HHS Public Access
Author manuscript
Psychol Rev. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:
Psychol Rev. 2016 January ; 123(1): 23–69. doi:10.1037/rev0000015.

A Four–Component Model of Age–Related Memory Change

M. Karl Healey and **Michael J. Kahana**
University of Pennsylvania

Abstract

We develop a novel, computationally explicit, theory of age–related memory change within the framework of the context maintenance and retrieval (CMR2) model of memory search. We introduce a set of benchmark findings from the free recall and recognition tasks that includes aspects of memory performance that show both age-related stability and decline. We test aging theories by lesioning the corresponding mechanisms in a model fit to younger adult free recall data. When effects are considered in isolation, many theories provide an adequate account, but when all effects are considered simultaneously, the existing theories fail. We develop a novel theory by fitting the full model (i.e., allowing all parameters to vary) to individual participants and comparing the distributions of parameter values for older and younger adults. This theory implicates four components: 1) the ability to sustain attention across an encoding episode, 2) the ability to retrieve contextual representations for use as retrieval cues, 3) the ability to monitor retrievals and reject intrusions, and 4) the level of noise in retrieval competitions. We extend CMR2 to simulate a recognition memory task using the same mechanisms the free recall model uses to reject intrusions. Without fitting any additional parameters, the four–component theory that accounts for age differences in free recall predicts the magnitude of age differences in recognition memory accuracy. Confirming a prediction of the model, free recall intrusion rates correlate positively with recognition false alarm rates. Thus we provide a four–component theory of a complex pattern of age differences across two key laboratory tasks.

Keywords

aging; memory; modeling; free recall; episodic memory

As we grow older many aspects of our cognitive functioning change. Some aspects like verbal ability (Verhaeghen, 2003), social skills (Carstensen, Isaacowitz, & Charles, 1999) and possibly wisdom (Baltes & Staudinger, 2000) show welcome increases. Unfortunately, most aspects of cognition show age–related declines rather than improvements (Baltes & Lindenberger, 1997; MacDonald, Dixon, Cohen, & Hazlitt, 2004; Park et al., 2002; Zelinski & Burnight, 1997). Perhaps the most salient of these declines occur in episodic memory, as revealed both by self-reports (Newson & Kemps, 2006; Zacks, Hasher, & Li, 2000) and laboratory studies (Craik & Jennings, 1992; Kausler, 1994; Light, 1991; Moscovitch & Winocur, 1992; Provyn, Sliwinski, & Howard, 2007; Salthouse, 1991). The central role played by episodic memory in aspects of cognition ranging from autobiographical recall

Correspondence concerning this article should be addressed to M. Karl Healey (healeym@sas.upenn.edu) or Michael J. Kahana (kahana@psych.upenn.edu) at University of Pennsylvania, Department of Psychology, 3401 Walnut St., Room 303, Philadelphia.

(Levine, Svoboda, Hay, Winocur, & Moscovitch, 2002) to abstract reasoning (Healey, Crutchley, & Kahana, 2014; Mogle, Lovett, Stawski, & Sliwinski, 2008; Unsworth, 2009) make age–related declines particularly disturbing. Therefore, as the average life span lengthens, developing interventions to prevent, slow, or reverse age–related memory decline has become one of the most important challenges facing cognitive science.

Success in developing such interventions will require a detailed understanding of how and why memory changes with age. Over the last several decades the field of cognitive aging has provided a comprehensive description of *how* memory changes with age by comparing older and younger adults on a wide variety of memory tasks (for reviews see Craik, 1977; Craik & Rose, 2012; Kausler, 1994; Light, 1991; Lindenberger & Ghisletta, 2009; Salthouse, 1991; Zacks et al., 2000). The field, however, has not yet converged on a common theoretical account of *why* these changes occur; that is, which memory processes underlie the broad pattern of age–related change. Here we suggest a new approach for adjudicating among existing theories and developing new ones.

Précis of Our Approach

Not all aspects of memory change uniformly with age. The clearest example of this is a gradation in the magnitude of age-related impairment from recognition tasks, which show modest impairments in accuracy (Craik, 1971; Jacoby, 1999; Ratcliff, Thapar, & McKoon, 2004; Schonfield & Robertson, 1966; Spaniol, Madden, & Voss, 2006), to recall tasks which show large impairments (Craik, 1968; Hultsch, 1969; Schonfield & Robertson, 1966). Even within a task, such as free recall, detailed measures reveal that some aspects of performance remain stable whereas others show varying degrees of impairment (Kahana, Howard, Zaromb, & Wingfield, 2002). To date no theory has been shown to account for these gradations with quantitative precision.

Researchers have proposed a variety of theories (for very different contemporary accounts see Bender, Naveh-Benjamin, & Raz, 2010; Benjamin, 2010; Craik, Luo, & Sakuta, 2010; Hasher, Lustig, & Zacks, 2007; Lindenberger & Ghisletta, 2009; Werkle-Bergner, Freunberger, Sander, Lindenberger, & Klimesch, 2012; West, 1996; Yassa, Mattfeld, Stark, & Stark, 2011). Most theories implicate deficits in particular cognitive processes, such as attentional resources (Craik et al., 2010), association formation (Naveh-Benjamin, 2000), inhibitory abilities (Healey, Hasher, & Campbell, 2013), and processing speed (Salthouse, 1996). Many of these theories provide a qualitative account of a broad range of findings, have served to organize knowledge in the field and have driven much empirical work. However, the theories are generally not evaluated by testing their ability to *simultaneously* account for multiple effects with quantitative precision (Benjamin, 2010).

For example, a deficit in attentional resources would be expected to impair performance on tasks that require effortful memory search (like free recall) and might be expected to spare performance on tasks that require less attentional resources (like recognition). Imagine older adults have d times less attentional resources than do younger adults, call d the coefficient of impairment. It is quite likely that some value of d would impair free recall performance and that some other value of d would spare recognition, but it is not necessarily true that the same value of d would simultaneously both spare recognition and impair free recall by the observed amount. It might be, for example, that the value of d that is sufficiently high to reduce recall by the appropriate amount would also produce a deficit on recognition. That is, it could be the case that no single coefficient of impairment can simultaneously account for both recall and recognition. These ideas are illustrated in Figure 1 which shows that whether or not a single value of d can simultaneously account for both recall and recognition depends on the details of the functions relating d to performance. In Figure 1A, there is a single value of d that fits both the recall and recognition data, but in Figure 1B there is not. Determining which of these two regimes a particular theory falls under is difficult because it requires that the theory make not just qualitative predictions (e.g., age deficits should be larger on free recall than recognition) but quantitatively precise predictions about task performance (i.e., it must specifiy the functions relating d to performance).

