Modeling prefrontal and medial temporal contributions to episodic memory.

Sean M. Polyn, Kenneth A. Norman & Jonathan D. Cohen
Department of Psychology, Princeton University, Princeton, NJ, USA

Overview

• A connectionist model of episodic memory captures many aspects of free recall performance in humans.

• The model of prefrontal cortex (PFC) implements a drifting context vector.

• During encoding, the model of hippocampus forms an association between an item representation and the current PFC state.

• During recall, PFC uses hippocampus to retrieve prior contextual states; PFC probes memory with these states.

• We explore various types of damage to the model with an eye towards capturing the deficits seen in the elderly.

Introduction

Evidence from neuroimaging and lesion studies indicates that both medial temporal lobe (MTL) structures and prefrontal cortex (PFC) contribute to episodic memory performance.

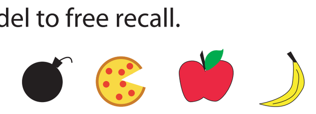
In this work, we present a computational neural network model of how interactions between MTL and PFC support performance on free recall tests. The MTL component of our model consists of hippocampal and perirhinal networks; this Complementary Learning Systems model has already been used to account for a wide range of recognition memory findings (Norman & O'Reilly, in press).

The PFC component of the model is based on recently developed models of how PFC supports cognitive control. These models (Frank, Loughry, & O'Reilly, 2001) posit that PFC actively maintains aspects of presented stimuli via multiple parallel "stripes" that can be updated separately. On each trial, information in PFC can be maintained, or can be replaced by aspects of the current stimulus. Over time, the pattern of activity in PFC can be viewed as an evolving "context vector".

At study, the current state of this PFC context vector is associated with item representations via the hippocampus. At test PFC context can cue memory for studied items. At a high level, this approach has much in common with more abstract models of temporal context memory (e.g., Howard & Kahana, 2002).

The paradigm

We apply the model to free recall.

Given study list: 

Say the subject recalls: "apple, pizza, banana"
That is, they recall item #3, then #2, then #4.

• Probability of First Recall (PFR) provides a measure of how subjects *initiate* recall.

We calculate the probability that a given item will be recalled *first* by the subject. Since subjects tend to initiate recall with end of list items, PFR provides a good measure of recency.

Here the first item recalled was "apple".

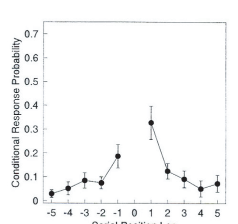
• Conditional Response Probability (CRP) provides a measure of how subjects carry out recall, once initiated.

CRP measures *transition probabilities*. Given that the subject recalled the N^{th} item from the study list on the previous recall attempt, we calculate the probability that the subject will recall item N+1, N-1, etc. on the current recall attempt.

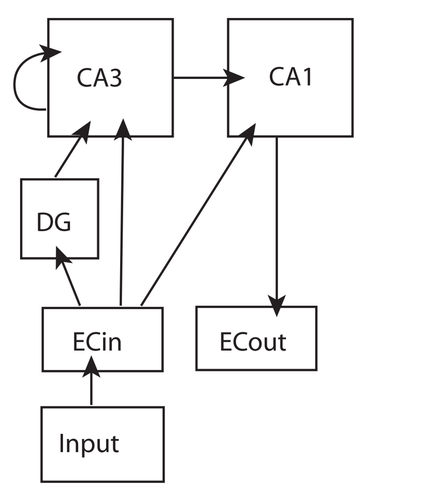
The transitions made in the example above were -1 (from apple to pizza) and +2 (from pizza to banana).

A sample CRP curve. Nearby items are more likely to be recalled on successive recall attempts.

Figure courtesy of M. Kahana



The hippocampal model



• Creates distinct representations for similar stimuli.

