How are Repeated Items Encoded in Working Memory?

Danke X. Shieh (dxie@crl.ucsd.edu)
Jeffrey L. Elman (elman@crl.ucsd.edu)
Department of Cognitive Science, University of California, San Diego
La Jolla, CA 92093, USA

Abstract

How are repeated items encoded and retrieved in the brain? This paper presents a working memory model of serial recall based on a novel theory of repetition encoding. Instead of creating a separate node for each repeated item, repeated items are assumed to be recognized as repetition patterns, such as X^n (immediate repetition), XYX (alternation), and $XYZX$ (2-separated repetition). These patterns encode lists with repetitions into a list of non-repeating representations, which solves the problem of repetition. The above patterns can encode arbitrary lists up to 4 items. With an additional mechanism of chunking, robust encoding of arbitrary lists within memory span (∼7) can be achieved. The *Repetition Encoding and Chunking* model based on this theory is constructed and simulated. Its performance shows serial position effects, list-length effects, and chunking effects found in human immediate serial recall tasks. It offers insights into representations of patterns, binding and hierarchy, and offers an explanation for various memory errors, including misplaced repetition (233 →223) and alternation errors (232→323).

Keywords: working memory, short-term memory, immediate serial recall, serial order, problem of repetition, binding.

Introduction

How does the brain temporarily hold a list of numbers, or plan a sequence of movements? These tasks require a *short-term* or *working memory* for serial order during the performance of the task. The *serial order* problem has been seen as fundamental to understanding the brain and cognition (Lashley, 1951). In recent years, many theories and models of working memory have addressed this problem, both in psychology (Burgess & Hitch, 1999; Henson, 1998b; Page & Norris, 1998; Farrell & Lewandowsky, 2002) and neuroscience (Melamed, Gerstner, Maass, Tsodyks, & Markram, 2004; Beiser & Houk, 1998).

One influential theory, the *ordinal theory*, assumes that order is represented by relative values of a continuous property (e.g., the activation levels of items) with the first item "strongest" and the last item "weakest". The order of these items is retrieved by iteratively selecting the most active item, and then suppressing it. Specifically, when order is represented by the activation levels of items, the theory can be referred to as the *activation gradient* theory.

This theory is most explicitly explored in the primacy model (Page & Norris, 1998), which can reproduce many dominant patterns of recall errors (omission, intrusion, and transposition) (Page & Norris, 1998). Other major working memory models of serial order has also explicitly or implicitly adopted an activation gradient, or a functionally equivalent counterpart, to encode serial order, such as the competitive queuing mechanism in Burgess and Hitch (1999), the start markers in Henson (1998b), and possibly also the learned mechanism of the neural network model (Botvinick & Plaut, 2006). Some researchers consider activation-gradient-based retrieval as a theoretical convergence achieved in working memory models in the recent 10 years (Farrell & Lewandowsky, 2002).

An issue that has arisen with the wide acceptance of the *activation gradient* theory is that the gradients are modeled in various ways, and there hasn't been a biologically-based account for how the brain generates it. In our recent Divergent-Reconvergent model (Shieh & Elman, 2006), we addressed this issue, and proposed a biologically plausible mechanism based on the connectivity of the basal ganglia (Graybiel, Aosaki, Flaherty, & Kimura, 1994). This idea is inherited but revised in this model, taking into account the crucial fronto-basal ganglia loops. It is worth noting that traditionally the basal ganglia are seen as a neural substrate for learning *long-term* motor sequences. But the view has changed, and many researchers now believe that the basal ganglia play similar roles in perceptual, linguistic and cognitive sequences based on neuropsychological and anatomical evidence (Lieberman, 2000).

Despite considerably wide acceptance of the activation gradient theory, there is a crucial issue that needs to be solved for the theory to capably perform immediate serial recall. The issue is the problem of repeated items.

