To appear in H. Bowman & C. Labiouse (2004). Proceedings of the 8th Neural Computation and Psychology Workshop. WorldScientific Press.

AN EXTENDED BUFFER MODEL FOR ACTIVE MAINTENANCE AND SELECTIVE UPDATING

EDDY J. DAVELAAR & MARIUS USHER

School of Psychology, Birkbeck College, University of London, Malet street, London, WC1E 7HX, United Kingdom

In previous work, we developed a neurocomputational model of list memory, based on neural mechanisms, such as recurrent self-excitation and global inhibition that implement a short-term memory activation-buffer. Here, we compare this activation-buffer with a series of mathematical buffer models that originate from the 1960s, with special emphasis on presentation rate effects. We then propose an extension of the activation-buffer to address the process of selectively updating the buffer contents, which is critical for modeling working memory and complex higher-level cognition.

1. Introduction

Many models of human memory have been developed in psychology since the early 1960s[1,2] addressing a variety of tasks such as the immediate free recall. Most of these models were abstract-mathematical (rather than neuro-computational models) and their advantage is being simple and transparent, thus easy to understand. Recently, a shift towards neurocomputational models is taking place[3-9], which due to their increased complexity can account for a wider range of data including the effects of neuropsychological dissociations[6,10]. Nevertheless such models are more complicated and therefore more difficult to understand.

Here we start (section 2) by comparing our previous neurocomputational model of active memory with a series of buffer models, suggesting a way to reduce it (or extend the buffer models) so as to capture some important data in immediate free recall. In section 3, we propose ways in which our activation-buffer could be extended in order to address working memory processes, such as selective updating.

2. Mathematical and Neurocomputational Buffer Models

In the field of memory research, the free recall paradigm has led to many theoretical viewpoints and debates. In the immediate free recall paradigm, participants are required to report, in any order, as many items from a list that has been presented to them. The typical result is better recall performance for items that were presented at the beginning and at the end of the list, the primacy and (S-shaped) recency effect[11], respectively. One view of the recency effect is that the end-of-list items still reside in a limited-capacity short-term buffer from which they are reported without error[1,2,4-6,12]. In this section, we compare mathematical buffer models, which have been used in early psychological theories to explain free recall performance, with our neurocomputational activation-buffer, with a special emphasis on the effect of presentation rate.

The models are compared on four measures. First, the serial position functions present the probability that an item is in the buffer at the end of a sequence of twelve items (1st column in Figures 1 and 4). Second, we compare the distributions of the number of items in the buffer at the end of the sequence (2nd column in Figures 1 and 4). Third, we compare the probability that a new item will enter the buffer as function of the presentation rate and the number of items already in the buffer (3rd column in Figures 1 and 4). This comparison will turn out to provide valuable information related to the effect of presentation rate. The fourth and last measure on which the buffer models are compared is the distribution of probabilities that an item will be displaced from the buffer as a function of the number of items already in the buffer and the relative recency of the displaced item (4th column in Figures 1 and 4).

2.1. *Mathematical Buffer Models*

Three mathematical buffer models that have been used in the psychological literature are the *random-buffer*[1,2], the *knock-out buffer*[1,13] and the *variable knock-out buffer*[14]. Due to space-limitations a thorough analysis including other buffer models will be left for a future project.

Random-buffer model (RB)

The first buffer model is that in which the buffer consists of a fixed number of slots, r. When the buffer is full to capacity, a displacement process randomly (and with equal probability) selects which of the r slots will be emptied and be occupied with the newly presented item. The top row of Figure 1 shows the results of such a model. The left panel shows the serial position function for a buffer with capacity 3 and with capacity 4. These are exponential functions with base (r-1)/r and exponent -(sp+1), where sp indicates the recency of the item (-1 being the most recent). The second panel shows the distribution of the effective capacity at the end of a sequence of twelve items. As this is a fixed-capacity buffer, the distribution is centred on r. The third panel shows the probability that a presented item will enter the buffer as a function of presentation duration and number of items already in the buffer. By definition, all the mathematical buffer models described here have a probability of unity that an item enters the buffer, regardless of presentation duration and current buffer contents. The right-most panel shows the distribution of the probabilities that a buffer-item will be displaced from the buffer, as a function of its relative recency (the item that has been in the buffer the longest will have a relative recency of –r, whereas the latest addition to the buffer contents has a relative recency of –1). No surprise here that with random displacement, the distribution is uniform with probability 1/r.