Recent models of episodic memory can predict task performance with the level of detail needed to address this issue. These models have been used extensively with younger adults and have allowed theorists to reach a broad consensus on several key aspects of memory such as the importance of similarity (Nosofsky, Little, Donkin, & Fific, 2011) and context (Polyn, Norman, & Kahana, 2009). Researchers have begun to apply these models, quite profitably, to the study of aging (e.g. Benjamin, 2010; Li, Naveh-Benjamin, & Lindenberger, 2005; Ratcliff, Thapar, & McKoon, 2004; Starns & Ratcliff, 2010; Surprenant, Neath, & Brown, 2006), however they have not yet been used to determine which specific cognitive processes account for the complex pattern of impaired and spared performance in memory tasks. Here we use one of these models to simultaneously account for a complex pattern of spared and impaired performance across recognition and recall tasks.

We elaborate our reasons for choosing these two tasks below. Briefly, free recall reveals a pattern of both preserved and impaired aspects of episodic memory performance. Older adults show no deficit in initiating recall and continue to exhibit primacy and recency effects. They do, however, show a deficit in using new temporally–defined (i.e., episodic) associations to guide recalls and recall fewer items overall. Older adults also produce many more intrusions (i.e., recalling words that were not on the current list) than younger adults. No existing theory has not been shown to account for this pattern with a single set of parameters. In contrast to free recall, standard item recognition tasks typically show a modest age deficit in accuracy, that can often be missed in studies with small sample sizes (Verhaeghen, Marcoen, & Goossens, 1993).

Our goal is to develop a theory that can simultaneously account for this pattern of within and across task effects using a single set of parameter values. To develop such a theory, we take a computationally explicit model of healthy episodic memory as a starting point and attempt to build a model of aging by systematically “lesioning” the cognitive processes in this healthy model memory system. We can then compare simulated data from the lesioned model with data from older adults, to determine which memory processes are critical in producing age effects.

We will build on existing work by lesioning the processes implicated by leading aging theories to determine if these processes can account for the pattern of age effects. These initial simulations show that whereas most existing theories provide precise fits to the effects when each effect is considered in isolation, they are unable to account for the effects simultaneously. We develop a theory that can account for the data by fitting the full model (i.e., allowing all parameters to vary) directly the the performance of individual participants and examining the resulting parameter distributions to determine which simulated cognitive processes show significant age differences. The result is a four–component model of aging implicating: 1) the ability to sustain attention across an encoding episode, 2) the ability to retrieve contextual representations for use as retrieval cues, 3) the ability to monitor retrievals and reject intrusions, and 4) the level of noise in retrieval competitions. Finally, we extend the model to simulate a recognition memory task using the same mechanisms the free recall model uses to reject intrusions. Without fitting any additional parameters, the four–component theory that accounts for age differences in free recall also predicts the magnitude of age differences in recognition hit and false alarm rates.

To facilitate other researchers in adopting similar approaches to theory development, all raw data used in the analyses and simulations presented in this paper, along with the model code, are freely available at http://memory.psych.upenn.edu.

Aging and Memory Change: The Data and The Theories

Much of the early work on cognitive aging compared the performance of older and younger adults on the major laboratory paradigms that had been developed in the broader memory literature (Craik, 1977). This work revealed a hierarchy of impairment: recognition tasks showed minimal age deficits in accuracy (Craik, 1971; Schonfield & Robertson, 1966), cued recall tasks showed moderate deficits (Smith, 1977), and free recall showed large and robust deficits (Craik, 1968; Hultsch, 1969; Schonfield & Robertson, 1966). This basic pattern has been well supported over the decades (Verhaeghen et al., 1993).

The gradient of age impairments across recognition, cued recall, and free recall led to one of the most influential early frameworks for thinking about age–related memory impairments: the environmental support framework of Craik and colleagues (Craik, 1983; Craik et al., 2010). The environmental support framework begins with the assumption that the processes underlying encoding and retrieval require cognitive resources. Absent external cues, participants must “self-initiate” processing to generate retrieval cues, or to refine vague cues. Self–initiated processes are assumed to be highly resource demanding. Tasks that provide strong external cues, such as recognition, reduce the load placed on self-initiated processes and thus make the task less resource demanding. If a task requires more self–initiated processing than can be accommodated by a participant’s available resources, performance decrements will be seen, a claim that has been supported in young adults by imposing a secondary task during retrieval (N. D. Anderson, Craik, & Naveh-Benjamin, 1998). If one assumes that cognitive aging reduces the available resources, this framework elegantly accounts for the hierarchy of impairments older adults show on recognition, cued recall, and free recall.

The finding that the sizes of age effects track the specificity of retrieval cues is perhaps the broadest and best supported generalization one can make about aging and memory and is consistent with data from a variety of memory tasks. For example, semantic memory, when measured using vocabulary tests such as the Shipley (1946) test or the WAIS-R, is generally unimpaired (Verhaeghen, 2003). These tests, however, provide very strong and direct retrieval cues. Fluency tasks also measure semantic memory, but do so using a less–focused retrieval cue. For example, in category fluency tests a participant may be given 60 seconds to generate members of the category “animals”. Fluency tasks generally show small but consistent age decrements (Loonstra, Tarlow, & Sellers, 2001; Troyer, Moscovitch, & Winocur, 1997).

