Can a chaining model account for serial recall?

Simon Dennis (simon.dennis@gmail.com)
225 Psychology Building, 1835 Neil Avenue
Columbus, OH 43210 USA

Abstract

Henson (1996) has argued that several results including fillin effects, patterns of protrusions and performance on lists of alternating similar and dissimilar items (the sandwich effect) preclude a model of serial recall that relies on chaining associations between items. However, this conclusion is at odds with other data showing that serial recall improves dramatically when study lists approximate language at the letter and word levels and also is improved when circular lists that maintain chaining information, but confound positional information are repeated. In this paper, I demonstrate that the objections to chaining models can be overcome if one assumes that associations act as constraints on a whole of list resolution process, rather than acting in a purely feedforward fashion.

Keywords: serial recall, chaining models, positional models, ordinal models, syntagmatic paradigmatic model

The problem of serial order has been a focus of memory research since its inception (Nipher, 1878). The touchstone task for studying serial order memory has been immediate serial recall. Subjects are presented with a series of letters, digits or words and are then required to reproduce them in order. In response to the extensive empirical database that has been collected a number of computational models have been developed (Henson, 1998; Burgess & Hitch, 1992; Lewandowsky & Murdock, 1989; Page & Norris, 1998). These models can be divided into three main classes: chaining, positional and ordinal (Henson, 1998).

Chaining models assume that study of a serial list creates associations between successive items (Ebbinghaus, 1885/1913; Lewandowsky & Murdock, 1989; Murdock, 1995). At test, each item is used to retrieve the subsequent item (Ebbinghaus, 1885/1913). The simplest form of this theory would suggest that once an error had occurred performance on the remaining items on the list should be serverely degraded - a pattern which is not usually observed. To account for this problem, Lewandowsky and Murdock (1989) proposed a distributed memory model based on the Theory of Distributed Associative Memory (TODAM, Murdock, 1982). In TODAM, the result of retrieval is a noisy version of each item that may be insufficient to enable recall of that item, but can nonetheless be used as a retrieval cue to allow recall of the remaining items to continue. More complex chaining models have also been proposed in which compound cues constructed from multiple preceding items are utilized (Murdock, 1995).

Positional models of serial memory propose that order information is maintained by a set of position cues (Lee & Estes, 1977; Burgess & Hitch, 1999; Brown, Preece, & Hulme, 2000; Henson, 1998). At study, the items from the list are associated with these position cues. At test, the cues are reinstated and the items associated with them retrieved. Models differ in the nature of the cues they propose. Lee and Estes (1977) proposed separate control nodes for each of the positions in the list. Burgess and Hitch (1999) and Brown et al. (2000) employ banks of oscillators of different periods whose phase is reset by the start of the list. Henson (1998) proposes that there is a start maker whose activation decreases as the list progresses and an end marker whose activation increases as the list progresses. The positional cue is the vector containing the activations of these two markers.

Finally, ordinal models assume that the activation of the items in the memory decreases as a function of the position in the list (Page & Norris, 1998). At each stage, a general retrieval is initiated that recalls the strongest item in memory. This item is then suppressed to allow retrieval of the next strongest item etc. Ordinal models do not employ associative processes at all, relying purely on the gradient of activation to code order information.

Henson (1998, 1996) has argued for positional models and against chaining models on several grounds including:

1. When an item is omitted in recall it tends to appear in the next location - a phenomena called fill-in. So, for instance, it is more common to see the pattern ACB than the pattern ACD (Henson, 1996). Simple chaining models, however, predict that ACD should be more common as C will be associated with D and should be used as the retrieval cue for the third item.

2. When an item from a preceding list is erroneously recalled (a protrusion) it tends to be recalled in the position that it appeared in the previous list (Conrad, 1960), which would not be predicted by simple chaining accounts.

3. When lists are constructed from alternating confusable and nonconfusable items such as DQTMPK (where D, T and P share a rhyme), performance is degraded for the confusable items, but not for the nonconfusable items (Henson, Norris, Page, & Baddeley, 1996). Simple chaining accounts would predict that an error on one of the confusable items should compromise performance on the subsequent nonconfusable item.

However, data also exists that suggests that some form of chaining must underlie serial recall. In particular,

1. When lists are constructed from high probability letter bigrams (Baddeley, 1964) or words sequences that approximate language (Miller & Selfridge, 1951) they are easier to learn despite the fact that the statistics that are relevant

to these phenomena are not confined to any particular positions. Indeed, if we wish to entertain the hypothesis that the representations that underlie serial recall are fundamentally the same as those that underlie language, then it is critical that the representation not be tied to any simple notion of position, as languages are universally structure dependent - not position dependent (Chomsky, 1965).