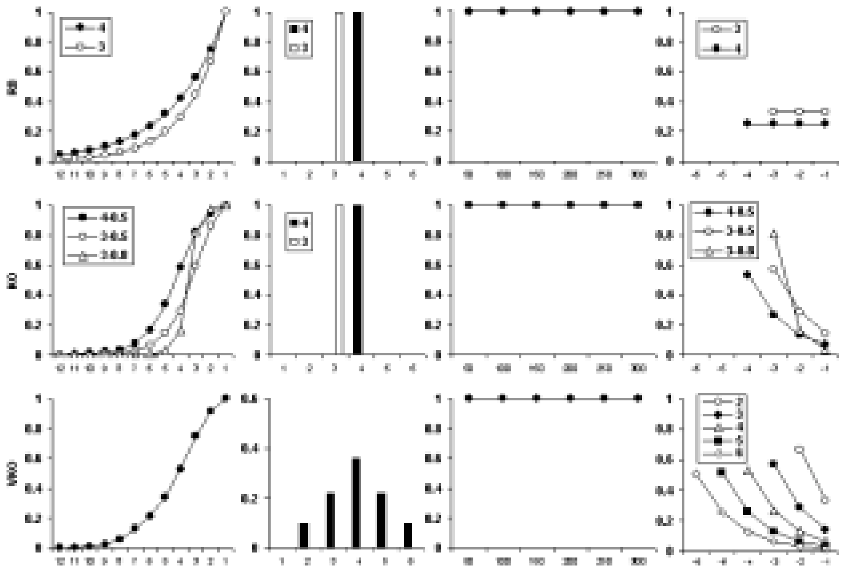


Figure 1. Comparisons of three mathematical buffer models. From top to bottom, results are presented for the random-buffer (RB), the knock-out buffer (KO) and the variable knock-out buffer (VKO). The results show serial position functions at different levels of capacity and displacement parameters (1st column), distribution of the number of items in the buffer after a twelve items sequence (2nd column), probability of a new item entering the buffer as a function of presentation duration (see abscissa) and number of items already in the buffer (3rd column) and the distribution of displacement probabilities as a function of the number of items in the buffer and the relative recency (4th column).

Knock-out buffer model (KO)

A variant of the random-buffer model is one in which the displacement process is such that the probability that an item is displaced from the buffer increases with the duration that the item has been in the buffer. This has been referred to as the knock-out buffer model[1,13]. The probability, d_i, of displacing item i, depends on the capacity r and a parameter δ that governs the slope of the displacement distribution (see Equation 1). The second row of Figure 1 shows the results. In the first panel, serial positions are presented for the model with capacity 3 and 4 and with δ=0.5 and δ=0.8. What is immediately apparent is that all serial position functions are S-shaped and that this increases with δ (compare 3-0.5 with 3-0.8). The distributions of the capacity (second panel) and the probability of entering the buffer (third panel) are the same as the random-buffer model. The right-most panel presents the distribution of displacement probabilities, which is a clear departure from the random process.

$$d_i = \delta(1-\delta)^{i-1}/[1-(1-\delta)^r] \qquad \text{(Eq. 1)}$$

Variable knock-out buffer model (VKO)

The third buffer model is one that extends the knock-out buffer and in which for every trial in a simulation the capacity r is drawn from a distribution of capacities[14]. This has the benefit of allowing more flexibility, as a participant's effective capacity may also depend on internal fluctuations in attention. The third row in Figure 1 shows the results of the variable-knock-out buffer model. The serial position function is basically a weighted aggregate of the various knock-out buffers within it. The second panel shows the distribution of capacities from which r was drawn. As an item always enters the buffer (third panel), the distribution of the displacement probabilities (fourth panel) is a collection of distributions at various capacities (which are all at unity).