Priming shows a similar pattern of cue–dependent age effects. Older adults tend to show intact priming, especially when the possibility for explicit retrieval attempts is minimized (D. B. Mitchell & Bruss, 2003). Notice that priming tasks generally provide an unambiguous cue; a particular recently presented item (e.g., bread) provides a clear cue for the primed item (e.g., butter). If, instead, several primes cue the target, older adults begin to show deficits. For example, Ikier, Yang, and Hasher (2008) had participants count the vowels in a list of words that included orthographically similar pairs (e.g., ALLERGY and ANALOGY). A later task required participants to complete word fragments that resembled both words but could actually be completed by only one (e.g., a _ l _ _ gy). Both older and younger adults showed priming in a control condition in which only the correct solution had appeared in the vowel counting task, allowing it to provide an unambiguous cue for the fragment. When both words are presented, however, there is an ambiguous cue and older adults now show reduced priming relative to younger adults.

One of the main weaknesses of the environmental support framework is that the notion of “resources” is ill defined. Therefore, a major focus of the cognitive aging literature over the past 30 years has been to identify specific cognitive processes that become dysfunctional with age. Much of this work has used novel experimental manipulations and sophisticated statistical modeling to determine the roles of various cognitive processes in the performance of a range of tasks. Candidate processes range from basic perceptual processes (Baltes & Lindenberger, 1997; Lindenberger & Ghisletta, 2009; Pichora-Fuller, Schneider, & Daneman, 1995), to representational fidelity (Benjamin, 2010), to processes subserved by the frontal lobes (e.g. the frontal theory of aging, West, 1996), to the ability to conduct pattern separation in the hippocampus (Stark, Yassa, & Stark, 2010; Yassa et al., 2011). Several of these theories have gained considerable support and have driven a great deal of research.

We want to build on this theoretical work in our own attempt to develop a model of aging. Therefore, we will focus on three of the most prominent aging theories: the Associative Deficit Hypothesis (Naveh-Benjamin, 2000), the Inhibitory Deficit Hypothesis (Hasher & Zacks, 1988), and the Cognitive Slowing Hypothesis (Salthouse, 1996). We choose to focus on these three theories for two key reasons. First, they are well established and widely viewed as foundational. Second, each makes clear claims about the locus of age deficits allowing them to be implemented as lesions in a model of the unaged memory system.

Associative Deficit Hypothesis (ADH)

Naveh-Benjamin (2000) suggested that older adults are selectively impaired in the ability to form (or use) new associations. This hypothesis was based on findings suggesting that aging does not spare all forms of recognition memory. Standard recognition tasks require a simple decision of whether a probe item was among a study set. Researchers working with younger adults introduced recognition tests that required more subtle distinctions, such as whether a given item was presented in a particular location in the display array or whether two items were paired together at study. Compared to younger adults, older adults showed a disproportionally large drop in performance on these associative tasks (Chalfonte & Johnson, 1996).

Naveh-Benjamin (2000) solidified this evidence with a series of experiments designed to decouple memory for items and memory for associations. For example, in one experiment participants had to study word–nonword pairs followed by an associative recognition test that required responding “old” to intact pairs (pairs in which the word and nonword had been paired at study) but not to rearranged pairs (i.e., pairs in which both the word and nonword had been studied but in different pairs). Compared to control conditions in which items but not associations had to be recognized, older adults showed a disproportionate associative deficit.

Support for a selective deficit in associations has come from a variety of paradigms that contrast memory for items with memory for associations. These include contrasting tasks that require memory only for the identity of an item with tasks that require memory both for the item (e.g., a word) and perceptual characteristics of the item (e.g., was this word presented in this particular font) or for face/name pairs (Naveh-Benjamin, Guez, Kilb, & Reedy, 2004). Other work has examined the selective effects of various manipulations on item versus associative memory, such as repetition (Kilb & Naveh-Benjamin, 2011; Overman & Becker, 2009) and the use of strategies (Naveh-Benjamin, Brav, & Levy, 2007). Many different types of associations have been found to be impaired with age. A meta-analysis by Old and Naveh-Benjamin (2008) identified 6 types of associations: memory source (e.g., which of several voices presented an item), associations among features of an item (e.g., the font and identity of a word), temporal order (e.g., which item was presented first), location (e.g., which part of the display the item appeared in), item-item pairings (e.g., intact versus rearranged word pairs), and presentation modality (e.g., visual versus auditory presentation). All but presentation modality were found to show a selective associative deficit.

Inhibitory Deficit Hypothesis (IDH)

Hasher and Zacks (1988) reviewed evidence that across a diverse set of tasks and materials, age deficits tend to be disproportionately large in situations that require overcoming interference. In many cases the interference is proactive interference from earlier in the task, but there is also evidence that older adults have difficulty dealing with distraction that is perceptually present (Gazzaley et al., 2008; Lustig, Hasher, & Tonev, 2006; for a review see Healey, in press). Early formulations of the theory suggested that inhibition was an attentional mechanism that regulated the flow of information into and out of working memory. More recent formulations do not assume a dual–store model of memory (Healey, Campbell, Hasher, & Ossher, 2010; Healey et al., 2013; Healey, Ngo, & Hasher, 2014).

Early work validating the theory focused on finding evidence that reducing the amount of interference on a memory task reduced or eliminated age differences (May, Hasher, & Kane, 1999). More recently, researchers have been pursuing evidence that distraction can actually have a beneficial impact on some tasks, such as creativity tasks in which the distracting information can aid in arriving at novel solutions (for a review see Healey, Campbell, & Hasher, 2008). This work has even been extended to show that under certain circumstances exposure to distraction can benefit older adults on memory tasks (Biss, Ngo, Hasher, Campbell, & Rowe, 2013; Campbell, Hasher, & Thomas, 2010; Campbell, Zimerman, Healey, Lee, & Hasher, 2012).

