In Kokinov, B. & Hirst, W. (Eds.), *Constructive memory* (pp. 129-137). Sofia, Bulgaria: New Bulgarian University (2003).

Constructive processes in immediate serial recall: A recurrent network model of the bigram frequency effect

Matthew Botvinick
Center for Cognitive Neuroscience and
Department of Psychiatry
University of Pennsylvania

David C. Plaut
Departments of Psychology and Computer Science and
Center for the Neural Basis of Cognition
Carnegie Mellon University

Contact information: Matthew Botvinick, MD, PhD
University of Pennsylvania
Center for Cognitive Neuroscience
3815 Walnut Street
Philadelphia, PA 19104-6196
United States

Telephone: (215) 898-2960
Email: mmb@mail.med.upenn.edu

Abstract

Short-term memory for serial order, like other domains of memory, is subject to constructive effects. In particular, background knowledge concerning regularities in sequential structure can affect serial recall. A clear example of this is the bigram frequency effect, whereby letter strings conforming to the transitional probabilities of English are better recalled than random strings. Such effects present difficulty for most current models of immediate serial recall, which rely on associations between individual items and context representations. At the same time, approaches that have proven effective in modeling sequence processing in other contexts, such as recurrent neural networks, have been viewed as inapplicable to short-term serial memory, since they are thought to depend fundamentally on inter-item associations or chaining. We present a recurrent neural network model of immediate serial recall that overcomes this apparent dilemma. The model is used to simulate data from Baddeley (1968) that are widely agreed to rule out chaining models. The model is then applied to the bigram frequency effect, as reported in Baddeley (1964) and Kantowitz, Ornstein and Schwartz (1972). The simulation results dispel the notion that recurrent neural networks are simply chaining models, and provide what appears to be the only available computational account of the bigram frequency effect.

Introduction

At least since the famous work of Bartlett (1932), it has been clear that constructive processes play an important role in memory. What we remember is determined not only by the material presented, but also by how that material fits with our background knowledge. Of course, this principle applies across a wide variety of domains and time-scales. In the present paper, we focus on the particular question of how constructive processes may operate in short-term memory for serial order.

While the issue of sequential structure plays a role in many studies of constructive memory, there are relatively few studies that have focused specifically on the topic. One of the earliest is a study by Miller and Selfridge (1951). This reported that immediate recall of word lists was positively related to how well those lists approximated the word-to-word transition probabilities of English. Subsequently, studies by Baddeley (1964, 1965), Mayzner and Schoenberg (1965), and Kantowitz, Ornstein and Schwartz (1972) demonstrated an analogous effect for lists of English letters. Here, recall was better for letter strings that mirrored the *bigram frequency structure* of English — i.e., the frequencies with which specific pairs of letters adjoin in English text — than for strings that conserved only the frequencies of individual letters (see Figures 3 and 4). In these studies, as in other instances of constructive memory, we find recall affected by domain-specific background knowledge, here, knowledge concerning the transition probabilities among sequentially-organized items.

Despite their interest, such findings have typically been marginalized, if not ignored, in recent work on immediate serial recall. A review of such work suggests a possible reason for this neglect, which is that constructive memory effects are problematic for currently popular theories of serial recall. A majority of recent models are based on the idea that short-term memory for serial order depends on temporary, rapidly formed links between individual item representations and representations of position or temporal context (e.g., Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1999; Henson, 1998; Houghton, 1990). Such a mechanism is insensitive to the presence of consistent serial structure across items, both because the associations it involves are transient, and because — for reasons we shall discuss in a moment — the mechanism eschews item-to-item associations. As a result, short of importing additional ad hoc mechanisms (e.g., Hartley & Houghton, 1996), current models are unable to address such constructive phenomena as the bigram frequency effect. As Henson (1998) directly states, such "interactions between short- and long-term memory pose problems for most models of serial recall" (p. 115).

A more effective approach to understanding short-term memory for serial order is suggested by network models emerging from neuroanatomical and neurophysiologic research (e.g., Beiser & Houk, 1998; Dominey, Arbib & Joseph, 1995; O'Reilly & Soto, 2002). According to these accounts, serial order memory arises not from temporary, weight-based associations between item and context representations, but instead from the sustained activation of processing units conjointly representing item and order information. A key element of this general account is the presence of recurrent or feedback connectivity that supports the preservation and integration of serial order information by allowing activation to circulate over feedback loops.