Repetition problem

The problem of repeated items is well-known for any models based on the activation (or primacy) gradient (Page & Norris, 1998). In these models, an item is typically represented by a single node. The second instance of an item will increase the activation level of the node that has already been activated, but this will not correctly represent the two instances (Bradski, Carpenter, & Grossberg, 1994). In order to encode and retrieve repeated items, additional mechanisms are needed. Some researchers remark that activation-gradient-based models may require a *tokenization* process (Henson, 1998b), which hasn't been addressed by existing models except (Bradski et al., 1994).

Technically, it is indeed possible to use a tokenizer to preprocess the input sequence so that it doesn't have repeated items and can be encoded/recalled by existing activation-gradient models (e.g., ABAD → A_1BA_2D where A_1 and A_2 are different nodes). Such a model is functionally identical to the STORE3 model of immediate serial recall (Bradski et al., 1994), which combines tokenization and activation gradients. However, from human performance, we know that different tokens of the same type have more intriguing interactions, which cannot be explained by tokenization alone.

Interesting phenomena can be observed from motor short-term memory outputs in typing, especially typing errors. Some typing errors are highly informative for how repeated items are encoded, for example, misplaced repetitions (*supper* → *suuper*) and wrong repetition numbers (*suppper*) (Rumelhart & Norman, 1982). These errors suggest that repeated items are represented by an item code and a separate repeating code. The errors occur when the repeating code is mistakenly applied to other items (misplaced repetitions), or perturbed by noises (wrong repetition numbers). Another error, the alternation reversal error (*these* → *thses*), suggests a special alternating mode, where two letters are bound with an alternating mechanism that produces the form XYX. When the wrong letter is produced first (e.g., due to binding error), the output will be the reversed alternation YXY.

The underlying principle seems to be avoiding the representation of multiple tokens, by other mechanisms that compensate for the representational loss. The brain seems to face the same challenge as the activation gradient models and has evolved/developed some mechanisms to remedy it. But imperfection of the solution can be seen in short-term memory tasks. It is well-known that lists containing repeated items are more difficult to recall than lists without repetitions – known as *repetition inhibition*, or the *Ranschburg effect* (Jahnke & Bower, 1986).

Some information about the coding can also be seen in an interesting exception to the above *Ranschburg effect*. When the repeated items are adjacent (e.g., 1RR294), facilitation happens instead of inhibition (Henson, 1998a). That is, the list containing a pair of doubled item is recalled better than lists with distinct items. Repetition facilitation further supports *doubling* coding, because it reduces the number of items needed to be remembered.

In the next section, we propose a theory based on the above observations of repetition encoding. It is hypothesized that a few repetition patterns are used to encode repeated items. Combining our previous model based on activation gradients and this theory, we develop a new model *Repetition Encoding and Chunking (REC)* model. The model's serial recall performance will be quantitatively fit and compared with human data.

Encoding Repetitions

Small patterns

In this paper, we propose that the brain uses two mechanisms to overcome the problem of repetition: *patterns* and *chunks*. The basic concept for patterns is that the brain uses them to the recode stimulus and remove repetitions. A pattern is a form in which the repeating items occur. For example, a *doubling* is a pattern with a the form X^2 and a variable X which denotes which item is repeated. Though there are two instances of X in the input, with the pattern only its identity (one instance) needs to be encoded. In general, each pattern has a pattern identity or form, such as X^2 or XYX, and one or more *variables* (e.g., X, Y, Z) that can be bound to other items or patterns.

It is not easy to determine what patterns are used by the brain for encoding. For example, should we include complicated patterns such as X^5 and $XYZWX$? In this research, we adopt two strategies in selecting the patterns: 1) only include repetition patterns with high recall accuracy in short-term memory studies; 2) only include patterns needed to support robust recall of arbitrary lists within some length. We choose the length to be 4 here, because people are exceptionally good with this length (Severin & Rigby, 1963). Further, four also reflects a binding capacity (number of feature conjunctions) that restricts the use of patterns for lists longer than about 4 items (Luck & Vogel, 1997).

From the two strategies, we hope to identify patterns that are both useful and necessary (under assumptions of the activation gradient theory). Using Strategy 2, we would find that the patterns in Table 1 are needed to systematically encode arbitrary lists of 4 items (verified below). This is a minimal set of patterns. With them, other patterns like XXY or $XYXZ$ are not necessary.