Research on the IDH has revealed many signs of vulnerability to interference. Complex span tasks (Daneman & Carpenter, 1980; Turner & Engle, 1989) have been used to determine how proactive interference influences age differences in memory span (May et al., 1999; Rowe, Hasher, & Turcotte, 2008). Garden path sentences, in which the initial part of the sentence suggests an ending (e.g., “She ladled the soup into her _____”), have been used to show that older adults are less able than young adults to suppress disconfirmed endings. For example, when given the unexpected ending of “lap”, older, but not younger, adults retain access to the disconfirmed ending (Hamm & Hasher, 1992; May & Hasher, 1998). Fragment completion tasks have been used to show that older adults suffer from interference in priming tasks (Ikier et al., 2008) and fail to suppress competitors during interference resolution (Healey et al., 2013; Healey, Ngo, & Hasher, 2014). Situation modeling tasks (e.g., remembering a series of object-location pairings) have been used to investigate the role of fan size in age differences (Copeland & Radvansky, 2007). In neuroimaging work selective remembering tasks which require participants to attend to only certain sections of a display or certain items in a series have been used to study neural signatures of the suppression of irrelevant stimuli (Gazzaley et al., 2008; Werkle-Bergner et al., 2012).

Cognitive Slowing Hypothesis (CSH)

Salthouse (1996) laid out a case that aging is associated with a general reduction in the speed with which cognitive processing occurs. Unlike the IDH and the ADH, which arose from experiments showing that age interacts with the effect of some manipulation (e.g., that age differences grow disproportionately with the level of interference), the CSH arose from large–scale psychometric studies that correlated age with performance on a wide range of tasks. Such studies generally find that most cognitive abilities, with the exception of those measuring crystallized intelligence (e.g., vocabulary tests), decline with age (Darowski, Helder, Zacks, Hasher, & Hambrick, 2008; Park et al., 2002, 1996). Among the tasks that show decline are those that require carrying out some simple operation as quickly as possible; generally, speeded perceptual tasks that are designed to require little use of memory or reasoning (Hertzog, Dixon, Hultsch, & MacDonald, 2003; Park et al., 1996; Salthouse, 1993). For example, the letter comparison task (Salthouse & Babcock, 1991) presents two strings of letters (e.g., RXL___RXI) on each trial and participants must determine if the two strings are identical or not. The time it takes participants to make these comparisons serves as a measure of processing speed.

Salthouse (1996) proposed two primary mechanisms by which slowed processing could lead to deficits on tasks like recall. The first is the limited time principle: If a task limits the amount of time a participant has to make a response, slowed cognitive processing makes it less likely that the required processing can be completed before the time limit. The second mechanism is the simultaneity principle: Slowed processing can be devastating in a system that depends on information from different subsystems being available at precisely the same time. An assumption here is that many tasks require the products of early processing stages to be available at the same time so that a later processing stage can operate on both simultaneously.

The primary evidence that reductions in processing speed underlie age differences on memory tasks comes from statistically controlling for age–related declines in these speed tasks and then determining whether any age–related differences remain on memory tasks. Across many studies it has been found that controlling for speed reduces the correlation between age and memory. For example, in a meta–analysis Verhaeghen (2011) found that processing speed explained almost 70% of the variance shared between age and episodic memory performance.

This overview of cognitive aging theories is far from comprehensive. Among much else, we have omitted important work on the motivational factors that influence age differences (Carstensen et al., 1999; May, Rahhal, Berry, & Leighton, 2005), task conditions that can mitigate age differences (Castel, 2005), age differences in strategy generation and implementation (Dunlosky & Hertzog, 1998), and age differences in meta-memory (Hertzog & Dunlosky, 2011). Our overview does, however, capture the major thrust of theory development in the cognitive aging literature, in particular the focus on identifying particular cognitive processes that underly age-related memory change.

Our Approach to Theory Building

Theories that assume that one or more cognitive processes become less efficient with age include a free parameter: the precise amount by which the process is less efficient for older adults. This free parameter, however, is generally not made explicit when evaluating a theory’s ability to account for a pattern of findings. To illustrate this point, consider the self-initiated processing resource hypothesis. The claim that reduced processing resources would spare recognition and impair recall implies a simple model under which performance on recall and recognition are both a function of the amount of available processing resources. The amount by which older adults’ available resources are reduced is, essentially, a free parameter in the model. Thus, the theory is really stating that there is some single value of this parameter that allows the model to simultaneously produce the level of both recall and recognition accuracy.

But, as illustrated in Figure 1, it need not be the case that such a parameter value exists. It may be, for example, that any parameter value that produced the appropriate level of recall accuracy, produces too large a deficit on recognition. Determining whether there is indeed a value of this parameter that allows the model to capture both recall and recognition data is impossible unless the theory can make predictions about the precise level of impairment a given reduction in processing resources produces. Our central argument is that to provide a truly compelling account of age-related change, a theory must make its parameters explicit and show that it can account for a broad range of effects with the same set of parameter values. In other words just as the aging memory system produces a particular pattern of behavior with a single set of impairments, a model of aging must account for the pattern with a single set of parameter values. Therefore, to develop a strong theory of aging we need a model that produces quantitatively precise predictions and a set of benchmark findings against which to evaluate it. We first describe the set of benchmarks and then describe our framework for modeling age-related change.

A Set of Benchmark Effects

For a set of benchmarks to be helpful in developing a model of age-related memory change, they should meet several criteria. First, the benchmarks should be composed of a large number of data points including instances of both age-related change and stability. The more data points included in the benchmarks, the more challenging it will be for a theory to simultaneously account for all of them, and the greater our ability to discriminate among theories. Including effects that show both change and stability with age will also help discriminate among theories as it ensures that a theory can not succeed by predicting a global drop in performance, but must instead correctly identify which processes decline and which remain stable (or become enhanced) with age.

Second, our set of benchmarks should reflect the operation of fundamental principles of the memory system so that any theory that fails to capture age differences on those effects would thus fail to provide a comprehensive account of age-related memory change. Yet, to ensure the theories are on a level field, it is important that none of the theories have already been extensively validated against the benchmarks. For example, were we to choose associative recognition tasks, the ADH would clearly have an advantage over the other theories.

The free recall task is an ideal match for these criteria. First, age differences in free recall have been the subject of extensive empirical investigation that has revealed a complex pattern of age effects with both impaired and spared aspects of performance. Second, these effects have been shown to reflect fundamental principles of the memory system that are remarkably consistent across individual younger adults (Healey & Kahana, 2014) and are highly predictive of both memory ability and general intellectual ability as measured by IQ (Healey, Crutchley, & Kahana, 2014). Yet, age differences on these free recall measures have received relatively little attention from aging theorists, so no theory will have the unfair advantage of already being tailored to the task. Finally, the processes that underlie free recall have been extensively modeled and are amongst the best understood of any memory task (Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Farrell, 2012; Polyn et al., 2009), which will allow us to derive precise predictions about the magnitude of age-related change.

