Constraints on Theories of Serial Order Memory Revisited: The Cases of the Fill-In and Protrusion Effects

Adam F. Osth (adamosth@gmail.com)
Simon Dennis (simon.dennis@gmail.com)
206 Psychology Building
1835 Neil Avenue
Columbus, OH 43210 USA

Abstract

In his seminal dissertation, Henson (1996) identified a number of constraints on theories of serial order memory. Two constraints, the *fill-in constraint*, in which an item that is erroneously recalled early is likely to be followed by its predecessor rather than its successor (recall of ACB is more likely than ACD), and the *protrusion constraint*, in which prior list intrusions are likely to be recalled in the same output position as their previous serial position, were considered evidence against chaining theories. We present results from two experiments which investigate the extent to which these effects are dependent on experimental methodology. When participants are given an open set of items, an equal ratio of fill-in and in-fill errors was observed and a protrusion effect was obtained. However, when a reconstruction of order task was used, a fill-in effect was observed. Implications for theories of serial order memory are discussed.

Keywords: serial recall; serial order memory; episodic memory; short term memory; working memory; chaining models; positional models; memory models

The problem of understanding serial order memory is possibly the oldest problem in the field of memory studies. Over a century of work has produced theories that can be broadly placed into two camps: *positional theories*, which claim that items in a sequence are coded with respect to their position in the sequence, and *chaining theories*, which claim that order is implicit in the associations among the items in a sequence; retrieval consists of using retrieval retrieved items as cues for their successors.

A watershed moment in theorization of serial order memory came with the seminal dissertation of Henson (1996). Through a series of experiments and meta-analyses, Henson outlined the regularities of the serial recall paradigm and dubbed them the constraints on serial order memory. The discussed regularities were diverse and ranged from effects of serial position to repetition errors and omissions. Of interest to the present investigation were regularities in error patterns that were particularly problematic for chaining models of serial order memory. One such regularity was the *fill-in effect*, which states that when an item is recalled one position too early, it is more likely to be followed by its predecessor than its successor. As an example, consider when a sequence *ABCDE* is studied and a subject erroneously recalls *AC*, skipping over the *B* item. Henson found that participants are more likely to recall *ACB* than *ACD*, as if they are "filling in" the missing response. The latter, the *ACD* case, is considered an *in-fill error*, and any chaining model with asymmetric associations is predicted to produce a greater incidence of in-fill errors than fill-in errors. The fill-in effect was replicated by Surprenant, Kelley, Farley, and Neath (2005) using a reconstruction of order paradigm.

Another such constraint was the *protrusion constraint*, which states that intrusions from prior lists tend to share the same position in both the current and prior lists. That is, if a subject is attempting to recall to item 3 on the current list and makes an intrusion from the prior list, the intrusion is most likely to be the third item on the previous list. Henson (1996) dubbed such in-position intrusions *protrusions*. The protrusion effect was initially discovered by Conrad (1960) and has been cited as evidence for positional coding.

Both the protrusion and fill-in constraints along with several other regularities that Henson (1996) discovered became canonical in the serial recall literature. Virtually every model that has been published since has no role of inter-item associations. The majority of the recent models are positional models; these include the the model of Burgess and Hitch (1999, 2006), Henson's own Start-End Model (Henson, 1998), the recurrent model of Botvinick and Plaut (2006), and the grouping model (Farrell, 2012). The constraints outlined by Henson (1996) were benchmark phenmonena for each of these models, as several of these models demonstrated how both fill-in and protrusion effects could be explained by positional representations.

The role of chaining in memory for serial order was revisited by Solway, Murdock, and Kahana (2012), who conducted a re-analysis of three serial recall studies and found a robust in-fill effect, contrary to the analyses of Henson (1996) and in agreement with the predictions of chaining models. They proposed a compound chaining model which yielded a good fit to the data, whereas the positional model of Burgess and Hitch (2006) did not yield an adequate fit. An open question remains as to why the analysis of Solway et al. yielded an in-fill effect whereas the analyses of Henson yielded the opposite pattern. In response to Solway et al., Farrell, Hurlstone, and Lewandowsky (2013) presented a re-analysis of over a dozen datasets, the vast majority of which yielded a fill-in effect. Farrell et al. posited that the datasets used by Solway et al. used longer list lengths than most serial recall experiments (minimum of ten items), which might be beyond the capacity of short-term memory and thus utilize different representations.