Table 1: Patterns that encode repetitions

Pattern	Description
X^n	immediate repetitions of an item X (n=2,3,4)
XYX	repetitions separated by one extra item
$XYZX$	repetitions separated by two extra items

Using strategy 1, we would examine what patterns are most accurately recalled in short-term memory tasks. They turn out to be the same patterns as above, which better supports the selection of these patterns. When the subjects recall repeated items, the recall accuracy is high when these items are in patterns X^2 (71%), XYX (52%) and $XYZX$ (54%), but quickly drops to 33% for the pattern $XYZWX$ (Henson, 1998a, Exp. 2A). Because patterns like $XYZWX$ have low recall accuracy, dropping them would not greatly affect performance of the model. In other words, even if the brain may occasionally use more complicated patterns, its performance can still be well approximated by this model.

With this said, it may be needed to point out that there is indeed a natural generalization of the above patterns (X^n, XYX and $XYZX$) that brain may all recognize: they are all cycles of up to 4 items with one item as the stopping point. This generalization would also cover patterns like XYZXY, whose cycle is XYZ, and the stopping point is Y (not the X as in the basic patterns)[1]. However, because they only apply to highly regular stimuli, we do not include them here.

[1]They can be seen as a prefix of the regular language $(\Sigma^n)^m$, where Σ is the alphabet, n is the cycle length and m is the number of cycles.

Robust encoding of lists of up to 4 items

Here we verify that our encoding scheme can indeed encode all lists of four items; that is, converting any list of 4 items into a list of non-repeating items and patterns. All possible cases of 4-item lists are shown in Table 2. In the table, repeated items are denoted by R and S, and non-repeating items are denoted by a and b. A bracket is a pattern.

Table 2: Encoding 4-item lists without repeated item/pattern

Type	Encoding
One repeated item	
Repeated 2 times	[RR]ab, a[RR]b, ab[RR]
	[RaR]b, a[RbR]
	[RabR]
Repeated 3 times	[RRR]a, [RR]aR, [RaR]R*, a[RRR]
Repeated 4 times	[RRRR]
Two repeated items	
Separate	[RR][SS]
Interleaved	[R[SS]R]

* RARR can also be encoded as Ra[RR].

It is important to notice that some patterns are nested, as in the case of RSSR. For this list, the pattern X^2 is first applied to the internal S's, and then XYX is applied to the remaining items. Without nesting, the pattern needs to be encoded by a new pattern.

It is possible that hierarchical encoding is essential of working memory organization. With hierarchical encoding, the working memory model can also represent syntactic structures in language processing. In theory, it is possible that the adaptive advantage of solving the sequence representation problem might lead to the evolution of neural mechanisms that are later used for language.

Chunks

When the list length is 5 or more items, some lists cannot be represented without using some items or patterns, for example, RABCR. To account for the brain's ability of encoding arbitrary lists up to about 7 items, extra mechanisms may be needed.

Chunking is a possible mechanism that overcomes the problem of repetition. It is known that the best encoding strategy for 7 items is to chunk them as 3 plus 4 items (Severin & Rigby, 1963). The effect of chunking can be seen as improving recall in a way that repetition patterns cannot. Though it is not immediately clear how chunking improves recall, it is assumed that items do not interfere as much as they are in the same chunk, at least under some conditions. One of such conditions is when the repeated items are in the same within-chunk position (Henson, 1998a). In the current model, the interference between items in different chunks is temporarily eliminated for simplicity, but will be included in future refinement of the model.

Using patterns and chunks, a long list can be divided into several chunks so that repeated items in a chunk can be represented by patterns (Figure 1). The branches show how a list is divided. In example A, X^2 and XYX are patterns that encode repetitions. In B, a long list is chunked into two short ones without repetitions within them.