2.2. *Neurocomputational Activation-Buffer Model*

We[4-6] developed a neurocomputational model of immediate free recall that is formulated within the Hebbian framework[15,16] with short-term memory (or primary memory[16]) being mediated by the current set of activated neuronal representations and long-term memory (or secondary memory[16]) being mediated by the connections between the activated subset and an episodic contextual system.

Figure 2 presents the architecture of our neurocomputational model. Each unit represents an assembly of interconnected neurons. When a stimulus is presented to the system, the corresponding representation will receive sensory input and increases in activation. To simplify, we use a single parameter, α, for the self-excitation. The self-excitation enables the unit to remain active above threshold after the sensory input has been taken away. Within the system, there is a global competition. This can be considered as originating from a general pool of inhibitory inter-neurons and has the effect of limiting the number of representations that can be active simultaneously.

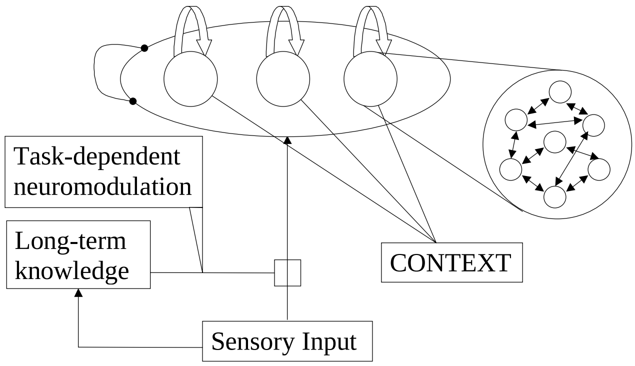


Figure 2. Architecture of our neurocomputational approach to list memory. The ellipse represents the activation-buffer, with units representing cell-assemblies and is addressed in section 2. The arrow ending in closed circles denotes global inhibition. The units form episodic links with a context representation. Sensory input goes directly into the activation-buffer after being neuromodulated. The specific architecture of how the neuromodulation is driven is arbitrarily chosen and does not change the discussion on selective updating in section 3.

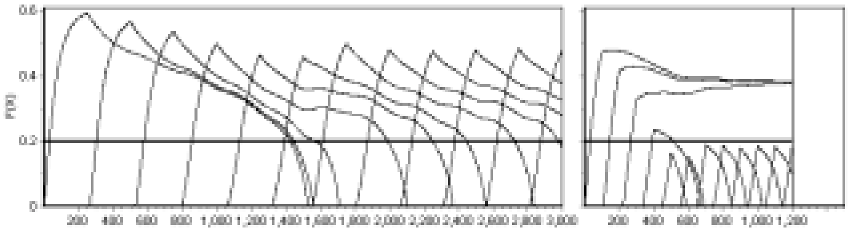
Our model is used in real-time, where all units are updated at every time-step according to a leaky integrator differential equation of which Equation 2 is the numerical solution in discrete time steps.

$$x_i(t+1) = \lambda x_i(t) + [1-\lambda][\alpha F(x_i(t)) - \beta \sum F(x_j(t)) + I_i(t) + \xi] \qquad \text{(Eq. 2)}$$

Here, λ=0.98 is the decay constant, α=2.0 the self-excitation, β=0.15 the global inhibition, I(t)=0.33 (0, when no input is presented) the sensory input at time t and ξ the zero-mean Gaussian noise with standard deviation σ=0.5. F(x) is the output activation function[8] MAX[0, x/(1+x)], which is similar to the threshold-linear function with the addition of a saturation non-linearity.