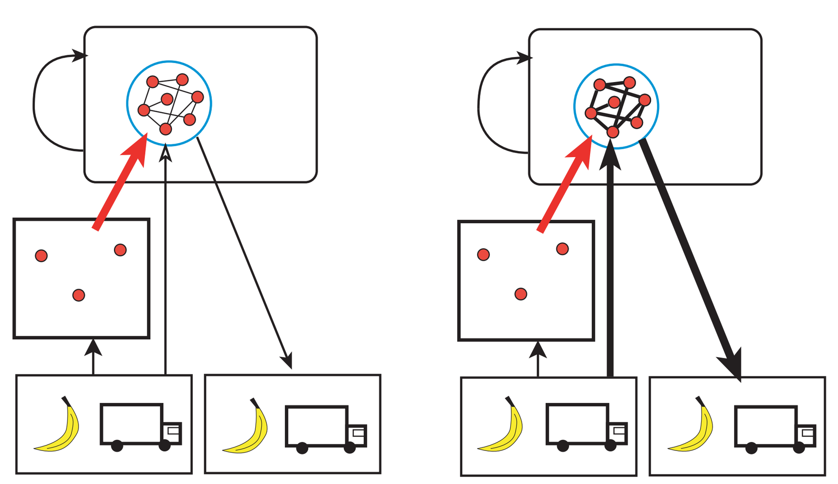
• Binds together features of an episode.

• Supports pattern completion and detailed recollection.

• Integrates information from diverse brain regions; both posterior and prefrontal areas.

• Changes in connection strength between encoding and retrieval follow those investigated by Hasselmo, Bodelon & Wyble (2002).

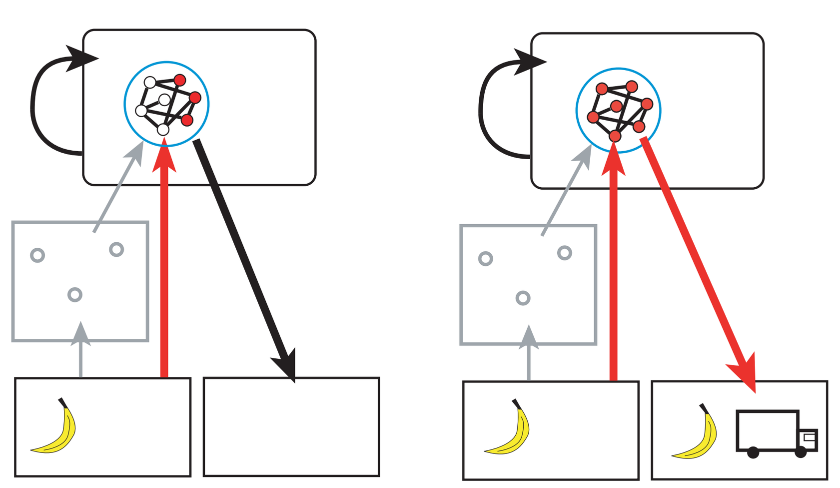
Hippocampus at encoding



A truck full of bananas drives past. The semantic representations of banana and truck activate a set of units in CA3. These activated CA3 units strengthen connections between each other, as well as to active input and output units via Hebbian learning.

During encoding:
• Strong Dentate Gyrus (DG) projections promote pattern separation. DG has many units and is sparsely active; this enhances small differences in input patterns.
• Once the strong DG projections determine which units are active in CA3, Hebbian weight change forms an attractor by linking together active units.

Hippocampus at retrieval



Later, you notice a banana on your desk, which activates the semantic representation of banana. This activates a subset of the CA3 units from encoding. The strengthened weights between the units allows for pattern completion to the original attractor. You recall the episode with the banana truck.

During retrieval:
• Weak Dentate Gyrus projections allow CA3 recurrent connections to promote pattern completion.
• Given an item stimulus, the hippocampus can retrieve any associated information.

Many existing models of memory retrieval posit unsticking mechanisms to prevent the model from recalling the same item again and again (inhibition of return). In the present model, when items are retrieved, a bias weight projecting to the active units becomes slightly more negative. Over a short time scale (2-3 trials) the bias weight decays to zero.

The context vector

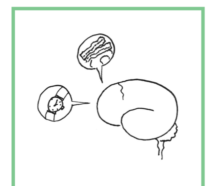
Our basic theory is that internal contextual states maintained by PFC (Schacter, 1987) can be used to target memory traces in medial temporal cortex. There have been a number of theories of the role of temporal context in memory over the last 50 years.

Early theories of temporal context memory posited a context vector that randomly drifts over time (Estes, 1955; Mensink & Raaijmakers, 1988).

In more recent theories (Howard and Kahana, 2002), the evolution of the context vector is driven by information associated with the studied items. This is also true of our model.

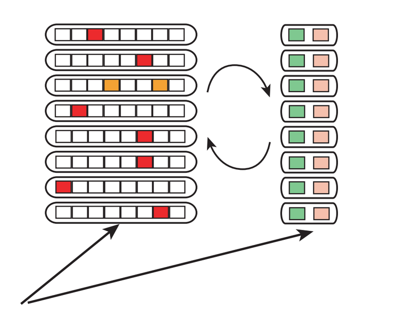
We believe that during stimulus processing PFC grabs features of items and actively maintains them; this information is used to bias processing in posterior areas (Miller & Cohen, 2001).