2. When rotated versions of a list are repeated (e.g. ABCDEF, EFABCD, CDEFAB), retaining chaining information but confounding position information learning is well above baseline for permuted lists (Addis & Kahana, 2005), although this advantage can be largely eliminated by interposing a few novel items at the cut point (Hitch, Fastame, & Flude, 2005).

The apparently conflicting data summarized above present a puzzle. Do associations from item to item play an important role in serial order memory or not? To resolve the impasse, we introduce a version of the Syntagmatic Paradigmatic model (Dennis, 2003, 2005). In the SP model, order information is retained purely by virtue of chained associations. Nonetheless, we will see that the model is capable of accounting phenomena that have been proposed to rule out chaining models.

The Syntagmatic Paradigmatic Model

The Syntagmatic Paradigmatic Model (SP) is a computational account of verbal cognition (Dennis, 2005). It proposes that many linguistic tasks can be conceptualized as the resolution of syntagmatic and paradigmatic associations that are retrieved from long term memory. Syntagmatic associations are thought to form between items that follow each other in sequence (e.g. run fast). Paradigmatic associations are thought to form between items that may not occur within the same sequence, but that fill similar slots in different sequences (e.g. deep shallow).

Dennis (2005) formalized the notion of syntagmatic and paradigmatic association in a Bayesian model that has been used to account for a broad range of phenomena. These include syntactic structural analysis, long term grammatical dependencies and structure sensitivity, transformations, generativity and systematicity, garden pathing, the extraction of statistical lexical knowledge, structural priming in comprehension, verbal categorization and property judgment tasks, analogical, rule-based and statistical inference and unsupervised thematic role analysis (Dennis, 2004, 2005).

In this paper, we will focus on the syntagmatic associations proposed by the SP model, which can be seen as identical to pairwise chaining associations that extend across all lags. Unlike classic chaining models, however, the SP model assumes that these associations are used as constraints to guide a global resolution process rather than operating in a purely forward fashion.

The idea that serial recall of short lists involves an assembly phase followed by the ballistic output of assembled items is supported by reaction time data. Thomas, Milner, and Haberlandt (2003) found that time for first retrieval increased as a function of list length, but that subsequent recalls occur at constant intervals despite the fact that accuracy decreases (see Ferreira (1991) for a similar result in the context of sentence processing). If lists were constructed as they are output, one item at a time, then one would expect that reaction times would more closely mirror error rates.

Previous incarnations of the SP model have employed a probabilistic version of string edit theory as the basic mathematical framework (Dennis, 2005, 2004). In the next section, however, we develop a simpler version of the model that facilitates interpretation.

The Formal Model

Let l_i be the representation of the item in the ith slot. To begin with we assume that each item is locally coded (zeros in all components except one which is a one). During study we form the following syntagmatic matrix:

$$S = \sum_i \sum_{i<j} l_i l_j^T \tag{1}$$

To determine the probability that a possible output list will be emitted, we construct the syntagmatic matrix for that list and then determine to what extent it satisfies the constraints defined by the study list as represented by S.

$$E(l) = \left\| \sum_i \sum_{i<j} l_i l_j^T - S \right\| \tag{2}$$

where $\|.\|$ is the city block metric. The probability that a given list will be produced at test follows the following distribution:

$$P(l) \propto e^{-E(l)} \tag{3}$$

Figure 1 gives a simple example. If the study list contained the items ABCDEF then the first matrix would be its syntagmatic representation. To determine the probability of producing ACBDEF, for example, we would construct the second matrix and note that there are two components on which it differs from the study matrix, that is, two constraints from the study set that are violated. As a consequence, $P(ACBDEF) \propto e^{-2}$. Similarly, $P(ACDEF) \propto e^{-5}$. By enumerating all lists with nonnegligible probability, we can normalize appropriately and avoid the use of monte carlo simulations.