However, another difference between the studies analyzed by Solway et la. and Farrell et al. is that in the former, items were not re-used across trials (this is referred to as an *open* *set* of items). The studies considered by Henson and Farrell et al. often used small sets of stimuli such that items were frequently re-used across trials (referred to as a *closed set* of items). This is often standard practice in serial recall experiments, as it is believed that when a large set of stimuli is used the participant has to remember both the items and the order in which they occurred. If a closed set of stimuli is employed, item memory quickly reaches ceiling and only the order among the items has to be remembered on a given trial (Healy, 1974). Another similar approach to obtaining a relatively "pure" measure of order memory is to employ a reconstruction of order task, wherein the participant studies a list of items but at retrieval they are provided with the items and asked to place them in the correct order (Healy, Fendrich, Cunningham, & Till, 1987); this task was used in the experiments of Surprenant et al. (2005) that demonstrated a fill-in effect.

Closed sets and reconstruction of order tasks may yield different results than serial recall experiments with open sets in that the former tasks are susceptible to guessing strategies and the latter has a high incidence of omissions. When a closet set or a reconstruction task is used, item memory is at ceiling, which may cause participants to guess as to the locations of the items without having any knowledge of their positions. Under some circumstances, this can lead to an artificially high degree of fill-in errors. Consider a case in which a participant studied a list *ABCD* and knows the locations of *A* and *B*, but cannot recall the positions of *C* and *D* and guesses on the third and fourth responses. In this circumstance, the participant can either a.) get the sequence correct (*ABCD*) or b.) produce a sequence with a fill-in error (*ABDC*). Under this circumstance, only fill-in errors can be produced and there is no possibility for making an in-fill error.

However, when an open set is used in a serial recall task, item memory is imperfect and there is a high incidence of omissions, which can disguise a fill-in effect as an in-fill effect. Consider if sequence *ABCD* is represented in memory as *ACBD*. This would yield a fill-in effect if all of the represented items were output, but if *B* is omitted at retrieval, sequence *ACD* is produced. This may be especially likely in the studies of Solway et al. since the lists in the experiments they analzyed were quite long and the performance of the participants was relatively poor.

Even if guessing strategies are responsible for the finding of the fill-in effect, the compound chaining model of Solway et al. has no recourse for producing protrusion errors. Nonetheless, a number of traditional studies of serial recall collect responses on a series of lined response grids in which participants can see the previous responses. A simple explanation for the protrusion effect is that when participants are not able to recall a given item, they glance at the output position from the previous trial and use that response in their answer.

The present study attempt to control for both of these possibilites using simple means. In Experiment 1, we conducted a serial recall experiment that used a large open set of words, such that items are not repeated from trial to trial making it unlikely that participants should be able to guess from a pool of possible responses. Additionally, participants entered their responses on a keyboard rather than on lined response grids and were not given access to their previous responses. In Experiment 2, we used the same procedure as Experiment 1 but employed a reconstruction task by showing the participants the list items at retrieval in a new randomized order.

Serial recall experiments using open sets have been conducted, but to date there have been no analyses of whether the data collected from these experiments exhibit a predominance of fill-in or in-fill errors or whether they exhibit the protrusion effect. Investigations of set size effects have instead been concerend with other issues, such as word frequency effects (Roodenrys & Quinlan, 2000) or the phonological similarity effect (Coltheart, 1993; Conrad, 1963). We found that the open set did affect the fill-in to in-fill ratio, as we observed equivalent levels of fill-in and in-fill errors in Experiment 1. In Experiment 2, we conducted a reconstruction of order task and replicated the fill-in effect found by Surprenant et al. (2005).

Experiment 1: Serial Recall

Experiment 1 was an immediate serial recall task using a large set of words such that every list a participant received was composed entirely of unique items. To capture different levels of performance, we employed lists of five and six items and list length was manipulated between subjects.

Method

Participants A total of 204 undergraduate psychology students (105 participants for the list length five condition, 99 for the list length six condition) from The Ohio-State University participated in this experiment in exchange for course credit in an introductory psychology course.