One factor that makes current neuroscience-based models appealing, from the point of view of constructive memory phenomena, is that they closely resemble recurrent neural network models that have been proposed in the psychological literature to account for numerous aspects of sequence processing. Recurrent networks have been successfully applied to sequence parsing (e.g., Elman, 1990), sequential prediction (e.g., Cleeremans, 1993), and sequence production (e.g., Botvinick & Plaut, submitted), to mention only a few examples. Such studies suggest that recurrent networks, unlike the context-association models often applied to immediate serial recall, are well suited to tasks where it is important to encode domain-specific regularities in sequential structure.

Despite this strength of recurrent networks, and despite their popularity among neuroscientifically minded theorists, recurrent networks have never been applied, in any detailed way, to behavioral data relating to short-term serial order memory. Indeed, many modelers in the field have expressed doubt that such an undertaking would be worthwhile. This is because recurrent networks are widely viewed as *chaining models*, that is, models based on inter-item associations. While chaining was considered to be a possible basis for serial memory in early work (e.g., Wickelgren, 1966), this has essentially been ruled out by subsequent empirical findings. Perhaps the most compelling data are those reported by Baddeley (1968; see also Henson, 1998). Baddeley studied memory for lists containing both acoustically confusable and non-confusable letters. A chaining account would predict that non-confusable letters positioned after confusable letters should be recalled less accurately than the same letters appearing in lists made up only of non-confusable items. The data contradicted this prediction: Non-confusable items were recalled just as accurately when they followed confusable items as when they did not (see Figure 2, left).

From the point of view of constructive memory, the discussion so far leaves us with an apparent paradox. The empirical data rule out item-to-item associations — chaining — as a mechanism for serial recall, and, as a result, current psychological theories avoid mechanisms involving such associations (including recurrent neural networks). However, without a mechanism for encoding the relationships among items, it seems impossible to account for constructive phenomena such as the bigram frequency effect. To pose the challenge concretely: It is unclear how a single mechanism could give rise both to the bigram frequency effect and to the pattern of behavior observed by Baddeley (1968).

We present here a computational model of immediate serial recall that does account for both of these phenomena. The model takes the form of a recurrent neural network, essentially identical in structure to networks that have been applied to other aspects of sequence processing (e.g., Botvinick & Plaut, submitted; Cleeremans, 1993; Elman, 1990). Elsewhere, we report simulations applying the model to a wide range of behavioral data in the domain of immediate serial recall (Botvinick & Plaut, in preparation). Here, we focus specifically on two empirical benchmarks: the findings of Baddeley (1968), and the bigram frequency effect, as reported by Baddeley (1964) and Kantowitz, et al. (1972). Our simulation results point to two basic conclusions: 1) Recurrent neural networks should not be regarded simply as chaining models; 2) Although recurrent networks need not chain, they are nonetheless sensitive to serial structure, allowing them to provide an account of immediate serial recall that extends to constructive phenomena, including the bigram frequency effect.

The model

Our model is based on the connectionist or parallel distributed processing framework (Rumelhart & McClelland, 1986), and implements the now well-known simple recurrent network architecture of Elman (1990). The network is made up of simple processing units, organized into three groups or layers (Figure 1). An input group sends connections to an internal or hidden group of 200 units, which in turns feeds into an output group. Recurrent connections (associated with a conduction delay of one time-step) are included from the output group to the internal group, and among the units within the internal group. These recurrent connections allow the network's state on each time-step to influence its state on the subsequent step. Further details of the achitecture are provided below, and in a separate article (Botvinick & Plaut, in preparation).

The model was used to simulate the task of immediate serial recall. Item presentation was simulated by activating input units corresponding to specific items (details below). During this phase, the network's task was to "shadow" the input, by activating the identical units in the output layer. Following presentation of the last list item, a special unit in the input layer was activated to serve as a recall cue. Beginning on this time-step, the model's task was to activate output units corresponding to the items in the presented list, one per time-step, and in the original order. The model was trained to perform the task using a version of the backpropagation learning algorithm, adapted to the simple recurrent network architecture (Rumelhart, Durbin, Golden & Chauvin, 1996). During training, lists of variable length were presented, ranging from 1 to 6 items for Simulation 1 and from 1 to 9 items for Simulation 2. Steepest descent learning was employed, with a learning rate of 0.001. Training continued until performance reached accuracy levels roughly approximating those for benchmark empirical data. It should be noted that length of training was the only free parameter involved in the simulations. Following training, model performance was evaluated by holding connection weights constant, and presenting further lists simulating task conditions from relevant studies (as detailed below).