Some Results with the Unparameterized Model

Primacy, Recency and the Locality Constraint

Perhaps the most fundamental result that a model of serial recall must account for is the pattern of errors that occur across position. Figure 2 shows the performance of the SP model. Each line shows the probability that items from each of the input positions will be output in that position. The model exhibits improved performance for initial items (the primacy

ABCDEF Matrix

	A	*B*	*C*	*D*	*E*	*F*
A	0	1	1	1	1	1
B	0	0	1	1	1	1
C	0	0	0	1	1	1
D	0	0	0	0	1	1
E	0	0	0	0	0	1
F	0	0	0	0	0	0

ACBDEF Matrix

	A	*B*	*C*	*D*	*E*	*F*
A	0	1	1	1	1	1
B	0	0	0	1	1	1
C	0	1	0	1	1	1
D	0	0	0	0	1	1
E	0	0	0	0	0	1
F	0	0	0	0	0	0

ACDEF Matrix

	A	*B*	*C*	*D*	*E*	*F*
A	0	0	1	1	1	1
B	0	0	0	0	0	0
C	0	0	0	1	1	1
D	0	0	0	0	1	1
E	0	0	0	0	0	1
F	0	0	0	0	0	0

Figure 1: Matrices for an ABCDEF list, a ACBDEF list and an ACDEF list. The ACDEF list induces greater error than the ACBDEF list.

effect) and for final items (the recency effect). Furthermore, when errors occur they tend to involve items from adjacent positions. This observation has been dubbed the locality constraint by Henson (1998).

Repetition Errors

Repetition errors typically occur infrequently. Henson et al. (1996) report an incidence of 2% and Vousden and Brown (1998) report 5%. The SP model produces a repetition error in 10% of lists. Just as important as the rate, however, is the distribution of errors as a function of lag. Typically, repetition errors do not occur immediately, but rather are separated by 3-4 items. Henson et al. (1996) found that the average lag was 3.4 items. Figure 3 shows the proportion of errors as a function of lag. The model produced an average lag of 3.446 and shows the bow shaped distribution of errors that is typically of the human data.

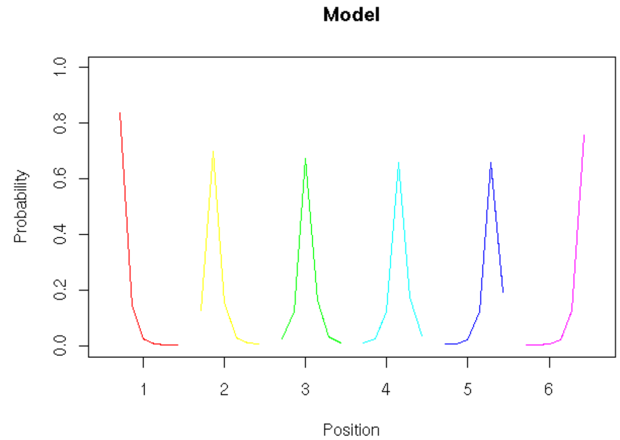

Figure 2: Serial position curve and position errors for the SP model. For each output position (labelled on the x axis), the probability of producing an item from a given input position is plotted.

The pattern of repetition errors is often used as evidence in favour of a response suppression mechanism. For instance, Lewandowsky (1999) uses a response suppression parameter to control the repetition error rate. In the SP model, however, there is no response suppression. Furthermore, there is no difference in the probability of recalling otherwise equivalent lists with repeated items at different lags (at least for the simple version of the model). For example, the error associated with the list AABC when ABC was studied is identical to that associated with ABCA (in both cases, there are three syntagmatic mismatches and one paradigmatic mismatch, so that $P(l) \propto e^{-3}$). What does differ, however, is how many potential output lists are associated with different lags. For example, the most probable lists containing a repetition for a three item study list are:

Lag 1	Lag 2	Lag 3
AABC	ABCB	BACB
ABBC	BABC	CABC
ABCC	ACBC	ABCA
	ABAC	

Given that these lists are equiprobable, the fact that there are more list types with lag two means that this will be the more probable outcome. Response suppression may still occur, but the above result demonstrates that the pattern of repetition errors does not provide compelling evidence.

Parameterizing the Model

The model presented above provides an interesting first approximation. However, to capture a number of other effects we need to introduce some parameters. In particular, we will define the error term as:

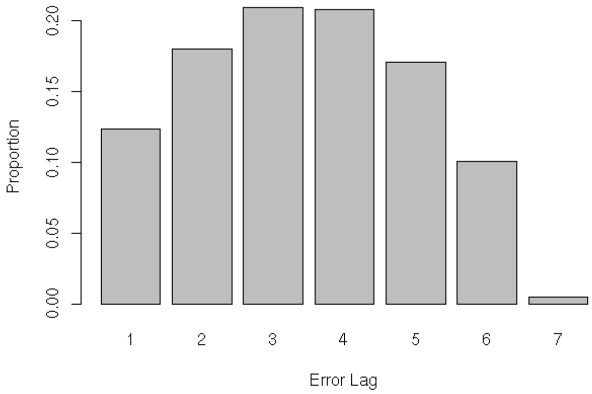

Figure 3: Repetition error rates as a function of lag.