We also assume that units that are activated above threshold are encoded in episodic memory, which comprises of a matrix of Hebbian connection weights between the items and a context system. However, here we focus primarily on the dynamics of the activation-buffer, which are illustrated in Figure 3. Twelve units are sequentially presented with sensory input for 250 (left panel) or 100 (right panel) iterations corresponding to a typical experimental procedure where a list of twelve items are presented sequentially on a computer screen at different presentation rates. Each line corresponds to the output activation, $F(x_i)$, of a given unit i. The left panel of Figure 3 shows the set of activation trajectories when the presentation rate is relatively slow and the right panel shows the set of trajectories at fast presentation rate. Two aspects can be observed. First, units remain active after stimulus offset, which is due to the self-recurrent excitation.

Second, several units can be active simultaneously and there is an upper limit to the number of units that are active above threshold, which reflects the capacity



limitation due to global inhibition.

Figure 3. Activation trajectories of twelve sequentially activated units at slow (left) and fast (right) presentation rates. Time-steps are set along the abscissa and the output activation on the ordinate. The horizontal line [at F(X)=0.2] represents the activation threshold above which an item is said to be in the buffer.

With increase in the presentation rate, the activation-buffer changes its behaviour. First, the units reach a lower level of activation compared to the condition with slow presentation rate. This merely reflects the limited time that is given for the units to accrue. Second, with the same structural parameters, the number of active units is smaller than the number of active units at slow presentation rates, implying that the effective capacity of the activation-buffer depends on external variables like presentation rate. Elsewhere[4], we have shown that the system will not exceed a certain upper limit given a wide range of presentation rates[a]. Third, whereas at slow presentation rates the unit to be displaced (de-activated) from the current buffer contents, is typically one that has been in the buffer (above threshold) the longest, at fast presentation rates the buffer only maintains the first few items and blocks out any subsequently presented item. In other words, at low presentation rates, the activation-buffer is a limited-capacity buffer system with a knock-out displacement process, while at fast presentation rates the probability of entering the buffer is greatly diminished. This prediction is fully due to the limited time available that a unit can be activated to the extent that it can overcome the amount of inhibition already in the system. As the first item enters an empty buffer, it will not have to overcome this sort of inhibition, giving it an advantage over subsequent activated units.

2.3. *The Effect of Presentation Rate on Buffer Dynamics*

Figure 3 shows activation trajectories for the activation-buffer under slow and fast presentation rates. In the top row of Figure 4 the results of the activation-buffer are presented on the same four measures we examined for the three

---

[a] In fact, the effective capacity shows an inverted U-curve with presentation rate.

mathematical buffer models, so that it can be compared with them (cf. Figure 1). First of all, the left panel shows the serial position functions for three presentation rates (here presented as durations). With slow presentation rates (250 iterations per item), the serial position function is recency-biased and S-shaped, whereas with intermediate rates (150 iterations per item), the function is recency-biased, J-shaped (exponentially-shaped) and some primacy items are maintained. However, with fast presentation rates (50 iterations per item), the serial position function is primacy-biased and J-shaped. It is important to remember that the serial position functions represent the probability of items presented at that position in the sequence still being active above threshold. No Hebbian weight-changes or other long-term memory processes are incorporated. This switch from recency to primacy with increase in presentation rate was verified in an experiment[6]. The activation-buffer maintains less items under fast than under slow presentation rates, as indicated by the shift in the distribution of the number of active items at the end of a twelve-item sequence. As mentioned before, in this range of presentation rates, the effective capacity is negatively correlated with the presentation rate.

Two major differences between the activation-buffer and the mathematical buffers were observed. First, for the activation-buffer, the probability that an item will enter the buffer depends on the presentation rate and the number of items already in the buffer. In the activation-buffer, increasing the presentation rate decreases the probability that a unit can be activated to such a level at which it can overcome the inhibition in the system, which increases with the number of items already in the buffer. This dual-relationship leads to the complex interaction depicted in the third panel. Second, with slow presentation rates the distribution of displacement probabilities for the activation-buffer suggest a knock-out displacement process (see fourth panel). With fast presentation rates the distribution becomes more flat (not shown). This suggests a displacement rule that is rate-dependent, such that with faster presentation rates δ decreases. We focus on extending the knock-out buffer with the rate-dependent probabilities that a presented item will enter the buffer and decrease δ for fast presentation rates.