In addition to free recall, we will examine age differences on a recognition memory task. Being able to capture the substantial age-related change on free recall while simultaneously capturing the more subtle changes on recognition is a strong test of an aging theory. Our strategy is to use free recall data to test a initial set of candidate theories and to use recognition data as an across-task test of any theory that passes that first hurdle.

Aging and Free Recall

One of the earliest and clearest findings in cognitive aging research was that older adults are impaired on the free recall task (Ceci & Tabor, 1981; Craik, 1968; Hultsch, 1969; Laurence, 1967; Schonfield & Robertson, 1966). Examination of the dynamics of free recall reveals a detailed picture of episodic retrieval and a complex pattern of age effects. Serial position curves (SPC) are shown in Figure 2A. The older and younger adult SPC are qualitatively similar, with both groups showing primacy and recency effects (Kahana et al., 2002). Older adults generally show lower recall levels across the SPC (Capitani, Della Sala, Logie, & Spinnler, 1992; Foos, Sabol, Corral, & Mobley, 1987; Kahana et al., 2002; Parkinson, Lindholm, & Inman, 1982; Poitrenaud, Moy, Girousse, Wolmark, & Piette, 1989; Rissenberg & Glanzer, 1987; Ward & Maylor, 2005), though occasionally reduced age differences are observed for recency items (Castel, Benjamin, Craik, & Watkins, 2002; Craik, 1968; Craik & Jennings, 1992; Raymond, 1971, we return to this point below).

More subtle age differences are revealed by decomposing the retrieval sequence into measures of how recall is initiated and measures of how transitions are made among items after initiation (Kahana et al., 2002). Probability of first recall (PFR) curves (Hogan, 1975; Howard & Kahana, 1999; D. Laming, 1999) show the probability of initiating recall at each serial position. Older and younger adults' PFR curves are virtually identical (Figure 2B), with both groups tending to initiate recall with the final item. After initiation, recalls are driven by associations between the just-recalled word and other words in the lexicon.

Temporal associations exert a powerful influence on transitions for younger adults (Kahana, 1996). Given that the item from position i has just been recalled, conditional response probability (lag-CRP) indicates the probability that item i + lag will be recalled next. Lag-CRPs show the *contiguity effect*, a tendency to successively recall neighboring list items, and the *asymmetry effect*, a tendency to recall items in forward order, regardless of variables such as presentation modality or encoding task. The idea of temporal contiguity captures the core of episodic memory—the ability to reconstruct the sequence of past events and thereby place them on an autobiographical timeline. There is also evidence that the ability to reproduce temporal order is a key element, not just in memory tasks, but complex cognition more broadly, as the extent to which young adults show temporal contiguity is correlated with general cognitive ability, as measured by WAIS IQ (Healey, Crutchley, & Kahana, 2014).

Older adults show reduced contiguity (Figure 2C), indicating that they are less influenced by newly-formed temporal associations than are younger adults (Kahana et al., 2002). Note that age differences are especially pronounced for lags of +1 and −1, with older adults being less likely to make these close temporal transitions. This reduced ability to use temporal information could make it difficult for older adults to situate events in time thus impairing autobiographical recall. If so, a temporal contiguity deficit is likely a large part of what makes memory difficulties so troubling for older adults. As we will see, it is this temporal contiguity deficit that poses the greatest challenge to aging theories.

Finally older adults show increased intrusion errors. An intrusion is the recall of a item that was not on the target list. Prior-list intrusions (PLIs) come from a previous study list in the session whereas extra-list intrusions (ELIs) were not studied on an earlier list. Older adults show increased rates of both types (Figure 2D). When PLIs occur, they tend to be words from recent lists—a PLI recency effect. Figure 2E shows that both older and younger adults show this effect, mirroring the finding that older adults have preserved recency as measured by PFR curves.

A Model-Based Approach

As we outlined above, our approach is to test and refine theories by evaluating their ability to account for the set of benchmark findings shown in Figure 2. The cognitive underpinnings of these effects have been extensively modeled in younger adults using many of the cognitive processes implicated by aging theories. Therefore, any comprehensive theory of aging should be able to account for age differences on these effects in quantitative detail.

A key feature of this set of benchmarks is that it is composed of multiple variables that show both impairments and stability across age groups. On one hand, the multivariate nature of the data is necessary to distinguish among theories, on the other hand, multivariate data makes it difficult to evaluate the theories' ability to account for the pattern. When considering univariate measures of performance such as overall accuracy (e.g., percent recall) it is relatively easy to check a theory's predictions against the data with simple thought experiments. For example, one can easily intuit that failing to inhibit words from previous lists will make it hard to selectively retrieve items from the current list, lowering overall recall. However, the multivariate nature of the measures in Figure 2 makes such thought experiments more difficult—it is difficult to intuit if or how the effect of reduced inhibition would vary with serial position. Thought experiments become intractable when trying to consider several multivariate effects at once (e.g., how can reduced inhibition simultaneously lower non-recency portions of the SPC while leaving recall initiation intact but making near temporal transitions less likely). The difficulty is magnified if we are interested in testing the ability of a theory to capture the quantitative level of the effects (e.g., not just predicting that older adults' SPCs are lower than younger adults', but predicting precisely how much lower they are).

A similar problem faced the researchers working on episodic memory in younger adults. Early theoretical and empirical work clearly established the importance of cues in determining which information is accessible for retrieval and pointed toward the importance of context in this cuing process (Bower, 1967; Estes, 1955; McGeoch, 1932; Tulving, 1972; Underwood, 1945). However, without detailed models of how items and context interact during study and recall, it was difficult to convincingly account for the complex dynamics of memory search. To a large extent this problem was solved by developing formal models of benchmark effects in a small number of laboratory paradigms. This modeling work has led to a broad consensus about many fundamental aspects of memory including the importance of context–based cues in retrieval (e.g., Bower, 1967; Davelaar et al., 2005; Estes, 1955; Farrell, 2012; Howard & Kahana, 2002a; Lohnas, Polyn, & Kahana, 2015; Mensink & Raaijmakers, 1988; Murdock, 1997; Polyn et al., 2009; Sederberg, Howard, & Kahana, 2008). We can leverage this existing computational framework to help make the complexity of the free recall aging pattern tractable.