$$E = \alpha \left\| \sum_i \sum_{i<j} l_i l_j^T e^{-\lambda j} - S^n \right\| + \beta \left\| \sum_i \sum_{i<j} l_i l_j^T e^{-\lambda j} - S^{n-1} \right\| \tag{4}$$

α parameterizes the strength of the list and can be manipulated to account for changes in exposure time etc. λ indexes the decrease in strength of encoding from the start of the list to the end and allows the model to capture the fact that the primacy effect is often stronger than the recency effect. At this point, we will remain agnostic as to what actually causes this difference. The chaining associations are assumed to be stored in long term memory and to be able to be retrieved given context cues. Of course, in immediate serial recall little retrieval is likely to be necessary as one is still within the same context. Nonetheless, there will be some retrieval from similar contexts both from general background language experience and also as a consequence of previous lists in the same experiment - in particular the preceding list. For our current purposes, we are going to be concerned with protrusions from the previous list (n-1) and so we assume that the error signal takes these associations S^{n-1} into account. β parameterizes the degree to which these associations are involved. Finally, in order to be able to talk about differences between similar and dissimilar items we must extend our representation of the input items to allow for overlap. Each item vector will consist of a unique component s_u and a common component s_c. Manipulating these parameters will allow us to change similarity (i.e. $l_A = (s_u, 0, 0, 0, 0, 0, 0, 0, 0, s_c)_T$, $l_B = (0, s_u, 0, 0, 0, 0, 0, 0, 0, s_c)_T$ etc. We are now in a position to address the arguments which might seem to be the most problematic for a model that relies on a chaining representation. All parameters were found using a best first recursive grid search optimizing a sum of squares error function. Unless otherwise noted, the values were α = 3.3, β = 1.05, λ = 0.12, s_u = 0.12, s_c = 2.6.

Fillin Errors

As suggested in the introduction, a critical result for chaining models of serial recall is the rate at which an omission in sequence is followed by the omitted item as compared to the next item in sequence. If a recalled list begins with AC, how often is the next item B versus D? Simple chaining models would predict that the D should be more common as it is the successor to C. Page and Norris (1998), however, found that ACB patterns were more common than ABD patterns. A higher proportion of fillin errors is a natural consequence of the SP model. Figure 4 shows the human proportions beside those for the SP model. The number of ACB errors exceeds the number of ABD errors.

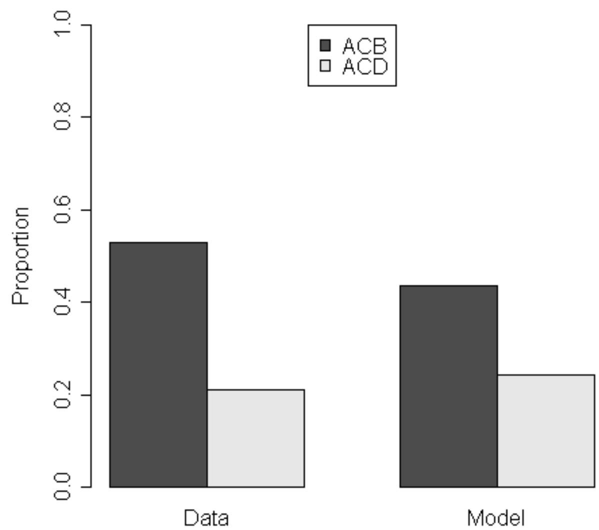

Figure 4: Fillin error rates from Henson (1996) for list lengths 7, 8, and 9 and from the model.

Figure 1 provides insight into why the model prefers ACB patterns to ACD patterns. The figure shows the syntagmatic matrices for a correct six item list (ABCDEF), and two error lists ACDEF and ACBDEF. The probability of a list is determined by the degree to which these matrices deviate from the correct matrices. So, in the unparameterized SP model $\frac{P(ACBDEF)}{P(ACDEF)} = \frac{e^{-2}}{e^{-5}} = 20.09$. The ACBDEF is more likely than ABDEF. Unlike simple chaining models, the SP model is affected by the syntagmatic associations to items following the omitted item.