2.4. *Extending the Knock-Out Buffer*

The above comparisons seem to suggest that the main reason why the mathematical buffers do not predict the shift from a recency-biased to a primacy-biased serial position function with increase in presentation rate is that in those models an item *always* enters the buffer. Although in the original Atkinson and Shiffrin[1] buffer model, a parameter was included that governed the probability of entering the buffer, simulations estimated its value at around unity, which is consistent with the activation-buffer at slow presentation rates. Re-introducing the parameter and making it dependent on the number of items already in the buffer and presentation rate would allow the mathematical model to accommodate the recency-to-primacy shift.

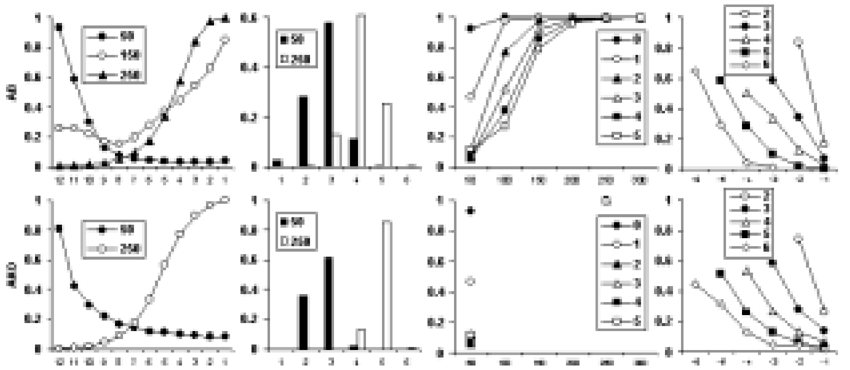


Figure 4. Results for the activation-buffer (AB; top row) and the activation knock-out buffer (AKO; bottom row) on the four measures for slow (250 iterations per item) and fast (50 iterations per item) presentation rates. For the activation-buffer, an intermediate presentation is also shown, indicating a gradual transition from recency-to-primacy bias. Note that for the activation knock-out buffer, the probability of entering the buffer at the two presentation rates are taken directly from those of the activation-buffer. The distribution of displacement probabilities as a function of relative recency is only shown for the slow presentation rate. In the AKO buffer, δ_{slow}=0.5 and δ_{fast}=0.01.

To test this assumption, we added a parameter to the knock-out buffer. We chose to extend the knock-out buffer as it contains the right kind of assumptions that lead to S-shaped serial position functions. Although we used the probabilities obtained with the activation-buffer, we did notice that the relationship between the probability of entering the buffer, the presentation rate and the current capacity can be approximated with a single sigmoidal function. Here, we are only interested in whether adding the probabilities will produce the two main predictions from the activation-buffer. As can be seen in the bottom row of Figure 4, adding the probabilities allows the model to predict the recency-to-primacy shift (first panel) and the decrease in effective capacity with increase in presentation rate (second panel). The rightmost panel shows the distributions of displacement probabilities, which are similar to those of the variable knock-out (third row, Figure 1) and activation-buffer (top row, Figure 4). It is important to realise that the variability in the effective capacity is a consequence of the probabilities that a newly presented item enters the system and the probabilities that an item is displaced from the system.

This exercise suggests that the initial conception of the knock-out buffer with the additional "entry-parameter" by Atkinson and Shiffrin[1] contained the relevant assumptions to predict the recency-to-primacy shift. These assumptions in turn follow naturally from the complex dynamics of the activation-buffer. To summarise, the activation-buffer shows that δ and the entry-probability are inversely related to the presentation rate.