Moving to a formal modeling framework aids us in another way. Theorizing in the cognitive aging literature generally follows a disease model in which it is assumed that the cognitive systems of older adults can be characterized as a version of the younger adult cognitive system in which one or more mechanisms are dysfunctional (for a somewhat different approach see Carstensen et al., 1999; Zimmerman, Hasher, & Goldstein, 2011). This view of aging is particularly well suited to testing within a computational model of younger adults' episodic memory. We can begin with a model that accurately simulates the performance of younger adults. Then, for a given aging theory we can identify which memory mechanisms are implicated by the theory and "lesion" the corresponding model mechanisms, and use the lesioned model to generate simulated data: If the theory is accurate the simulated data should resemble that of older adults.

To use this lesioning approach, we must select a particular computational model of episodic memory. The task of comparing the explanatory power of different aging theories will be much easier if the theories are embedded in a common model of the memory system. The IDH, ADH, and CSH do not provide strong guidance in this regard, as they have not made commitments about the nature of the underlying memory system at the computational level. Existing modeling work in the aging literature has generally either developed novel models designed specifically to account for age effects (Benjamin, 2010; Li et al., 2005), or used existing models (e.g. Ratcliff, Thapar, & McKoon, 2004; Surprenant et al., 2006) to develop unique theories of aging rather than to test existing theories. Therefore, we take the approach of implementing each of the theories within the retrieved context framework. The retrieved context framework has been extensively used to model the free recall effects outlined above in young adult samples (Howard & Kahana, 2002a; Lohnas et al., 2015; Polyn et al., 2009; Sederberg et al., 2008). The retrieved context framework provides a good test bed for at least two reasons. First, most contemporary models of episodic memory share the core features of the retrieved context framework (i.e., that associations between items and a drifting context representation are critical in accounting for recall dynamics, e.g., Davelaar et al., 2005; Farrell, 2012). Second, the retrieved context framework accounts for many findings that cannot be accounted for by competing models (e.g., Lohnas & Kahana, 2014a, 2014b).

The Retrieved Context Framework

The notion that memory for an item is intimately tied to the context in which it occurred has a long history in the science of memory (Bower, 1967; Estes, 1955; Howard & Kahana, 1999, 2002a; McGeoch, 1932; Mensink & Raaijmakers, 1988; Murdock, 1997; Tulving, 1972; Underwood, 1945). Yet, exactly what is meant by "context" can be difficult to define. Therefore, we start with an example to provide an intuition of the retrieved context framework (Appendix A provides a full formal description).

Imagine I have just presented you with a picture of your childhood home. Perceptual processes will create a very clear representation of the physical features of the picture. This representation will in turn trigger a cascade of episodic and semantic associations. You may recall details of the neighborhood; you may recall specific episodes that took place at the house; you may remember friends you had while you lived there, and all of these semantic and episodic memories will be colored by various emotions. This rich ensemble of activated representations is what we call context—a milieu of brain activity that is not identical with the representation of the picture itself.

The power of the retrieved context framework lies in the dynamics of how item and context representations interact as events occur in the environment. To continue with our example, imagine I present a new picture: One of your Psych 101 professor. Just as the "childhood house" picture did, this new picture will trigger a cascade of associations; you may remember where you sat in class; the cover art of the course textbook; your favorite study place. But this new cascade does not completely erase the activated representations that formed the "childhood home" context, rather it builds upon it; resulting in a new ensemble of activated representations that is neither purely "childhood home" nor purely "Psych 101", but a blend of the two. In essence, this blending of contexts means that thoughts related to the previously presented picture remain in mind as the next picture is presented. The fact that the professor item representation is active alongside the "childhood house" context representation allows new episodic associations to form between the two representations. That is, aspects of the professor item representation will become linked to the house context, and, reciprocally, aspects of the house context will become associated with the professor item representation. These reciprocal associations between item's and contexts allow for the formation of new episodic memories (Gallistel, 2008; Howard & Kahana, 1999, 2002a).

If I continued showing you pictures, they would continue to initiate cascades of associations, with each new picture adding to the ensemble of activated representations. Of course, these representations do not stay active indefinitely; rather mental context continually drifts as new events activate new representations, with representations associated with the most recent events being strongly active and those associated with more distant events fading into the background. Context is thus a recency–weighted history of past events.

Moreover, you would continue associating each newly presented item with the state of context that prevailed when it was presented (and forming reciprocal associations between contexts and items). The strength of these newly formed associations is unlikely to be uniform across the list, however. Instead, associations formed early in the list are likely to be stronger than those formed later in the list. That is, there is a primacy gradient. This primacy gradient reflects stronger encoding of early–list items, due either to more frequent rehearsals (D. Laming, 2008; D. L. Laming, 2006; Marshall & Werder, 1972; Modigliani & Hedges, 1987; Rundus, 1980; Tan & Ward, 2000; Ward, 2002) or to more efficient encoding during the item's initial presentation (Atkinson & Shiffrin, 1968; Neath & Crowder, 1990; Sederberg et al., 2006).

Now imagine that I ask you to recall all of the pictures you have seen (i.e., to do the free recall task). How do you begin recalling the items? The retrieved context framework proposes that you do so by using the current state of the context representation as a retrieval cue. Such cuing happens by allowing the context representation to activate those item representations to which it is associated. The last picture you saw is likely to be strongly activated because its contextual state is still strongly represented in the contextual ensemble. Therefore, consistent with the PFR data (Figure 2B), you are likely to recall the last item first.