Crucially, some information about prior states remains in PFC upon presentation of subsequent items. This allows PFC to act as a drifting context vector. The mechanisms are covered in more detail below.



Internal PFC states are associated with the current episode.

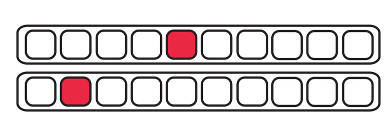
The Prefrontal Model



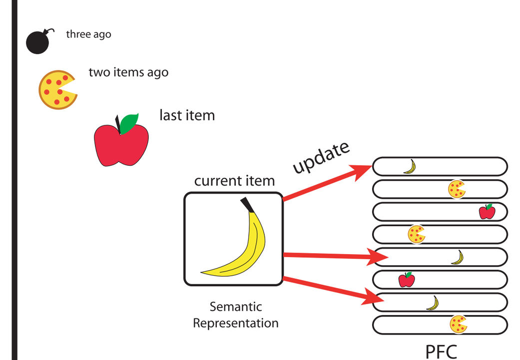
• Activity in PFC projects to posterior areas and biases processing in those areas.

• PFC maintains features of item stimuli.

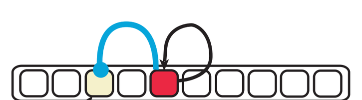
• PFC representations are robust to interference but can rapidly update when needed.



In the model, each stripe is an attractor system with 10 stable states. Each state can be thought of as a high-level representation of item features.

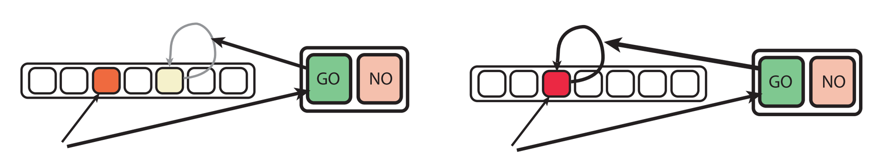


PFC can maintain the features of multiple items simultaneously. By gating a few features at a time, the overall representation in PFC changes slowly.



A stripe can maintain its current state or update to reflect new information in the environment. Maintenance is accomplished through a self connection and lateral inhibition. Here, a potential input is blocked by maintained activity in the stripe.

Each stripe has a gate, which consists of a GO unit and a NO unit. The GO unit acts as a toggle which can unlock or lock the gate. If the NO unit is activated, nothing happens.

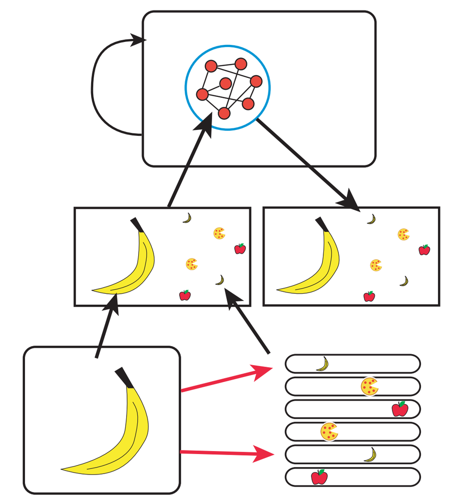


Inputs project to both the stripe and the gate. If the input activates the GO unit, the current self connection is removed, allowing the new input to gain access to the stripe. The input then activates the GO unit again, which locks the new input.

The interaction

The model can be used to simulate the operations undertaken by PFC to manipulate the memory system. We use the combined PFC/MTL model to investigate the dynamics of free recall.

Encoding Mode



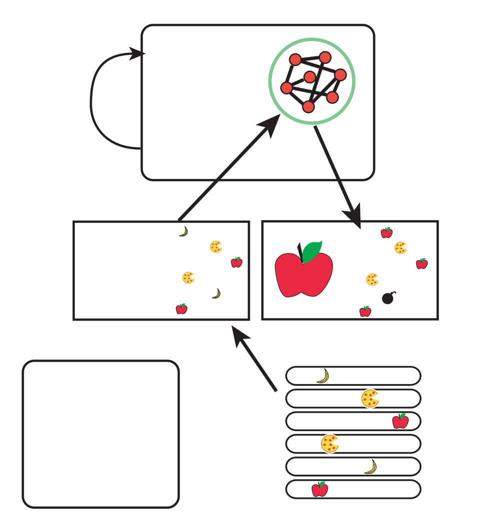
During study, items are presented to the model for encoding.