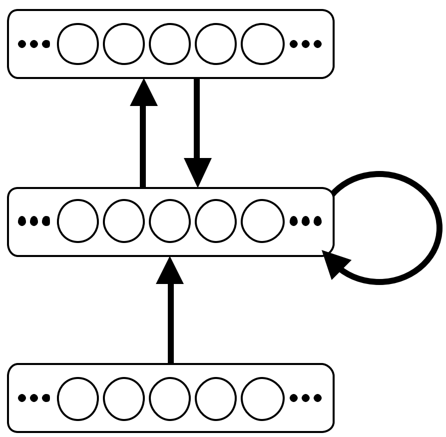

Figure 1. Schematic of the simulation model. Arrows represent full projections, containing connections from every unit in the sending layer to every unit in the receiving layer.

A brief note is warranted concerning the training procedure. It is, of course, not our view that the ability to perform immediate serial recall is based on explicit training specifically on this task. Instead, we suggest that the relevant processing mechanisms develop through experience with a wide variety of tasks — most notably language use — that demand the integration of information over time, and the subsequent production of ordered sequential responses. In addition, neurodevelopmental processes that are independent of experience may contribute to establishing at least the rudiments of these mechanisms (see Beiser & Houk, 1998). Our approach to training in the present simulations is thus, to some extent, a computational convenience, allowing us to bypass the simulation of much more diverse and long-term developmental processes.

All simulations were conducted using the LENS neural network simulator (Rhode, 1999).

Simulations

We present here the results of two simulations. In the first, the model was applied to the experimental paradigm of Baddeley (1968). This simulation functioned as a test of the claim that recurrent networks necessarily function through chaining, a claim that, if true, would rule out recurrent networks as models of serial recall. The second simulation evaluated the effect of domain structure on serial recall, by simulating two studies of the bigram frequency effect (Baddeley, 1964 and Kantowitz et al., 1972).

Simulation 1: Testing the chaining claim

Methods. A key aspect of Baddeley's (1968) study was that it involved both confusable and non-confusable items. In order to simulate the experiment, it was thus necessary to use item representations that involved some similarity structure. This was accomplished by representing items using a two-dimensional code. Each item was represented by one unit from each of two sub-groups within the input (and output) layers, containing 6 and 36 units respectively. The units in these two groups were used, collectively, to represent 36 items. Each item included a unique unit from the 36-unit group and one unit from the 6-unit group. Each unit in the smaller group featured in a total of 6 items. Items sharing a unit from this group were considered to be confusable, given their representational overlap. Items involving different units from the 6-unit group were considered non-confusable. Items were represented identically in input and output layers.

The model was trained on randomly structured (but non-repeating) lists for a total of 202,000 lists of each length. It should be noted that over 1.4 billion different lists of length 6 could be constructed from the 36 input items. Thus, training exposed the network to well under 1/6000 of potential target lists. At test, the model was presented with four types of 6-item list: pure non-confusable (PN) lists, involving only non-overlapping items; pure confusable (PC) list, involving items all of which were mutually overlapping; lists in which items 1, 3, and 5 were mutually confusable (AC, for alternating/confusable); and lists in which items 2, 4, and 6 were confusable (AN, for alternating/non-confusable). During testing, responses were gathered by identifying the item representation most closely resembling the pattern of activation in the output layer, with equal weighting given to the two dimensions used in representing items.

Results. Following Baddeley (1968), performance was evaluated by calculating the proportion of trials on which items at each position were recalled incorrectly. Results are diagramed in Figure 2 (right). As in Baddeley (1968), non-confusable items were recalled more accurately than confusable items, leading to better performance on PN than PC lists, and a "sawtooth" pattern for AC and AN lists. As in many studies of serial recall, a pronounced primacy effect, and a smaller recency effect, were also evident. The critical finding, however, was that, as in Baddeley (1968), accuracy for non-confusable items in AC and AN lists was virtually identical to that for items in the corresponding positions in PN lists.