What happens when you successfully recall one item? Recalling an item activates its item representation which in turn, like studying an item, activates its associated contextual state. But there is a key difference between studying an item for the first time and recalling an item (or studying it for a second time; Lohnas & Kahana, 2014b): When an item is studied for the first time it activates contextual states to which it was associated before the experiment, the *pre–experimental* context. When an item is recalled, it activates this same pre–experimental context, but because new learning occurred during study it also reinstates the context that was active when the item was studied. This reinstated *experimental* context is a blend of the pre–experimental context states of all items studied prior to the just-recalled item. After successfully recalling an item, its pre–experimental and the experimental contextual states are incorporated with the existing state of context, and this updated context is used to cue the next recall.

Which items are likely to be cued by the updated context? The experimental component of the reinstated context was active when the just-recalled item was studied, but because context drifts slowly, somewhat similar states of context were active while items presented before and after the just-recalled item were studied. The degree of similarity decreases as the lag from the just-recalled item increases. Therefore the experimental context provides a strong cue for items studied near the just-recalled item and a weak cue for items studied far form the just recalled item, giving rise to the contiguity effect (Figure 2C). During study, the pre–experimental component of the reinstated context was not active until the just-recalled item was presented. Therefore, the pre–experimental component formed associations only with those items that came later in the list. Thus, the pre–experimental component is a good cue for items that were presented after the just–recalled item, but is a very poor cue for items presented before the just–recalled item. The fact that the experimental component of the reinstated context is a good cue for both earlier and later list items, but the pre–experimental component provides a cue only to later list items means that forward transitions are more likely than backward transitions, giving rise to the forward asymmetry in the lag-CRP context reinstated by retrieved items provides a cue to both earlier and later list.

This example captures the spirit of the retrieved context framework and the core mechanisms that produce the free recall patterns shown in Figure 2. The retrieved context framework has been implemented in several neural network models that represent items and contextual states as nodes and represent associations as the connections between nodes (Howard & Kahana, 2002a; Lohnas et al., 2015; Polyn et al., 2009; Sederberg et al., 2008). For the simulations in this paper we use the continuous memory version of the context maintenance and retrieval model (CMR2; Lohnas et al., 2015), an implementation that can simulate an entire session of free recall lists within a common associative network. By contrast, most previous free recall models have reset their associative networks after each simulated list, a simplification that prevents them from modeling PLIs in a natural way. Here we provide an overview of the model with a focus on key mechanisms that will be used to implement the aging theories. Appendix A provides a full formal description of CMR2 (see also Lohnas et al., 2015).

For illustration we will use a scaled–down version of the model that simulates a single free recall trial with the 4–item list *apple, cat, boat, dog*. Figure 3 shows the basic structure of the model: items are represented as nodes on an feature layer (which can be thought of as the perceptual features of the items) and context is represented as nodes on a context layer. Each word has a corresponding node on both the feature layer and the context layer. A word's feature node stands for the low–level representation of the item triggered by its presentation. In the version of the model used here, each item has a single feature node, but items could be represented by multiple feature nodes. The corresponding context node represents the ensemble of activated semantic and episodic associates we defined as context. Each feature node has a directional connection to each context node. Each of these connections has a value that encodes the strength of the association between the feature and a context. In the model these feature-to-context associations are stored in associative matrices. Each feature has a strong connection to its corresponding context node and very weak connections to the remaining context nodes. Just as each feature has a directional connection to each context, each context node has a directional connection to each feature node. These context-to-feature associations are stored in a separate associative matrix.

Figure 4 shows how the network responds as items are presented during the encoding phase. In the first panel *apple* has just been presented, which causes its node to become active on the feature layer (represented by the shading of the node). Activation flows through the network via the feature-to-context associations. Because the *apple* feature node is strongly associated with the *apple* context node, but weakly associated with all other context nodes, only the *apple* context is activated.

The second panel shows what happens when the next item, *cat*, is presented; first notice that the *cat* feature node has been activated and the *apple* feature node is now completely deactivated, reflecting the fact that activation on the feature layer lasts only while the stimuli is perceptually present. In contrast, both the *cat* and *apple* nodes are active on the context layer: activating *cat* on the feature layer causes its context node to activate via the feature–to–context associations, and this new activation is added to the existing activation of the *apple* node. Allowing multiple context nodes to be active at the same time is the critical feature that allows the context layer to simulate the blending of the contextual states of recently presented items. But the two nodes are not equally active; instead, the context node associated with the most recently presented item is strongly activated and the activity of the *apple* context node has decayed somewhat (represented by the size of the icons in the context layer). This strong activation of recently presented items and progressively decaying activation of more distant items allows the layer to serve as a recency–weighted history of past events. The amount by which activated nodes decay with each new item presentation is governed by a model parameter, β_{enc}. When β_{enc} is set to one, each new item completely erases existing activation and when it is set to zero, new items have no effect on the context layer. Thus, β_{enc} controls the degree of recency weighting in the context layer.

The second panel also illustrates how episodic memories are formed in the model. The model uses a Hebbian learning rule so that associations form between feature nodes and context nodes that are active at the same time. Because the activation of the *apple* context node persists when *cat* is presented, the *cat* feature node and the *apple* context node become associated via Hebbian learning. This new association is represented by the shading of the corresponding cell in the associative matrix in the second panel of Figure 4.

The efficiency with which new associations are formed is governed by several factors. First, two parameters, γ_{FC} and γ_{CF} control the strength of new feature–to–context and context–to–feature associations, respectively. When the value of γ_{FC} is low, weak feature–to–context associations are formed; when it is high, strong feature–to–context associations are formed (γ_{CF} similarly influences the strength of context–to–feature associations). The second factor governing association formation is attention. Consistent with the notion that primacy results from changes in the efficiency of encoding across the list (Serruya, Sederberg, & Kahana, 2014; Tulving & Rosenbaum, 2006), there is a primacy gradient such that early items are encoded more strongly than later items. Two model parameters govern the primacy gradient. ϕ_s controls the size of the initial primacy boost, and ϕ_d controls the rate at which it fades across the list. As described in Appendix A, the values of ϕ_d, ϕ_s, and γ_{FC} scale the strength of new associations.