• PFC gates in features of items and the PFC state evolves over time.

• Semantic representations and PFC contextual states project to hippocampus.

• Hippocampus memorizes conjunctions of item and internal context information.

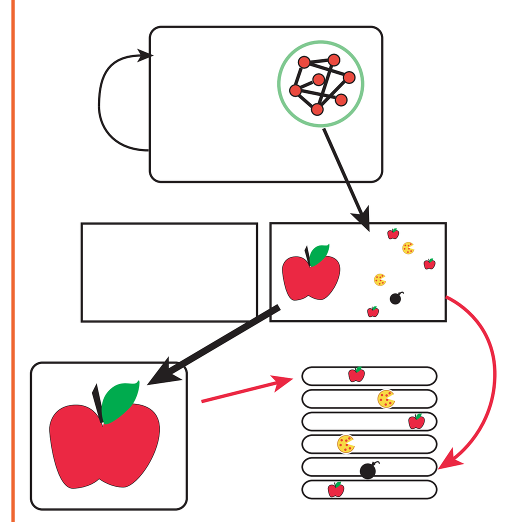
Retrieval Mode



At test, there is no external cue to drive recall.

• PFC is still maintaining the set of features that were present for the final study item (the end of list state).

• This information is projected to hippocampus as a retrieval cue. Since PFC state drifts slowly, the end of list state is a good cue for the last few items in the list

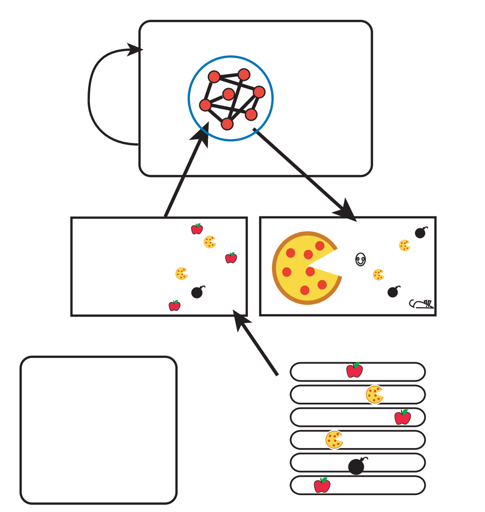


Here, the model successfully retrieves an item and its associated context. The defining characteristic of retrieval mode is that retrieved context is used to update PFC state.

• The retrieved context can then be used to update the PFC state.

• The item representation is reactivated in semantic memory.

• The retrieved semantic representation can also be used to update PFC in a similar way as at study.

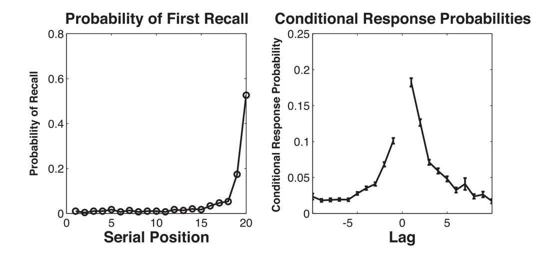


The newly updated PFC can attempt another hippocampal retrieval. The cycle repeats.

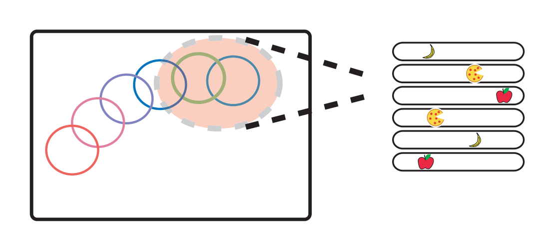
• In a sense, the model 'jumps back in time' (Tulving, 2002). Hippocampus reactivates a PFC state from earlier (during study), and uses that prior state to cue for items.

Applying the model to free recall

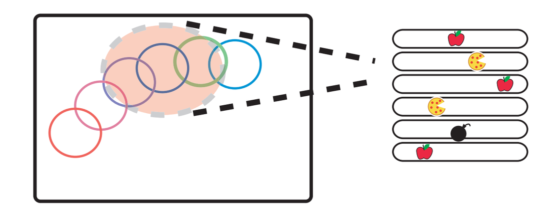
As elucidated by Kahana (1996), the Probability of First Recall (PFR) and Conditional Response Probability (CRP) provide all the information needed to follow the dynamic process of free recall.