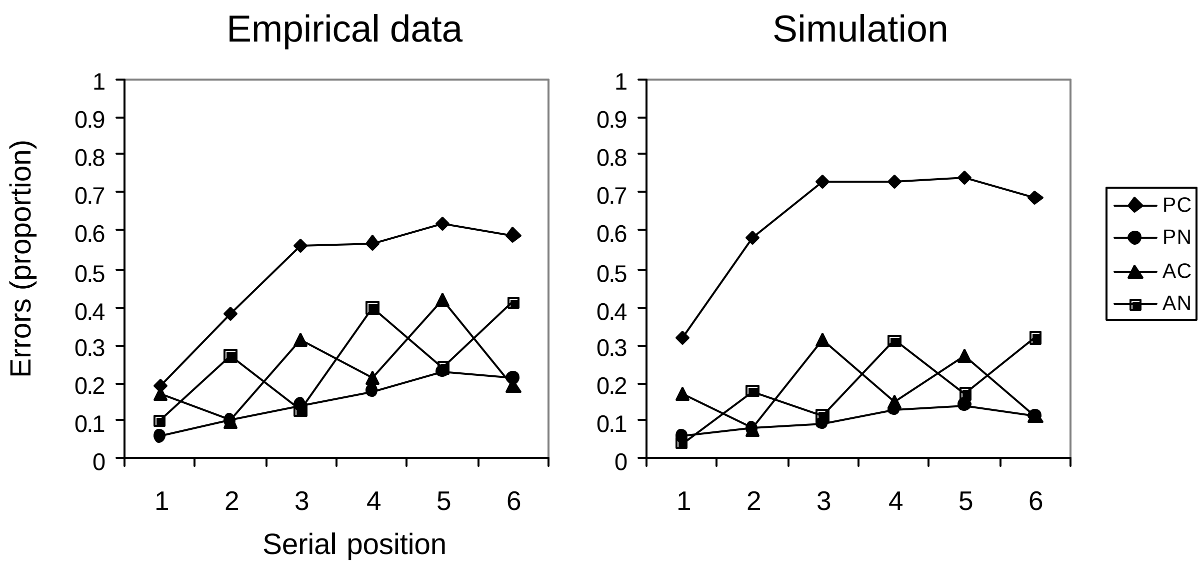

Figure 2. Error rates, by serial position, for lists including confusable and/or non-confusable items. Left: Empirical data, estimated from Figure 4 of Baddeley (1968). Right: Simulation results, based on a sample of 5000 trials. PC = pure confusable, PN = pure non-confusable, AC = alternating, first item confusable, AN = alternating, first item non-confusable.

Discussion. There is a widely held view (e.g., Brown, Preece & Hulme, 2000; Henson, 1998; Houghton, 1990), that recurrent networks cannot be considered as candidate models of immediate serial recall, because such networks function through inter-item chaining. This simulation put that claim to the test, by using a recurrent network to simulate the experiment of Baddeley (1968), a study whose results are widely acknowledged to be incompatible with chaining accounts. The model captured the key aspects of the empirical data rather precisely. We conclude from this that the claim that recurrent networks must operate through chaining can be set aside.

Simulation 2: Effects of domain structure on serial recall

Having shown, in Simulation 1, that the present model does not function simply through item-to-item chaining, the question arises: Is the network nonetheless sensitive to consistent item-to-item transition probabilities? If trained in a domain involving serial structure, would the model's recall performance be influenced by the degree of fit between target lists and the domain's sequencing constraints? In human behavior, this phenomenon is evident in the bigram frequency effect. We asked whether the model, if trained on letter strings reflecting the transition probabilities of English, would display the bigram frequency effect, as reported by Baddeley (1964) and Kantowitz et al. (1972).

Methods. For this simulation, input items were represented using single units, one for each letter of the alphabet. Training lists were constructed based on letter-to-letter transition probabilities, computed based on a large corpus of text from the Wall Street Journal (see Marcus, Santorini, & Marcinkiewicz, 1993). Two full training runs were conducted. During the first, lists involving all letters of the alphabet were permitted. For this simulation, testing involved 8-item lists that reflected the transition probabilities used during training ("first-order" lists), and, for comparison, lists that were randomly sequenced, preserving only the frequencies of occurrence of individual letters ("zero-order" lists). Results from this simulation were compared with data from Baddeley (1964). The second simulation run was intended to simulate Kantowitz et al. (1972). In that study, subjects were presented with 9-item lists always involving the same 9 consonants (in varying permutations), and performance was compared between a subset of the lists that contained high-probability letter transitions (high bigram-frequency lists) and a subset that contained less frequent transitions (low bigram-frequency lists). In our simulation, the model was trained on lists involving the same letters used in the empirical study, generated using the corpus-based transition probabilities. At test, non-repeating lists including the same 9 letters were presented. These were divided into two groups, based on a median split on cumulative bigram frequency.