Protrusions

Errors that occur as a consequence of protrusions from earlier lists show a systematic positional pattern. Figure 5 shows data from Conrad (1960) beside the fit of the SP model. At first this may seem counter intuitive. Recall, however, that each item vector now contains a common component s_c. This common component accumulates across items. So, items that appear at the start of the list have a strong association to the common component appearing to their right. Items that appear at the end of list have a strong association to the common component appearing to their left. Items that appear in the middle of the list have about equal associations to the common component appearing to each side. These observations are true both for items from the current list and for items from the previous list, so when protrusions occur they will tend to occur in place.

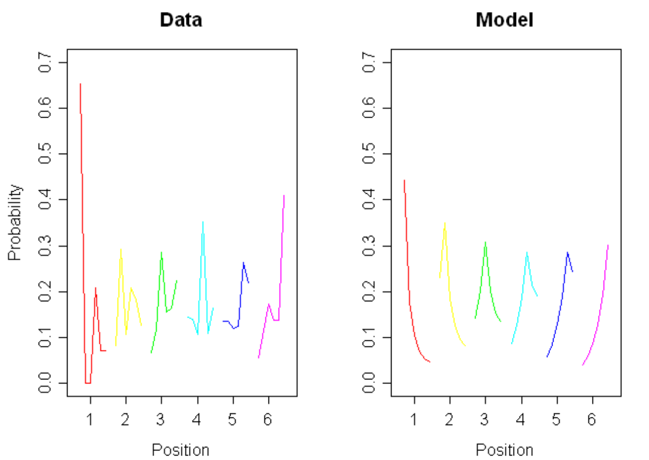

Figure 5: Patterns of protrusions. Data from Conrad (1960) on the left and the model fit on the right.

Effects of Inter-item Similarity

When lists are constructed from items that are acoustically similar (e.g. B, E, T, C, V, D) performance in serial recall drops (Conrad & Hull, 1964). As outlined above, a more discriminating result for models of serial recall, however, occurs when lists are constructed from alternating confusable and nonconfusable items (e.g. X, B, Y, E, K, T). Unsurprisingly, the confusable items are recalled less accurately. What is problematic for simple chaining models, however, is that performance on the nonconfusable items does not seem to be compromised as compared to a pure nonconfusable list (e.g. X, Y, K, L, Q, H). A simple chaining model would predict that performance on items immediately following confusable items should suffer. In fact, Farrell and Lewandowsky (2003) suggest that when conditions are designed to limit intrusions and omissions the nonconfusable items actually benefit from being embedded within confusable items as compared against pure lists of nonconfusable items.

Figure 6 shows data from Baddeley (1968) beside the performance of the model with the following parameters $\alpha = 4.7$, $\lambda = 0.2$, dissimilar items - $s_u = 1.0$, $s_c = 0.0$, similar items - $s_u = 0.3$, $s_c = 1.2$. The model demonstrates reduced performance for lists of confusable items and the saw pattern for the alternating lists. Importantly, performance on the nonconfusable items in the alternating lists is equivalent to that in pure nonconfusable lists. Furthermore, the model can also account for data in which performance on the nonconfusable items in confusable lists is superior than in pure nonconfusable lists as Farrell and Lewandowsky (2003) suggest is the case, when intrusions and omissions are minimized ($\alpha = 5.5$, $\lambda = 0.33$, similar items - $s_u = 0.39$, $s_c = 1.25$). The later result is possible because the common component of the similar items accumulates and allows the dissimilar items to be better localized.

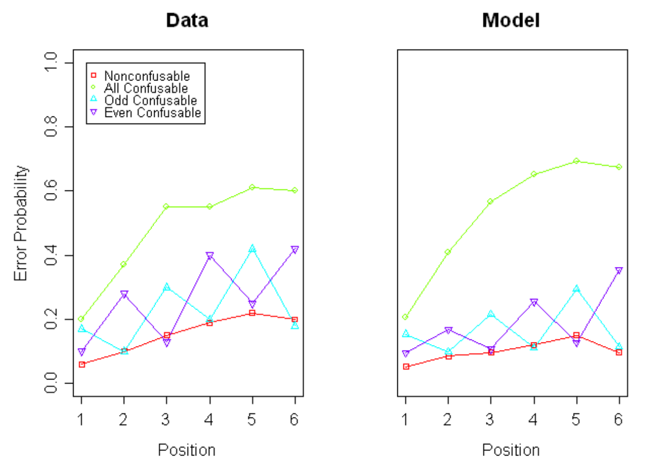

Figure 6: Probability correct on lists containing confusable items as a function of position in the human data (left) and the SP model (right). The four list types are pure nonconfusable, pure confusable, and alternating confusable - nonconfusable lists in which either the odd or even items were confusable. Data from Baddeley (1968).