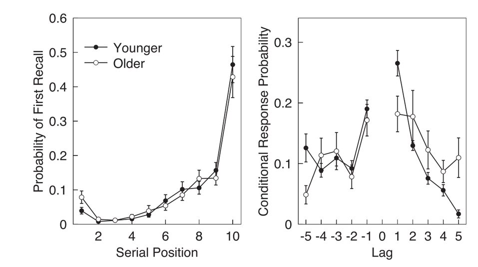
Simulation data (300 runs)
The model was trained on a set of 20 items, and was then put into retrieval mode. Testing consisted of 40 retrieval attempts.



The end of list PFC state acts like a spotlight searching memory. The first item recalled comes from the end of the list because the PFC state at the beginning of test is most similar to traces that were seen recently.



After a successful recall the spotlight recenters on the retrieved item. Subsequent recalls will come from the items studied close in time to the recalled item.



Empirical data (Howard, Wingfield and Kahana, submitted)
Younger and older subjects were run on a test of free recall. Overall, the older subjects recalled fewer items.

Figure reproduced with permission from M. Howard.

• The Probability of First Recall (PFR) curve. Notice that elderly subjects are ***unaffected*** in this measure. Thus, elderly subjects ***initiate*** recall appropriately.

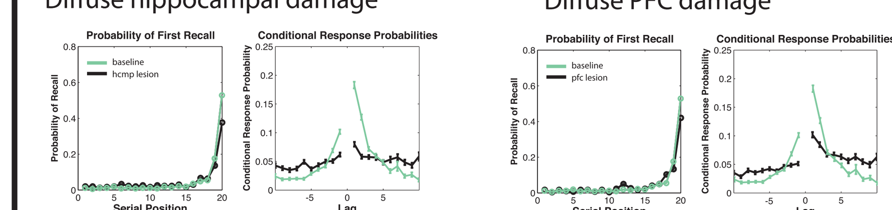
• The Conditional Response Probability (CRP) curve. In general, the CRP curve for the elderly is ***flattened***. Thus, elderly subjects ***do not transition*** appropriately.

• We explore various lesions to the model that might produce this pattern of data.

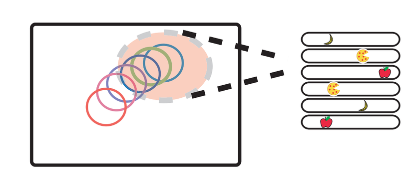
Lesioning the model

The undamaged PFR curve in the elderly is a very constraining data point, that cannot be captured with just any type of damage to the model. Here we investigate two types of broad damage: a lesion of 70% of the units in areas DG, CA3 and CA1 of the hippocampus, and a lesion of 70% of the units in the PFC.

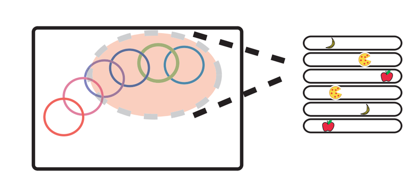
Diffuse hippocampal damage | Diffuse PFC damage



In each case the PFR and CRP curves are flattened. These lesions do not capture the pattern of deficits seen in the elderly



The hippocampal lesion reduces the representational capacity of the network. Thus, it pushes all of the hippocampal representations closer together. Now the spotlight shines on more memory traces at once, reducing its ability to sharply target traces.



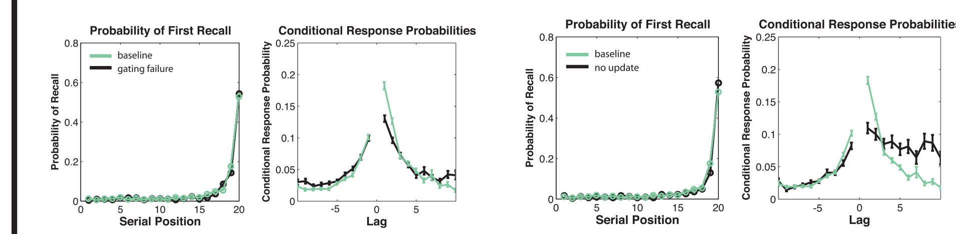
The PFC lesion reduces the sharpness and effectiveness of the retrieval cue. Thus, it broadens the spotlight; this is a similar deficit to the one seen above, but due to a different mechanism. Again, the spotlight shines on more traces at once, reducing its ability to sharply target traces.