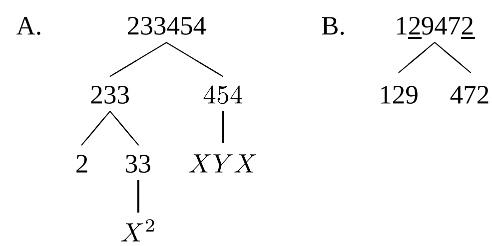


Figure 1: Encoding of lists with repeated items

The Repetition Encoding and Chunking Model

Theoretical Neuroscience pioneer David Marr advocated three levels of analysis: computational, algorithmic, and implementational. They can be understood as the goal of the computation, representation and algorithm, and neural implementation. For this research, we implement the model at the algorithmic level. The focus is to model the strategies and methods used by the brain, but for crucial components of the model, we resort to neural network models for better biological relevance.

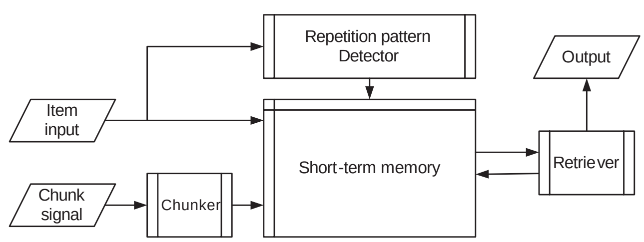


Figure 2: Components of the Repetition Encoding and Chunking model

The *Repetition Encoding and Chunking* model has several interactive components. *Item input* presents items to the system one at a time. The input item enters both the *short-term memory (STM)* and the *repetition pattern detector*. When the *repetition pattern detector* detects one of the three patterns in Table 1, it reorganizes the items in *STM* to reflect the pattern. The *chunk signal* component can receive chunking requests, and create new chunks in *STM*. *Retriever* can recursively (not just iteratively as in CQ models) select the most active item, pattern or chunk in *STM*, and produces the corresponding output.

Representation

The basic representation of information is a unit. A unit x has an activation level $A(x)$, a phase $\phi(x)$, and a response inhibition value $r(x)$. When an item is presented, an initial activation level is assigned to it, which is later subject to spontaneous decay. The phase represents the unit's relation to other units, units in the same group and having the same phase are *bound* together. The representation of binding by synchrony is used to bind items and variables in *patterns*. The response inhibition suppresses a unit after it is retrieved.

The content of a unit is either a basic item, or a *group*. A group is a flat structure containing a collection of items, with the property that items in a group compete for response during recall. The group is a general structure used for representing *patterns* and *chunks*.

A special group, the *primary group*, is the major space of information storage. All items must first enter the primary group, before they can be organized into a pattern, or a chunk. The primary group in this model is equivalent to item-level representations in many other working memory models.

When items are organized into a pattern or a chunk, a new group is created. The affected items are transferred from the primary group to the new group with their original activation levels and phases. Then, a unit that links to the new group is inserted into the primary group. An example of the internal representation of the list 233454 is given in Figure 3.

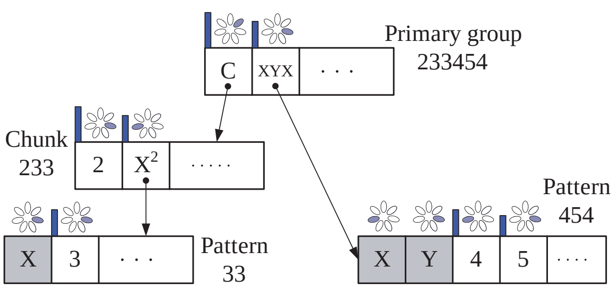


Figure 3: An example of internal representation of the model. The bar over an item shows its activation level. The flower shape with a dark/blue petal shows its phase. In patterns, an item having the same phase as a variable is bound with it.

Serial order

In the model, serial order is represented by relative activation levels of active items. This mechanism has been investigated in the Primacy Model (Page & Norris, 1998), Burgess and Hitch's (1999) model, and the Divergent-Reconvergent model (Shieh & Elman, 2006). But there are differences in the way the activation gradients are generated. Since the Divergent-Reconvergent model is based on neural anatomy, we will also use this mechanism in this model. To do this, we first derive an analytical form of the activation gradient function from the Divergent-Reconvergent model.