3. Selective Updating of the Buffer

The neurocomputational activation-buffer captures the complex dynamics of short-term memory that are needed to explain the data found in immediate free recall. Within this neurocomputational level of description, it is possible to model the dynamical process of updating the contents of the buffer in accordance with a given task set, as needed to account for cognitive control and working memory[3,7]. The updating task we examine here is one in which a sequence of concrete and abstract nouns is presented with the instructions to remember only those words that represent small things[17]. For example, in the sequence *car, desk, idea, key, plane, staple, giraffe* only the words *key* and *staple* need to be reported. In this example, it is not until *key* is presented that one knows that *car* and *desk* belong to the category of large things and the contents of the buffer is to be updated. However, when *plane* is presented it is already apparent that this belongs to the large-things category and will not even enter the buffer.

As in previous work[3,18], we assume that neuromodulation of sensory input introduces sufficient flexibility to support task-dependent selective updating. In Figure 2 the architecture illustrates a configuration that could lead to task-dependent neuromodulatory control. Sensory input enters the activation-buffer and activates long-term knowledge about the presented item, such as magnitude. With the instruction that small things need to be maintained, words representing small things will provide larger modulated input to the buffer than words representing large things or abstract nouns. In order to capture the essence of the neuromodulation, we represented a sequence of concrete nouns as a sequence of items that vary in the amount of input (I_{target}=0.33, $I_{non\text{-}target}$=0.21). A more detailed model of selective updating with an actual implemented neuromodulatory system is due to space limitations left for future work.

In the left-hand side of Figure 5 the activation trajectories are shown for a sequence of twelve nouns in which nouns 4, 5 and 6 are target nouns and all others are non-targets. As can be seen, the model maintains the first three non-targets until the three targets are presented. After the three targets, none of the non-targets displace the target items: the system has updated the current contents and maintains the targets in the face of distractors. This is due to the targets receiving sufficient effective input to overcome the inhibition driven by the initial non-targets, whereas the non-targets presented after the targets do not receive enough effective input to overcome the inhibition that is then driven by the targets.

In our work on free recall memory[4-6], we assumed that in addition to maintenance in the activation buffer, Hebbian connections are formed between items that are active above threshold and an episodic context system. The strength of these connections is proportional to the integral under the activation trajectories and the threshold. At retrieval, participants can report items from the buffer or trigger a slower competitive retrieval process that uses the episodic Hebbian connections. In Figure 5, it can be seen that non-targets presented before the targets will have stronger episodic connections than the many non-targets presented after the targets, which could lead to more intrusions of non-targets presented before the targets than of those presented after the targets, as reported by Palladino and co-workers[17] (right-hand side of Figure 5).

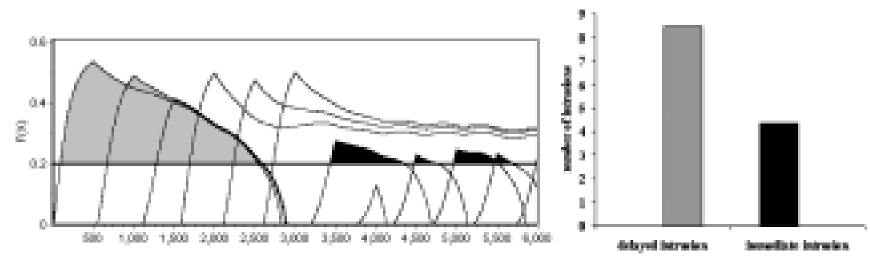


Figure 5. Left: Activation trajectories for a sequence of twelve items in which items 4, 5, and 6 are target items. Note that the targets displace the preceding non-targets and that non of the subsequent non-targets are maintained. The shaded areas correspond to the episodic strengths for the non-targets presented before (grey) and after (black) the target items. Right: Results from Palladino and coworkers on the number of non-target intrusions. Delayed intrusions are before-target non-targets and immediate intrusions are after-target non-targets.