For the Baddeley (1964) simulation, training proceeded for a total of 116,000 lists of each length (for 8-item lists, this meant exposure to well under 1/1,000,000 of potential target lists). For the simulation of Kantowitz, et al. (1972), training involved a total of 24,000 lists at each length (allowing for presentation of less than 1/15,000 of potential target lists of length 9).

Results. Baddeley (1964) compared recall for 8-item lists that conformed to the bigram frequency structure of English with recall for lists that preserved only isolated letter frequencies. As illustrated in Figure 3 (left), first-order lists were recalled more accurately than zero-order lists. The performance of the present model is illustrated in Figure 3 (right). As in the empirical data, recall was more accurate for first-order lists. Our goal in the present studies was to reproduce qualitative patterns of performance, rather than seeking quantitative fits. However, it is worth noting that the bigram frequency effect is, if anything, stronger in the simulation than in the empirical data.

Similarly confirmatory results were obtained in our simulation of the study of Kantowitz, et al. (1972). Their data was presented in the form of positional accuracy curves (Figure 4, left), which indicated that high bigram-frequency lists were recalled more accurately than low bigram-frequency lists, with the discrepancy being small at early list positions, and growing over time. Our simulation results, plotted in a similar manner, are shown in Figure 4 (right). Once again, the model's behavior closely resembled that reported in the empirical study.

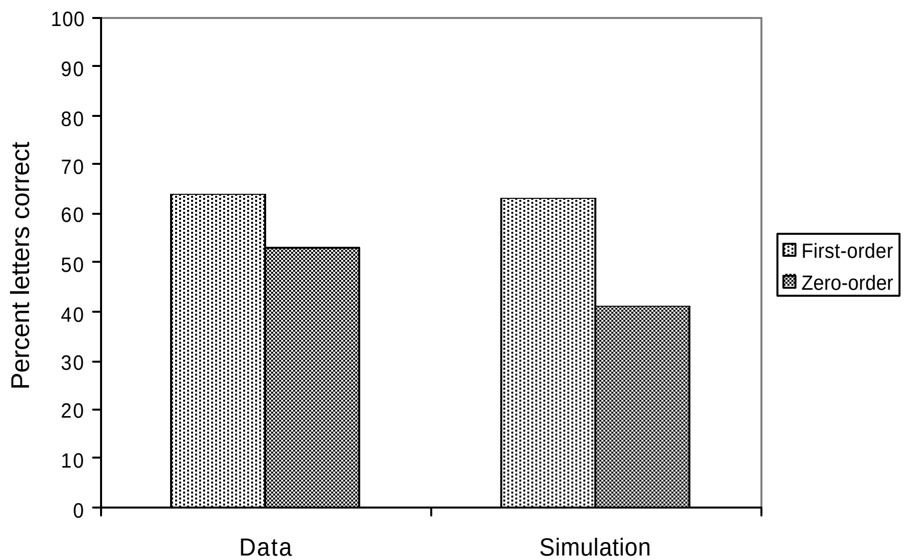

Figure 3. Percentage of letters recalled in the correct position, for lists based on first-order transition probabilities, and zero-order letter probabilities. Empirical data are estimated from Figure 1 of Baddeley (1964). Simulation data are based on a sample of 5000 trials.