The basic idea of the Divergent-Reconvergent model (with some revision) is that input the prefrontal (PFC) area passes through the bottle neck of the basal ganglia, and projects back to the prefrontal area. The function of the fronto-basal ganglia loop is to add an activation-gradient over the active items in the PFC, so that they can be competitively selected by the (pre-)motor cortex for output.

The activation gradient is generated in the following way. Each item in PFC connects to a large number of units in the striatum. Each of these striatal units, in turn, has lateral inhibition on several neighboring units. Due to this lateral inhibition, late items do not have less strong activity in the striatum, forming an activation gradient that encodes the serial order.

The activation gradient depends on two parameters: the density ρ of striatal units each item connects to, and the factor η of lateral inhibition (the average number of neighbors inhibited by one striatal unit). Let A_n denote the density of striatum units activated by the n^{th} item. A_n can be recursively calculated as

$$A_n = \begin{cases} \rho & n = 1 \\ \rho \prod_{i=1}^{n-1} (1 - \eta\, A_i) & n > 1, \end{cases} \tag{1}$$

where $1 - \eta\, A_i$ is the proportion of units *not* inhibited by the i^{th} item. Eq. (1) is used in this model to compute the initial activation of a unit when it is activated in the primary group. When a unit is transferred from the primary group to other groups, it keeps its current activation level. After a unit is activated, its activation level $A(x;t)$ at time t is subject to decay:

$$A(x;t) = (1 - \gamma)\; A(x;\; t - \Delta t), \tag{2}$$

where Δt is time step of the simulation, the decay rate γ is chosen so that a fully active item decays to the retrieve threshold θ in T seconds.

Patterns

A pattern in the system is represented by a pattern unit associated with a group (see the X^2 unit in Figure 3). In the group, it has a number of variables, and items associated with this pattern. Associated items are bound to the variables having the same phases.

A pattern is formed when it is detected. For example, when the pattern 4-5-4 is recognized, a new pattern unit XYX is first created. Then, the active items 4 and 5 in the primary group are transfered to the new group and bound with corresponding variables. At last, the pattern unit XYX is stored and activated in the primary group. Its activation level is set to that of the most active item in the pattern.

While a pattern is active, binding errors may occurred if the phases of variables and bound items are desynchronized. This accounts for errors where the list 232 is retrieved as 323 by mistake.

Chunks

When the model receives a chunk signal during the presentation of a list, it creates a new chunk unit associated with a new group (see the unit C in Figure 3). Then, it transfers all unchunked items in the primary group into the new group, and store and activate the chunk unit in the primary group. The process is similar to the creation of a pattern, except that no variable binding occurs here. The activation level of the chunk unit is also set to that of its most active item.

Retrieval

Retrieval starts from the primary group. The *retriever* selects the unit x with the strongest output strength $s(x)$ and above the retrieve threshold θ:

$$s(x) = A(x)/\rho - r(x) + \epsilon, \tag{3}$$

where ϵ is the noise term with $\epsilon \sim \mathcal{N}(0, \sigma^2)$ and σ is a noise level parameter. After outputting the item, the unit is depressed by a high $r(x)$ value. If the content of the unit is a pattern or a chunk, items in it are recursively retrieved.

Parameter estimation

Parameters of this model are as shown in Table 3. There are 5 parameters. For each we have chosen 4-10 levels based on pilot experiments. There are about 30,000 combinations of parameters. Instead of fitting all parameters at once, we separated parameters in two groups (ρ, η, σ) and (T,θ) and alternatively fit each group, until the RMS converges. The optimal sets of parameters for different experimental data vary in a range, due to differences in materials, procedure and subject population, but the parameters ρ, η about the neural mechanism are relatively stable. The ranges of optimal parameters are given in Table 3.