Materials Participants studied words that were randomly selected from a word pool of 625 words from the Google search counts. Words ranged from 4 to 7 letters in length and from 30 to 250 counts per million in word frequency.

Procedure To familiarize participants with the nature of the task, all participants began each session with four unscored practice lists of three, four, five, and six items in order. Participants were given feedback upon completion of each of the practice lists. If any errors were made, participants were reminded that they have to recall the items in the order in which they were presented. No feedback was given at any point in the experiment after the practice session was completed.

Upon completion of the practice session, participants were given 62 trials with lists of either five or six items. During the study phase, participants were presented with each word for one second followed by a blank screen for 250 ms. Following completion of the study list, participants were presented with a recall prompt that was a series of three question marks ("???") on the center of the screen. Participants were instructed to begin recalling the items upon seeing the prompt by typing their respones on a keyboard and were given 20 seconds to recall the sequence. After the first key was pressed, the question marks disappeared and replaced with the letters typed by the participant. Participants signaled completion of a word by hitting the ”ENTER” key on the keyboard. Upon completing a response, the response disappeared from view on the computer screen and was replaced by the question marks. Participants signaled completion of the recalled sequence by typing the word ”done” and hitting ”ENTER.” Upon completion of each trial, participants signified readiness to begin the next trial by hitting the ”ENTER” key.

Halfway through the experiment, participants were given a break in which they played a digital card game for 180 seconds. Stimulus presentation and response collection was handled using the Python experimental library (Geller, Schleifer, Sederberg, Jacobs, & Kahana, 2007).

Results

Serial Position Effects The proportion of correctly items recalled in their correct serial position are shown in Figure 1. Performance was significantly worse in the length six condition than in the length five condition, $t(201.37)$[1] = 8.26, $p <$ .001. A mixed analysis of variance (ANOVA) using serial position as a within subjects factor and list length as a between subjects factor revealed that performance varied as a function of serial position, $F(5, 906) = 850.40$, and there was a serial position by length interaction, $F(4, 906) = 34.99$, both ps $<$ .001. Post hoc comparisons revealed a negative recency effect (poorer performance for the last item than the second to last item) in both the length five, $t(104) = -10.99$, and length six, $t(98) = -14.12$, conditions, both ps $< .001$.

Fill-In and In-Fill Errors Henson (1996) classified fill-in errors by focusing on all responses following the first item that was recalled one position too early. If the following item was from one serial position earlier than the just recalled item, this was considered a fill-in error. If the following item was from one serial position later than the just recalled item, this was considered an in-fill error. However, one need not just consider transitions that only traverse one serial position. Rather, the initial error could be more than one position too early and the subsequent fill-in or in-fill could be more than one position earlier or later, respectively. We will henceforth refer to the one position restriction as the *lag 1* analysis and the latter case which considers longer transition as the *any lag* case; both will be considered in the present analysis.

The analyses of Solway et al. (2012) and Farrell et al. (2013) both used a strict scoring procedure in which fill-in errors and in-fill errors are only considered for cases where the initial skip was the first error in the trial. We follow their example here and use strict scoring in our analyses.

The mean number of fill-in and in-fill errors for each classification style can be seen in Figure 2. Separate mixed ANOVAs with error type (fill-in or in-fill) as a within subjects factor and list length as a between subjects factor for both the any lag and lag 1 transition analyses. Results indicated there was neither a fill-in effect or an in-fill effect when transitions of any lag were analyzed, $F(1,202) = .246$, or when analyses were restricted to lag 1 transitions, $F = 1.291$, both ps $> .05$.

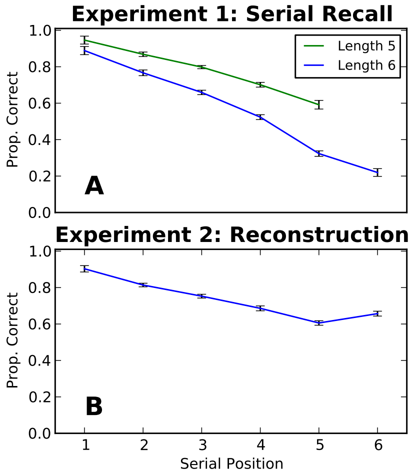

Figure 1: Serial position curves for Experiment 1 (A), which used a traditional serial recall procedure, and Experiment 2(B), which used a reconstruction task. Error bars indicate 95% within subjects confidence intervals.