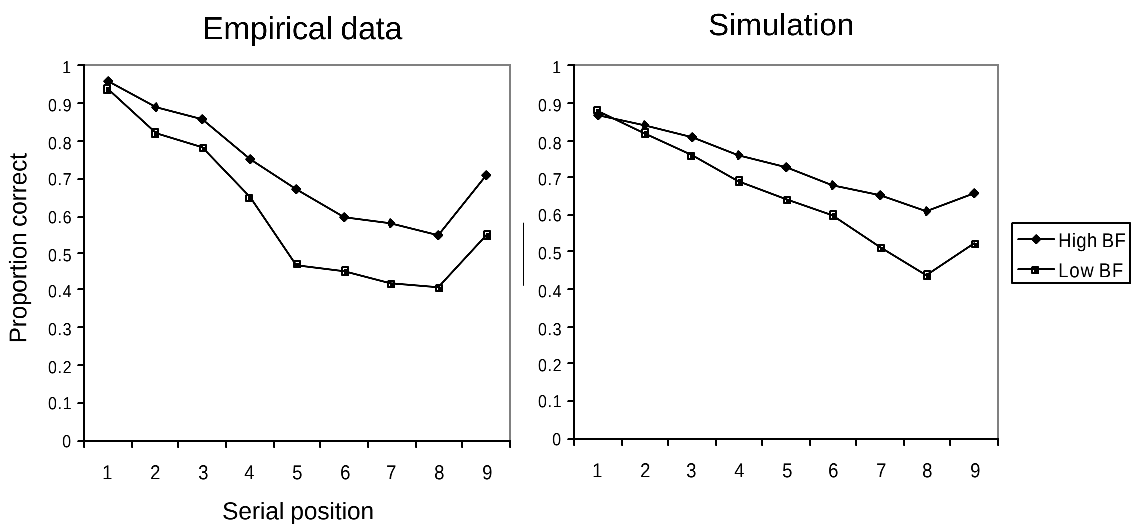

Figure 4. Proportion correct responses, by position, for letter lists involving high and low bigram-frequencies (BF). Data are estimated from Figure 1 of Kantowitz, Ornstein, and Schwartz (1972). Simulation data are based on a sample of 5000 trials.

General discussion

It is well established that short-term memory for serial order is subject to constructive effects. Background knowledge concerning the sequential structure of a domain influences the recall of items drawn from that domain. This fact, illustrated by such phenomena as the bigram frequency effect, presents a puzzle for models of serial recall. On the one hand, it requires that the relevant memory mechanisms be sensitive to distributional patterns governing item-to-item transitions. On the other hand, this must be accomplished without making the encoding of specific items wholly dependent on inter-item chaining.

We have presented a recurrent neural network model of immediate serial recall, which shows how both of these constraints might be met by a single mechanism. The model simulates Baddeley (1968), indicating that it does not function simply by chaining. However, the model also clearly displays the bigram frequency effect, showing that its recall performance is influenced by domain-specific background experience. To our knowledge, this is the only published model of immediate serial recall that accounts for this combination of findings.

It is a key feature of the model's sequencing mechanism that it is shaped by domain-specific experience. Importantly, the specific, tuned sequencing mechanism that develops through this experience is neither obliged to work through inter-item chaining, nor restricted from doing so. Instead, the degree to which item-to-item transitions are utilized depends on the nature of the domain and the task. Where inter-item associations are unhelpful for the task of recall, the mechanism will learn to neglect them. However, in cases where such associations can be leveraged for performing recall, the sequencing mechanism will learn to capitalize on them. In situations where the system does learn to utilize sequential dependencies, recall is improved for items that fit well with the sequential structure of their source domain, and hindered for items that do not fit well. It is this aspect of the model that leads it to display the type of constructive memory effects we have sought to address.

Some further analyses of the model’s performance have led to an interesting observation, and in turn, a prediction. As we have shown, the model's recall performance depends on the goodness-of-fit between the sequential structure of target lists and the sequencing constraints associated with their source domain. For brevity, we may refer to this goodness-of-fit as a list's "grammaticality." A closer look at the model's performance indicates that, in fact, it is not precisely the grammaticality of the target material that determines accuracy of recall. Rather it is the grammaticality of this material *relative to other potential responses.* This is because alternative responses that are much more grammatical than the target are prone to 'intrude' on recall. As a result, recall of a target item depends not only on the grammaticality of the target list, but also on the grammaticality of all other potential responses. This fits well with empirical evidence from bigram frequency studies, which have indicated a tendency to regularize; when recall is inaccurate, the list reported tends to be more grammatical than the original list (see, e.g., Mayzner & Schoenberg, 1964). However, this is still not the entire story. On the model's account, the tendency of alternative responses to intrude depends not only on their relative grammaticality, but also on their *distance* from the target. Items that are similar to the target are more likely to appear as responses than items that are dissimilar. Of course, this tendency, considered on its own, is a well-known aspect of serial recall. The novel observation from the model is that *recall in structured domains is determined by an interaction, for all alternative responses, between 1) their grammaticality relative to the target, and 2) their distance from the target*.