Table 3: Model parametrization

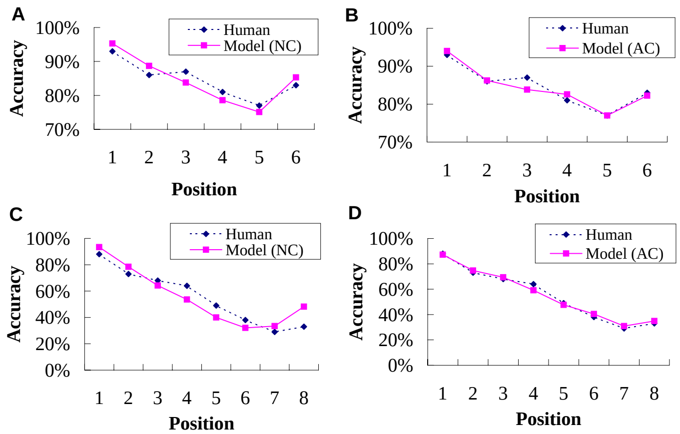
Param.	Range	Description
ρ	.02-.03	Divergent density
η	3-6	Lateral inhibition factor
σ	.02-.04	Std. Dev. of Gaussian noise
T	50-100 sec	Activation timescale
θ	5-20%	Retrieve threshold

Simulations

Serial position curves

We first examine the models' capability of fitting human experimental data. The material for this simulation is lists of non-repeating items (lists with repeated items will be analyzed next). We used two versions of the model: the first is an activation-gradient model without repetition patterns and chunks (NC, no chunking), and the second chunks the lists into two groups (AC, averaged chunking). The chunking boundary is randomly chosen for each list from 2-4, 3-3 and 4-2 (for 6-item lists), and 3-5, 4-4 and 5-3 (for 8-item lists). The two models have the same set of parameters, optimized with the above method.

The results are shown in Figure 4. Both models fit the data fairly well. This confirms the validity of activation-gradient-based serial recall (Page & Norris, 1998; Shieh & Elman, 2006). Since there are no repeating item, the new mechanism (patterns and chunks) is not expected to make a big difference. Nonetheless, the new model fits human data a little better. If this effect is true, it may suggest that people may voluntarily chunk the lists into 3 or 4 items even when this is not required.



RMS errors of A: 2.50%, B: 1.54%, C: 8.28%, D: 2.34%.

Figure 4: Serial-position curves of the model and human data. The top panels (A,B) are based on lists of phonologically dissimilar letters in Experiment 1 of Henson et al. (1996). The bottom panels (C,D) are based on Experiment 1 of Page and Norris (1998).

Recall of lists with repetitions

In this simulation, we include lists with and without repeated items. A basic activation-gradient model would systematically fail to encode repeated items. Here, we measure how much the new model improves recall of lists with repeated items.

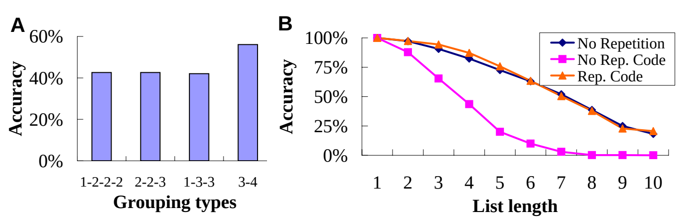


Figure 5: Recall accuracy of lists with repetitions

After the model parameters are optimized by the serial position curve above, its recall accuracy for lists of different lengths but without repeated items is plotted (Figure 5B, No repetition). This list-length curve exhibits the typical reversed S-shape as in people's serial recall. When we fit the model against a human list-length curve (Guildford & Dallenbach, 1925), the fit is very tight with RMS error as low as 1.2%. This suggests a good match of the model to human performance.

Next and crucially, we repeated this simulation with lists containing repeated items. Performance of the model without repetition coding (patterns and chunks) dropped dramatically

(No. Rep. Code). However, the model with repetition encoding (Rep. Code) remained at same performance level as for lists without repeated items (No repetition). The difference is dramatic, as shown in Figure 5B. With the new mechanism, the activation-gradient model can recall lists with repeated items as accurately as lists without repeated items. The mechanism effectively remedied representational deficiency in basic activation-gradient models.

