, *119*(2), 223-271.
- Fischer-Baum, S., & McCloskey, M. (2015). Representation of item position in immediate serial recall: evidence from intrusion errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (Second ed.). Thousand Oaks CA: Sage. Retrieved from <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>

- Geller, A. S., Schleifer, I. K., Sederberg, P. B., Jacobs, J., & Kahana, M. J. (2007). PyEPL: a cross-platform experiment-programming library. *Behavior Research Methods*, *39*(4), 950-958.
- Guérard, K., & Saint-Aubin, J. (2012). Assessing the effect of lexical variables in backward recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(2), 312-324.
- Guérard, K., Saint-Aubin, J., Burns, S. C., & Chamberland, C. (2012). Revisiting backward recall and benchmark memory effects: a reply to Bireta et al. (2010). *Memory & Cognition*, *40*, 388-407.
- Guitard, D., Saint-Aubin, J., Poirier, M., Miller, L. M., & Tolan, A. (2019). Forward and backward recall: different visuospatial processes when you know what's coming. *Memory & Cognition*, *48*(5), 111-126.
- Haberlandt, K., Lawrence, H., Krohn, T., Bower, K., & Thomas, J. G. (2005). Pauses and durations exhibit a serial position effect. *Psychonomic Bulletin & Review*, *12*(1), 152-158.
- Henson, R. N. A. (1998). Short-term memory for serial order: the Start-End Model. *Cognitive Psychology*, *36*(2), 73-137.
- Henson, R. N. A. (1999). Positional information in short-term memory: relative or absolute? *Memory & Cognition*, *27*(5), 915-927.
- 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*, *49A*(1), 116-139.
- Hulme, C., Stuart, G., Brown, G. D. A., & Morin, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: evidence for associative effects in serial recall. *Journal of Memory and Language*, *49*(4), 500-518.
- Johnson, N. F. (1969). Chunking: associative chaining versus coding. *Journal of Verbal Learning and Verbal Behavior*, *8*(6), 725-731.
- Johnson, N. F. (1972). Organization and the concept of a memory code. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (p. 125-159). Washington, DC, USA: Winston & Sons.
- Jou, J. (2011). Two paradigms of measuring serial-order memory: two different patterns of serial-position functions. *Psychological Research*, *75*(3), 202-213.
- Kahana, M. J., Mollison, M. V., & Addis, K. M. (2010). Positional cues in serial learning: the spin-list technique. *Memory & Cognition*, *38*(1), 92-101.
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral*

- mechanisms in behavior* (p. 112-146). New York, NY: John Wiley and Sons.
- Lee, C. L., & Estes, W. K. (1981). Item and order information in short-term memory: evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 7(3), 149-169.
- Lewandowsky, S., & Brown, G. D. A. (2005). Serial recall and presentation schedule: a micro-analysis of local distinctiveness. *Memory*, 13(3/4), 283-292.
- Lewandowsky, S., Brown, G. D. A., Wright, T., & Nimmo, L. M. (2006). Timeless memory: evidence against temporal distinctiveness models of short-term memory for serial order. *Journal of Memory and Language*, 54, 20-38.
- Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review*, 96(1), 25-57.
- Li, S.-C., & Lewandowsky, S. (1993). Intralist distractors and recall direction: constraints on models of memory for serial order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(4), 895-908.
- Li, S.-C., & Lewandowsky, S. (1995). Forward and backward recall: different retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 837-847.
- Liu, Y. S. (2015). *Human order memory: insights from the relative-order task* (Unpublished doctoral dissertation). University of Alberta.
- Liu, Y. S., Chan, M., & Caplan, J. B. (2014). Generality of a congruity effect in judgements of relative order. *Memory & Cognition*, 42(7), 1086-1105.
- Madigan, S. (1980). The serial position curve in immediate serial recall. *Bulletin of the Psychonomic Society*, 15(5), 335-338.
- Madigan, S. A. (1971). Modality and recall order interactions in short-term memory for serial order. *Journal of Experimental Psychology*, 87(2), 294-296.
- Manning, S. K., & Pacifici, C. (1983). The effects of a suffix-prefix on forward and backward serial recall. *American Journal of Psychology*, 96(1), 127-134.
- Martin, E., & Noreen, D. L. (1974). Serial learning: identification of subjective subsequences. *Cognitive Psychology*, 6, 421-435.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81-97.
- Murdock, B. B. (1995). Developing TODAM: three models for serial-order information. *Memory & Cognition*, 23(5), 631-645.



- Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review*, *104*(4), 839-862.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, *7*, 308-313.
- Ng, H. L. H., & Maybery, M. T. (2002). Grouping in short-term verbal memory: is position coded temporally? *Quarterly Journal of Experimental Psychology*, *55A*(2), 391-424.
- Norris, D., Hall, J., & Gathercole, S. E. (2019). How do we perform backward serial recall? *Memory & Cognition*, *47*(3), 519-543.
- Olivia, B., Saint-Aubin, J., Guérard, K., & Pâquet, M. (2017). Are lexical factors immune to response modality in backward recall? the effects of imageability and word frequency. *Canadian Journal of Experimental Psychology*.
- Osth, A. F., & Dennis, S. (2015). Prior-list intrusions in serial recall are positional. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1893-1901.
- Plant, R. R., & Turner, G. (2009). Millisecond precision psychological research in a world of commodity computers: New hardware, new problems? *Behaviour Research Methods*, *41*(3), 598-614.
- Richardson, J. T. E. (2007). Measures of short-term memory: a historical review. *Cortex*, *43*, 635-650.
- Ritchie, G., Tolan, G. A., Tehan, G., Goh, H. E., Guérard, K., & Saint-Aubin, J. (2015). Phonological effects in forward and backward serial recall: qualitative and quantitative differences. *Canadian Journal of Experimental Psychology*, *69*(1), 95-103.
- Ryan, J. (1969a). Grouping and short-term memory: different means and patterns of grouping. *Quarterly Journal of Experimental Psychology*, *21*(2), 137-147.
- Ryan, J. (1969b). Temporal grouping, rehearsal and short-term memory. *Quarterly Journal of Experimental Psychology*, *21*(2), 148-155.
- Serra, M., & Nairne, J. S. (2000). Part-set cueing of order information: implications for associative theories of serial order memory. *Memory & Cognition*, *28*(5), 847-855.
- Solway, A., Murdock, B. B., & Kahana, M. J. (2012). Positional and temporal clustering in serial order memory. *Memory & Cognition*, *40*(2), 177-190.
- St. Clair Thompson, H. L., & Allen, R. J. (2013). Are forward and backward recall the same? a dual-task study of digit recall. *Memory & Cognition*, *41*, 519-532.
- Surprenant, A. M., Brown, M. A., Jalbert, A., Neath, I., Bireta, T. J., & Tehan, G. (2011). Backward

- recall and the word length effect. *American Journal of Psychology*, *124*(1), 75-86.
- Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 37-55.
- Thomas, J. G., Milner, H. R., & Haberlandt, K. F. (2003). Forward and backward recall: different response time patterns, same retrieval order. *Psychological Science*, *14*(2), 169-174.
- Tremblay, A. (2013). LMERConvenienceFunctions: a suite of functions to back-fit fixed effects and forward-fit random effects, as well as other miscellaneous functions (version 2.5) [Computer software and manual]. <http://cran.r-project.org/web/packages/LMERConvenienceFunctions/index.html>.
- Vousden, J. I., & Brown, G. D. A. (1998). To repeat or not to repeat: the time course of response suppression in sequential behaviour. In J. A. Bullinaria, D. W. Glasspool, & G. Houghton (Eds.), *Proceedings of the fourth neural computation and psychology workshop: connectionist representations* (p. 301-315). London: Springer-Verlag.
- Wickelgren, W. A. (1964). Size of rehearsal group and short-term memory. *Journal of Experimental Psychology*, *68*(4), 413-419.
- Wickelgren, W. A. (1967). Rehearsal grouping and hierarchical organization of serial position cues in short-term memory. *Quarterly Journal of Experimental Psychology*, *19*(2), 97-102.