4. Conclusion

In this paper, we compared our neurocomputational activation-buffer with a series of mathematical buffers used in the earlier literature. We found that these buffer models were lacking the flexibility needed to enable them to predict presentation rate effects and we proposed an extension of the knock-out buffer, which may be seen as a reduction of our activation model. We suggest that this illustrates how starting from neurocomputational principles (before reducing to an abstract model) may be productive in modeling psychological processes, since it can ground relatively arbitrary assumptions (in this case the buffer properties). For example, the buffer properties and its capacity limitation follow from mechanisms of recurrent self-excitation (interconnectivity of neurons within an assembly) and the global inhibition (originating from a pool of interneurons). This balance is dynamic and leads to a distribution of the capacities instead of a single capacity value and is affected by external factors like presentation rate, leading to the recency-to-primacy shift.

We have also presented a conceptual extension to the activation-buffer that addresses processes, such as selective updating of the buffer contents. Recently, we[10] showed how the model can account for deviant serial position functions found with neuropsychological patients. We believe that a neurocomputational approach to (short-term) memory not only allows a way to understand how neural principles underlie cognitive behaviour, but also provides a promising platform on which natural extensions can allow for more complex higher-level cognitive processes to be addressed.

Acknowledgments

This work is supported by the Economic and Social Research Council (T026271312). Send correspondence to e.davelaar@psychology.bbk.ac.uk or m.usher@bbk.ac.uk.

References

1. Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: a proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation, Vol 2.* New York: Academic Press.
2. Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: a theory of probabilistic search of associative memory. In G. Bower (Ed.), *The psychology of learning and motivation, Vol 14*. New York: Academic Press.
3. Braver, T. S., & Cohen, J. D. (2000). On the control of control: the role of dopamine in regulating prefrontal function and working memory. In S. Monsell & J. Driver (Eds.), *Attention and Performance XVIII. Control of cognitive processes*. Cambridge, MA: MIT Press.
4. Davelaar, E. J. (2003). *Active memory: its nature and its role in mnemonic and linguistic behaviour*. Unpublished doctoral dissertation. London: Universtiy of London.
5. Davelaar, E. J., & Usher, M. (2002). An activation-based theory of immediate item memory. In J. A. Bullinaria & W. Lowe (Eds.), *Connectionist models of cognition and perception: proceedings of the seventh neural computation and psychology workshop.* Singapore: World Scientific.

6. Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., & Usher, M. (submitted). A context/activation model of list memory: dissociating short-term from long-term recency.
7. O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically-based computational model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press.
8. O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience. Understanding the mind by simulating the brain.* Cambridge, MA: MIT Press.
9. Taylor, J. G., & Taylor, N. R. (2000). Analysis of recurrent cortico-basal ganglia-thalamic loops for working memory. *Biological Cybernetics, 82*, 415-432.
10. Davelaar, E. J., & Usher, M. (2003, July). *Modeling serial position functions of neuropsychological patients*. Poster presented at the meeting of the Experimental Psychological Society. Reading, UK. Available at http://www.geocities.com/ejdavelaar
11. Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Verbal Learning and Verbal Behavior, 64*, 482-488.
12. Glanzer, M. (1972). Storage mechanisms in recall. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation, Vol 5*. New York: Academic Press.
13. Philips, J. L., Shiffrin, R. M., & Atkinson, R. C. (1967). Effects of list length on short-term memory. *Journal of Verbal Learning and Verbal Behavior, 6*, 303-311.
14. Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory and Cognition, 24*, 103-109.
15. Hebb, D. O. (1949). *The organization of behavior: a neuropsychological theory*. New York: Wiley.
16. James, W. (1890). *Principles of psychology*. New York: Henry Holt.
17. Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory and Cognition, 29*, 344-354.
18. Usher, M., & Davelaar, E. J. (2002). Neuromodulation of decision and response selection. *Neural Networks, 15*, 635-645.