At last, we tested the model's recall accuracy on pre-chunked lists of 7 items, in order to find out the best chunking strategy for the model. Four grouping types are compared as shown in Figure 5A. The 3-4 pattern is the most suitable for the model, consistent with the result in human experiments (Severin & Rigby, 1963). The reason, we believe, is that the chunk sizes of three and four make the best use of the encoding power of the repetition patterns. This result further supports the validity of the patterns used.

Conclusion

This research proposes a new theory of how the brain encodes and retrieves sequences with repeated items, and its processing of repeated items in general. Using patterns and chunks, any span-length lists with repeated items can be robustly encoded, in a way that is suitable for a biologically plausible serial order mechanism (Shieh & Elman, 2006).

The above theory is implemented in our computational model of immediate serial recall, the *Repetition Encoding and Chunking (REC)* model. The model can produce serial position effects, list-length effects, and chunking effects as found in human immediate serial recall tasks.

The repetition encoding scheme gives it the potential to account for many errors that have not been adequately studied: misplaced repetitions (233→223), alternation errors (858→585), wrong repetition numbers (cheese → cheeese), grouped transposition (233 → 332). It also gives working memory models the potential for language processing, by virtue of the activation-based serial order and repetition encoding, chunking, and variable binding mechanisms. Many aspects of this model deserve further investigation, including but not limited to position coding, temporal grouping, cognitive control and language processing. They are directions for future research.

Acknowledgments

Thanks to Hal Pashler, Marty Sereno, Gary Cottrell, Jochen Triesch, Marta Kutas, and Ginny de Sa for insightful and supportive comments on this research. This research was supported in part by grant NIH MH60517 to the second author.

References

Beiser, D., & Houk, J. (1998). Model of cortical-basal ganglionic processing: Encoding the serial order of sensory events. *Journal of Neurophysiology*, *79*, 3168-3188.

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

Bradski, G., Carpenter, G. A., & Grossberg, S. (1994). Store working memory networks for storage and recall of arbitrary temporal sequences. *Biological Cybernetics*, 469-480.

Burgess, N., & Hitch, G. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106 (3)*, 551-581.

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychon Bull Rev*, *9*(1), 59-79. (1069-9384 Journal Article)

Graybiel, A. M., Aosaki, T., Flaherty, A. W., & Kimura, M. (1994). The basal ganglia and adaptive motor control. *Science*, *265*(5180), 1826-31.

Guildford, J. P., & Dallenbach, K. M. (1925). The determination of memory span by the method of constant stimuli. *American Journal of Psychology*, *36*, 621-628.

Henson, R. N. A. (1998a). Item repetition in short-term memory: Ranschburg repeated. *J Exp Psychol Learn Mem Cogn*, *24*(5), 1162-81.

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

Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *49A*(1), 80-115.

Jahnke, J. C., & Bower, R. E. (1986). Are there two ranschburg effects? *The American Journal of Psychology*, *99*(2), 275-288.

Lashley, K. (1951). The problem of serial order in behavior. In L. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley.

Lieberman, P. (2000). *Human language and our reptilian brain: The subcortical bases of speech, syntax, and thought*. Cambridge: Mass: Harvard University Press.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279-281.

Melamed, O., Gerstner, W., Maass, W., Tsodyks, M., & Markram, H. (2004). Coding and learning of behavioral sequences. *Trends in Neurosciences*, *27*(1), 11-14.

Page, M., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105 (4)*, 761-781.

Rumelhart, D. E., & Norman, D. A. (1982). Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science*, *6*, 1-36.

Severin, F. T., & Rigby, M. K. (1963). Influence of digit grouping on memory for telephone numbers. *Journal of Applied Psychology*, *47*(2), 117-119. (English)

Shieh, D. X., & Elman, J. L. (2006). The divergent-reconvergent model of serial order encoding and retrieval. In *Proceedings of the 29th annual meeting of the cognitive science society.*