This aspect of the model's function can be illustrated using the following scenario, which translates into a testable prediction. Consider two target lists with an identical degree of grammaticality (for example, two consonant lists fitting equally well with the constraints of English). Now assume that, for one of these, there is an alternative response that is far more grammatical than the target, and that is also very similar to the target. Meanwhile, for the other target list, there is no such enticing neighbor. Every alternative list that is at all similar to this target is highly ungrammatical. If the account provided by the model is correct, the second of these two targets should be recalled better than the first. This prediction, and the hypothesized interaction that it reflects, is currently being put to experimental test in our laboratories.

References

Baddeley, A. D. (1964). Immediate memory and the "perception" of letter sequences. *Quarterly Journal of Experimental Psychology,* **16,** 364-367.

Baddeley, A. D. (1965). Predictability and immediate memory for consonant sequences. *Quarterly Journal of Experimental Psychology,* **17,** 175-177.

Baddeley, A. D. (1968). How does acoustic similarity influence short-term memory? *Quarterly Journal of Experimental Psychology,* **20,** 249-265.

Bartlett, F.C. (1932). *Remembering: An Experimental and Social Study*. Cambridge: Cambridge University Press.

Beiser, D. G. & Houk, J. C. (1998) Model of cortical-basal ganglionic processing: encoding the serial order of sensory events. *Journal of Neurophysiology*. **79**, 3168-88.

Botvinick & Plaut (in preparation). Short-term memory for serial order: a recurrent neural network model.

Botvinick, M. & Plaut, D. C. (submitted). Doing without schema hierarchies: a recurrent network model of routine sequential behavior.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review,* **107,** 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,** 551-581.

Cleeremans, A. (1993). *Mechanisms of implicit learning: connectionist models of sequence processing*. Cambridge, MA: MIT Press.

Dominey, P., Arbib, M. & Joseph, J.-P. (1995). A model of corticostriatal plasticity for learning oculomotor associations and sequences. *Journal of Cognitive Neuroscience*. **7,** 311-336.

Elman, G. (1990). Finding structure in time. *Cognitive Science,* **14,** 179-211.

Hartley, T. & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language.* **35,** 1-31.

Henson, R. N. A. (1998). Short-term memory for serial order: The Start-End Model. *Cognitive Psychology,* **36,** 73-137.

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Mellish & M. Zock., (Eds.), *Current research in natural language generation* (p. 287-319). London: Academic Press.

Kantowitz, B. H., Ornstein, P. A. & Schwartz, M. (1972). Encoding and immediate serial recall of consonant strings. *Journal of Experimental Psychology.* **93,** 105-110.

Marcus, M., Santorini, B. & Marcinkiewicz, M. (1993) Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics,* **19,** 313-330.

Mayzner, M. S. and Schoenberg, K. M. (1964). Single-letter and digram frequency effects in immediate serial recall. *Journal of Verbal Learning and Verbal Behavior.* **3,** 397-400.

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

O'Reilly, R. C. & Soto, R. (2002). A Model of the Phonological Loop: Generalization and Binding. In T. G. Dietterich, S. Becker & Z. Ghahramani, (Eds.), *Advances in Neural Information processing Systems (NIPS),* 14, , MIT Press, Cambridge, MA.

Rhode, D. (1999). *LENS: The light, efficient network simulator.* Technical Report CMU-CS-99-164, Department of Computer Science, Carnegie Mellon University, Pittsburgh, PA. (Available at http://tedlab.mit.edu/~dr/Lens/)

Rumelhart, D. E., Durbin, R., Golden, R. & Chauvin, Y. (1996). Backpropagation: The basic theory. In P. Smolensky, M.C. Mozer, & D. E. Rumelhart (Eds.) *Mathematical perspectives on neural networks* (p. 533-566). Hillsdale, NJ: Lawrence Erlbaum.

Rumelhart, D. E. & McClelland, J. L. (1986). Parallel distributed processing: explorations in the microstructure of cognition. Cambridge, MA: MIT Press.

Wickelgren, W. A. (1966). Associative intrusions in short-term recall. *Journal of Experimental Psychology,* **72,** 853-858.