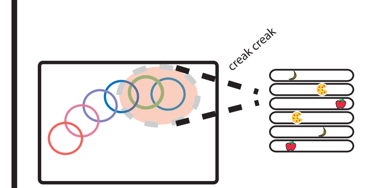
Both the hippocampal and PFC lesions alter the way the model ***initiates*** recall. This flattens the PFR curve. These lesions do not capture the elderly deficit. Any candidate lesion can not affect the first recall attempt. We turn our attention to a second class of lesions: those affecting the PFC gating system.

Lesioning the model

Capturing the elderly deficit in free recall.
• Howard, Wingfield and Kahana (submitted) capture the elderly deficit with a mathematical model by adding noise to the contextual retrieval process.
• We explore this possibility by introducing gating failures at test. On 50% of test trials, retrieved information fails to update the PFC. In the graphs below, it is clear that this has no effect on the PFR curve, but significantly flattens the CRP curve.

Gating failure (50%) | No PFC update at test





Gating failures make it hard to update the context vector with retrieved context. This is like a rusty spotlight that has difficulty being pushed about to search for memory traces. Recall is initiated normally, but then the model has trouble "jumping back in time".

• Even when the gating mechanism undergoes 100% failure (no PFC update at test), the model still shows an intact PFR curve. This is because PFC ***initiates*** recall normally.

• During study, PFC is updated by gating in features present in the environment. However, at test, PFC must update by gating in retrieved information that has no bottom-up support. A damaged gating system might not be driven reliably in this situation, where the information driving updating is weak.

• The dopamine (DA) system is intimately involved with the gating system in the brain, and a number of studies have shown disturbances in the DA system in healthy aged humans and monkeys (as reviewed in Braver et al, 2001). Thus, further investigation of this component of the model is underway.

Conclusions & future directions

We extend an existing model of medial temporal contributions to episodic memory by adding a prefrontal component. This combined model is used to investigate prefrontal and medial temporal interactions in episodic memory. Here, the model is used to simulate free recall performance.

The model is successful in capturing a number of aspects of free recall performance including how subjects initiate and carry out recall. This is achieved by the representation of internal context stored in the prefrontal component of the model.

Importantly, the model can inform us about the effects of various types of lesions. A set of model lesions are compared to the performance of healthy aged humans. Large diffuse lesions to the hippocampus and prefrontal cortex do not produce the correct pattern of elderly data. A dysfunctional gating system does produce the proper pattern of elderly data.

By including a functional PFC component in our model, we gain the ability to investigate a vast array of memory tasks, including list learning, interference paradigms, directed forgetting, recency judgements and temporal order.

We are also exploring how medial temporal cortex contributes to more classically "frontal" tasks such as task-switching.

References

Estes WK (1955). Statistical theory of distributional phenomena in learning. Psychological Review, 62, 369-377.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between the frontal cortex and basal ganglia in working memory: A computational model. Cognitive, Affective, and Behavioral Neuroscience, 1, 137–160.

Hasselmo ME, Bodelon C, Wyble BP (2002). A proposed function for hippocampal theta rhythm: separate phases of encoding and retrieval enhance reversal of prior learning. Neural Computation 14, 793-817.

Howard MW & Kahana MJ (2002). A distributed representation of temporal context. Journal of Mathematical Psychology, 46, 269-299.

Howard MW, Wingfield A, Kahana MJ (submitted). Modeling the age-related associative deficit in self-initiated recall.

Kahana MJ (1996). Associative retrieval processes in free recall. Memory & Cognition, 24(1), 103-109.

Kahana MJ, Howard MW, Zaromb F, Wingfield A (2002). Age dissociates recency and lag recency effects in free recall. Journal of Experimental Psychology: Learning, Memory & Cognition, 28(3), 530-540.

Mensink, G-JM, & Raaijmakers, JGW (1988). A model for interference and forgetting. Psychological Review, 95, 434-455.

Norman KA & O'Reilly RC. Modeling hippocampal and neocortical contributions to recognition memory: a complementary learning systems approach. Psychological Review, in press.

Schacter DL (1987). Memory, amnesia, and frontal lobe dysfunction. Psychobiology, 15(1), 21-36.

Tulving E (2002). Chronesthesia: conscious awareness of structured time. In DT Stuss & RT Knight (Eds.), Principles of Frontal Lobe Function (ch. 20). Oxford University Press: Oxford.

http://www.princeton.edu/~psych/PsychSite/compmemory