Conclusions

The SP model takes a fundamentally different perspective on the serial recall task from existing models. Rather than focus on the item to item recall process, the model looks at the reconstruction of the list as a whole (c.f. Hitch et al., 2005; Botvinick & Plaut, 2006), as is suggested by observed pattern of reaction times (Thomas et al., 2003). The probability of producing a particular list is determined by the number of syntagmatic constraints established in the study episode that must be violated. The success of the model suggests that this shift in perspective may add important insight into the governing principles of serial recall.

In particular, fillin effects (Henson, 1996), patterns of protrusions (Conrad, 1960) and the sandwich effect (Henson et al., 1996) can all be accommodated within a model that uses chaining representations. Furthermore, because the model is based on chaining associations it is straightforward for it to account for approximation to language (Baddeley, 1964; Miller & Selfridge, 1951) and circular list Hebb effects (Addis & Kahana, 2005), by assuming that background associations like those responsible for protrusions are becoming stronger as a function of exposure to language at the letter and word levels both pre-experimentally and during the experiment. In this way, then the model resolves the apparent conflict between these different datasets.

There are, of course, many other serial recall phenomena that have yet to be considered. However, the current results suggest that as a representational substrate chaining models remain a viable alternative.

References

Addis, K. M., & Kahana, M. J. (2005). Circular lists cast negative spin on positional models of serial recall. In *Proceedings of the annual meeting of the psychonomics society.* Toronto, ON: Psychonomics Society.

Baddeley, A. (1964). Immediate memory and the "perception" of letter sequences. *Quarterly Journal of Experimental Psychology*, *16*.

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

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

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*, 127-181.

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

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

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

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

Conrad, R., & Hull, A. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, *55*.

Dennis, S. (2003). An alignment-based account of serial recall. In R. Alterman & D. Kirsh (Eds.), *Twenty fifth Conference of the Cognitive Science Society* (Vol. 25). Boston, MA: Lawrence Erlbaum Associates.

Dennis, S. (2004). An unsupervised method for the extraction of propositional information from text. *Proceedings of the National Academy of Sciences*, *101*, 5206-5213.

Dennis, S. (2005). A memory-based theory of verbal cognition. *Cognitive Science*, *29*(2), 145-193.

Ebbinghaus, H. (1885/1913). *Memory*. New York: Dover. (H.A. Ruger & C. E. Bussenius Trans. Original work published 1885)

Farrell, S., & Lewandowsky, S. (2003). Dissimilar items benefit from phonological similarity in serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*.

Ferreira, F. (1991). Effects of length and syntactic complexity on initation times of prepared utterances. *Journal of Memory and Language*(30), 210-233.

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.

Henson, R. N. A., Norris, D., Page, M., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology*, *49A*, 80-115.

Hitch, G., Fastame, M., & Flude, B. (2005). How is the serial order of a verbal sequence coded? some comparisons between models. *Memory*, *13*.

Lee, C. L., & Estes, W. K. (1977). Order and position in primary memory for letter strings. *Journal of Verbal Learning and Verbal Behavior*, *16*, 395-418.

Lewandowsky, S. (1999). Redintegration and response suppression in serial recall: a dynamic network model. *International Journal of Psychology*, *34*, 434-446.

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

Miller, G. A., & Selfridge, J. A. (1951). Verbal context and the recall of meaningful material. *American Journal of Psychology*, *63*, 176-185.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*(6), 609-626.

Murdock, B. B. (1995). Developing todam - 3 models for serial-order information. *Memory & Cognition*, *23*(5), 631-645.

Nipher, F. E. (1878). On the distribution of errors in numbers written from memory. *Transactions of the Academy of Sciences of St. Louis*, *3*, CCX-CCXI.

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

Thomas, J. G., Milner, H. R., & Haberlandt, K. F. (2003). Forward and backward recall: Different response time patterns, same retrieval order. *Psychological Science*, *14*, 169-174.

Vousden, J. I., & Brown, G. (1998). To repeat or not to repeat: The time course of response suppression in sequential behavior. In D. W. Bullinaria, D. Glasspool, & G. Houghton (Eds.), *Proceedings of the fourth neural computation and psychology workshop: Connectionist representations.* London: Springer-Verlag.