Protrusion Errors All analyses on intrusion rates were restricted to immediate intrusions (intrusions from only one list prior to the current study list). A visualization of the intrusion rates for the serial position of each intruding item at each output position can be seen in Figure 3. As can be seen by the spiked nature of the graph, intrusions tend to appear in the same output position as their serial position in the prior list.

A statistical analysis was performed by calculating the proportion of immediate intrusions that were protrusions for each participant. Because there were some participants that exhibited no prior list intrusions, this analysis was restricted to participants that made at least one immediate intrusion error. The proportion of immediate intrusions that were protrusions (in-position) was above chance for both the list length five, $t(69) = 4.49$, and list length six conditions, $t(82) = 3.50$, both ps $< .001$, which is consistent with the findings of Henson (1996) and Conrad (1960).

Discussion

We conducted a serial recall experiment that used experimental parameters that were highly similar to previous experiments with the exception that stimuli were not reused from trial to trial in an effort to gauge the the generality of the constraints on serial recall established by Henson (1996). In contrast to Henson's data and the analyses of Solway et al., we observed roughly equivalent numbers of fill-in and in-fill er-

---

[1] t test degrees of freedom (df) are corrected df from the Welch-Satterwaithe equation.

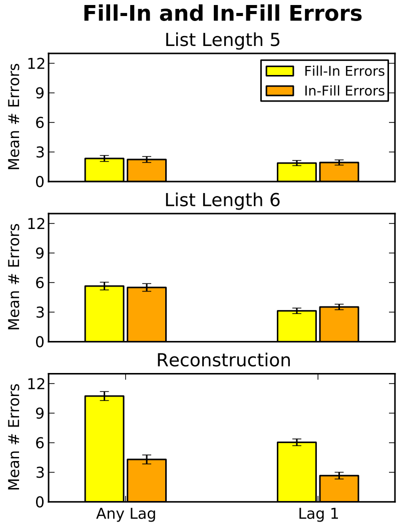

Figure 2: Mean number of fill-in and in-fill errors for each classification type for list length 5 (top) and list length 6 (middle) conditions of Experiment 1, along with the reconstruction task used in Experiment 2 (bottom). Error bars indicate 95% within subjects confidence intervals.

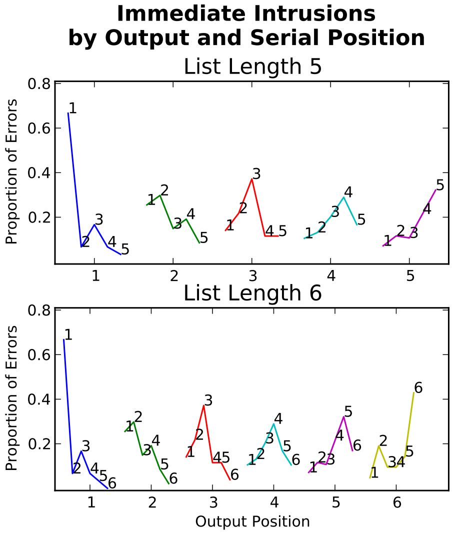

Figure 3: Proportions of immediate intrusions from each serial position in the prior list for each output position of the current trial for both the list length five (top) and list length six (bottom) conditions. The serial position of the prior list intrusion is indicated by the numbers on the lines.

rors. In other words, when an item was recalled too early, participants were equally likely to continue in a forward order or to return to the skipped item. The protrusion effect, in contrast, was robust in our experiment. The spiked, positional nature of immediate intrusions strongly resembled the data displayed by Henson (1996, 1998).

Nonetheless, it was still somewhat surprising that we did not observe a fill-in effect in our experiment because Surprenant et al. (2005) found a robust fill-in effect with both a small and large set of items. Nonetheless, their experiments used a reconstruction of order paradigm, and as previously discussed, such a task might introduce guessing strategies in the same manner as a small set size would in a standard serial recall paradigm. Thus, we repeated our experiment using the same parameters but employed a reconstruction of order task at retrieval and hypothesized that the results would favor a fill-in effect.

Experiment 2: Reconstruction of Order

Experiment 2 was identical to that of Experiment 1 with the exception that a reconstruction of order task was used at retrieval. During the test phase, participants were presented with all of the study list items in a randomized order and told to recall the words in the order in which they were presented. Because performance was far superior to that of the serial recall task in Experiment 1, we only collected data with a list length of six items.

Method

Participants A total of 95 undergraduate psychology students from The Ohio-State University participated in this experiment in exchange for course credit in an introductory psychology course.

Procedure The procedure was identical to that of Experiment 1 with the exception that at retrieval, participants were presented with all of the studied words at the top of the screen in a randomized order. The words remained fixed on the screen through the test and did not disappear as participants entered the corresponding words. Participants were instructed in advance of the experiment that they would be presented with the words and were told that they should type out the words in the order in which they had appeared in the study list.

Results

Due to the severe rarity of prior list intrusions in Experiment 2, analyses were restricted to serial position effects and fill-in and in-fill errors.

Serial Position Effects A one way repeated measures ANOVA revealed a main effect of serial position, $F(5, 470) = 225.6$, $p < .001$. Post hoc inspection of the serial position data revealed a positive recency effect for the last item in the list, $t(94) = 6.45$, $p < .001$.

Fill-in and In-Fill Errors Mean number of fill-in and in-fill errors were calculated for each participant and can be seen in the bottom panel of Figure 2. Fill-in errors outnumbered in-fill errors when any lag transitions were considered, $t(94) = 14.02$, and when only lag 1 transitions were considered, $t = 9.80$, both $ps < .001$. Thus, in contrast to Experiment 1, a fill-in effect was observed.

Discussion

Consistent with our hypothesis, usage of a reconstruction of order task at retrieval produced a robust fill-in effect. Our results are consistent with the findings of Surprenant et al. (2005), who found a fill-in effect with a large set size using a reconstruction task.

General Discussion

The two experiments we conducted found that a previously established regularity of serial order memory, the fill-in effect, depended on the type of task used to gauge serial order memory. In Experiment 1, we used a traditional serial recall paradigm with the exception that items were not repeated across trials and found equivalent levels of fill-in and in-fill errors. In Experiment 2, we employed a reconstruction of order task and found the traditional fill-in effect. These results are theoretically relevant because the fill-in effect has been used to argue against chaining models of memory, and while an in-fill effect was observed in the analyses of Solway et al., our results using shorter lists of words do not show a preponderance of in-fill errors over fill-in errors.

This begs the question as to what class of models the results in these experiments support. While an in-fill effect is predicted by chaining models with asymmetric associations (stronger associations in the forward than in the backward direction), chaining models with symmetric associations (equal strength associations in the forward and backward direction, such as in the TODAM model of Lewandowsky & Murdock, 1989) would be perfectly compatible with the equal ratios of fill-in and in-fill errors observed in Experiment 1 and a guessing response strategy could be appended to such models to produce fill-in effects as observed in the reconstruction task in Experiment 2. However, one should be reminded of the fact that the protrusion effect applied in both conditions of Experiment 1 and any simple chaining model with either symmetric or asymmetric associations has no recourse for predicting in-position prior list intrusions without positional representations.

Is there a role of inter-item associations in serial recall?

We reject the central claim made by Solway et al. (2012) that participants' recall sequences are indicative of a chaining model with asymmetric associations, as we were unable to replicate their finding of an in-fill effect using shorter lists of words. Nonetheless, a number of other studies in the literature have exhibited findings that are in accordance with the predictions of chaining models. During retrieval, when participants are given study list items in the same order as they were presented at study, participants perform better than when they're given unordered list items as cues, suggesting that they're using the cues to retrieve neighboring items from the list (Serra & Nairne, 2000; Basden, Basden, & Stephens, 2002). Similarly, when participants are given the same list of words from trial to trial but the starting point of the list differs (the spin list paradigm: Sequence *ABCDE* might be repeated as *CDEAB*), participants are only slightly worse than when the same list is repeated with all of the items in the same positions (Kahana, Mollison, & Addis, 2010). Positional models, in contrast, would predict a more dramatic impairment to performance in the spin lists than in the repeating lists. While positional models might be extended to produce both of these sets of results, these results follow intuitively from the predictions of chaining models.

While one might be inclined to suggest a hybrid model that incorporates inter-item associations to account for the above findings and positional representations to account for the protrusion effect [2], such a framework is not only inelegant but is an ad hoc solution to the problem. A more elegant approach is the constraint satisfaction (CS) model proposed by Dennis (2009). In the CS model, asymmetric associations in the forward direction are stored among all of the list items. The model differs from the aforementioned chaining models because at retrieval, the stored representation of the list is compared to all possible ordered list constructions; the distances between the possibilities and the list representation determine the probability of outputting a given sequence.

The basic principle behind the model is that the more differences there are between a candidate sequence and the studied sequence, the less likely it will be that the candidate sequence will be output. Dennis demonstrated that fill-in errors are more frequent than in-fill errors for this reason because a sequence with a fill-in such as *ACBDEF* only misses the *B-C* connection and erroneously introduces a *C-B* connection (two differences from the original sequence), whereas a sequence such as *ACDEF* misses all of the connections between *B* and its subsequent items (four differences from the original sequence). The model might be able to produce a lower incidence of fill-in errors if it's assumed that list items from the retrieved sequence are only output if they're sufficiently strong, producing more omissions when item memory is poor and lowering the fill-in to in-fill ratio. Dennis also demonstrated that introducing a component of similarity that is common to all of the items allows the model to capture key phenomena that have used to argue for positional representations, such as the protrusion effect as well as the mixed-list phonological similarity effect (e.g.: Baddeley, 1968).

Conclusion

Our work evaluating two of the constraints on theories of serial order memory established by Henson (1996) uncovered a generality of one (the protrusion constraint) and a limitation of the other (the fill-in constraint). These results may indicate a need to re-evalute whether inter-item associations are sufficient to support memory for serial order.

---

[2] The model of Burgess and Hitch (1992) incorporated both inter-item and positional associations. However, later versions of the model did not include inter-item associations.

Acknowledgments

We would like to thank Gina Gerardo, Mengze Zheng, Katherine Bubeleva, and Rebecca Kirchner for their assistance with data collection. We would also like to thank Per Sederberg for helpful discussion about these experiments.

References

Baddeley, A. (1968). How does acoustic similarity influence short-term memory. *Quarterly Journal of Experimental Psychology*, *20*.

Basden, B. H., Basden, D. R., & Stephens, J. P. (2002). Part-set cuing of order information in recall tests. *Journal of Memory and Language*, *47*, 517–529.

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review*, *113*, 201–233.

Burgess, N., & Hitch, G. J. (1992). Towards a network model of the articulatory loop. *Journal of Memory and Language*, *31*(4), 429–460.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*.

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, *55*, 627–652.

Coltheart, V. (1993). Effects of phonological similarity and concurrent irrelevant articulation on short-term memory recall of repeated and novel word lists. *Memory & Cognition*, *21*(4), 539–545.

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

Conrad, R. (1963). Acoustic confusion and memory span for words. *Nature*, *197*, 1029–1030.

Dennis, S. (2009). Can a chaining model account for serial recall? In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the XXXI Annual Conference of the Cognitive Science Society* (pp. 2813–2818).

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review*, *119*(2), 223–271.

Farrell, S., Hurlstone, M. J., & Lewandowsky, S. (2013). Sequential dependencies in recall of sequences: Filling in the blanks. *Memory & Cognition*.

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, Instruments & Computers*, *65*(1), 50–64.

Healy, A. F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 644–655.

Healy, A. F., Fendrich, D. W., Cunningham, T. F., & Till, R. E. (1987). Effects of cuing on short-term retention of order information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *13*, 413–425.

Henson, R. N. A. (1996). *Short-term memory for serial order.* (Unpublished doctoral dissertation. MRC Applied Psychology Unit, University of Cambridge, Cambridge, England.)

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

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.

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review*, *96*, 25–57.

Roodenrys, S., & Quinlan, P. T. (2000). The effects of stimulus set size and word frequency on verbal serial recall. *Memory*, *8*(2), 71–78.

Serra, M., & Nairne, J. S. (2000). Part-set cuing 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*, 177–190.

Surprenant, A. M., Kelley, M. R., Farley, L. A., & Neath, I. (2005). Fill-in and infill errors in order memory. *Memory*, *13*(3), 267–273.