The third and fourth panels of Figure 4 show the evolution of both the context layer and the associative matrix as the remaining list items are presented. As each successive item is presented, its context node becomes strongly activated and the activation of the context nodes of more distantly presented items decays. When each new item is presented, its feature node becomes associated with all active context nodes (and vise versa), with the strength of the associations being determined both by how active the context node is (features become strongly associated with highly active nodes and weakly associated with weakly activated nodes) and the value of the primacy parameters and γ_{FC}. Notice that feature nodes are only associated with the context nodes of items that preceded them in the list; *apple* is not associated with any new context, whereas *dog* has been associated with all of the context nodes. This fact helps explain the asymmetry of the contiguity effect.

Figure 5 illustrates the retrieval process. Once all items have been presented, recall begins by using the current contextual state to cue retrieval via context-to-feature associations. Each context node sends activation to each feature node via the context–to–feature associative matrix. The specific amount of activation sent from a given context node to a given feature node is determined by the activation level of the context node multiplied by the strength of the context–to–feature association between the nodes. The total amount of activation received by a given feature node is simply the sum of all activations arriving from all context nodes.

These activation levels on the feature layer provide the initial input to a retrieval competition in which all items race to cross a threshold, following the dynamics of the competitive accumulator model of Usher and McClelland (2001). The evidence accumulation process includes noise, so that although the item with the strongest activation on the feature layer has the best chance of winning, items with lower activation can sometimes win. As we will see, the amount of noise in the process, which is governed by a model parameter, turns out to be an important component in capturing age differences.

Before the winning item is actually output by the model, it undergoes a post–retrieval editing phase to screen for possible intrusions. Items that were actually studied on the list should be associated with states of context that are very similar to the state of context that prevailed during the recall period. By contrast, intrusions will tend to be associated with contexts less similar to the test context. Therefore, CMR2 screens for intrusions by allowing each potential retrieval to reinstate its context, and comparing that reactivated context to the existing context. If the similarity of the two exceeds a threshold, the item is recalled.

Once an item is recalled it activates its associated context state via the feature-to-context associations (see the second panel of Figure 5). This activated context includes both the item's own context node (to which it was associated before the experiment) and any context nodes to which it formed new associations during the study period, with the degree of activation being governed by the strength of the feature–to–context associations. Just as during encoding, the newly activated context state does not completely erase the existing state of context, but blends with it. A parameter, β_{rec}, controls how rapidly old states of context decay as new states are activated during retrieval (i.e., the rate of context drift). The rate of context drifts has a powerful influence on the size of the contiguity effect. The updated context representation is used to drive another recall competition. Note that once an item is recalled it is still allowed to compete in subsequent retrieval competitions (i.e., like participants, the model can recall the same item more than once), but is handicapped by having to pass a higher threshold than never–recalled items (see Appendix A for full details). The cycle of context cuing, retrieval competition, editing, and context reinstatement continues until the end of the recall period (e.g., 30 seconds in the first study considered here) is reached, at which point the next trial begins. The number of cycles that can take place during the retrieval period is determined by how long each retrieval competition takes: The accumulator proceeds in a series of iterations and a model parameter governs how many seconds each iteration takes.

It is thought that the transition from one trial to the next is accompanied by a change in mental context (Sahakyan & Kelley, 2002). CMR2 simulates this context shift by activating a unique "disruption" item on the feature layer and allowing this item to update context. One effect of this interlist context shift, which will become important later, is to reduce the accessibility of items presented before the shift, as the activation of contexts associated with prior–list items is significantly reduced. That is, the shift serves to contextually isolate successive lists.

Under CMR2, the free recall effects shown in Figure 2 derive from the dynamics of context drift, in particular the ability of presented and recalled items to reinstate contexts with which they have been associated. Recency occurs because the context at the end of the list, which is used as a retrieval cue, most closely matches items presented near the end of the list. Primacy occurs because the effectiveness of encoding processes changes across a trial. At the start of a trial they are highly efficient but become less effective with each successive word, following an exponential decay to a baseline level. As a consequence, early list items form stronger associations with the context that prevails when they are presented than do items presented later in the list. Temporal contiguity occurs because each recalled item retrieves a context that is similar to the contexts associated with its near neighbors in the list and less similar to its more distant neighbors and therefore the retrieved context most strongly cues neighbors of the just-recalled item. Intrusions occur because the context retrieved by a current list item is similar to a context associated with previous list items (existing semantic associations produce ELIs and new episodic associations produce PLIs).

We focus first on simulating age differences in free recall before examining a recognition task. The retrieved context framework has not been applied to recognition so we develop an extended version of CMR2 to simulate the task. Recognition requires determining whether a probe item (e.g., a word) was recently studied (i.e., old probes) or not (i.e., new probes). The difficulty of the task is that both items have representations in memory and are distinguished only by whether they were studied in the experimental setting. This challenge is very similar to the challenge of distinguishing intrusions from genuine list items during free recall (K. J. Mitchell & Johnson, 2009). Therefore, we simulate recognition using the same mechanism used to screen for intrusions during free recall: each probe is allowed to activate its associated contextual representation which is then compared with the current state of context. Old probes became associated with the context that prevailed when they were studied, and because context drifts slowly, this study-phase context will be similar to the state of context during the recognition test. Therefore, old probes will be expected to reinstate a contextual representation that is similar to the current state of context. By contrast, new probes formed no association with the study-phase context because they were not studied and therefore are expected to reinstate a context representation that is somewhat dissimilar the the current state of context. Just as with intrusions, probes are endorsed as having been studied if the similarity between the reinstated context and the current context exceeds a threshold. It would be possible to develop a more complex model of recognition, but for our initial simulations we wanted to determine if this simple thresholded context comparison mechanism, which is a direct extension of the mechanisms used to simulate free recall, is sufficient to account for age differences (see below for a discussion of alternative models and the relationship of this mechanism to dual process recollection/familiarity models; Yonelinas, 2002).

For simplicity, our description of the model omitted several details. Although a full description of the model is available in Appendix A, we highlight a few important details here. In our four–item example, feature-to-context associations and context-to-feature associations had the same strengths; however, this need not be the case in the full model. In our example we assumed that associative matrices begin the simulated trial with each item being associated to its context node, but no other context nodes (and vise versa for context–to–feature associations). In the actual model the matrices encode pre–experimental semantic associations among items (e.g., the *dog* feature node is most strongly associated with the *dog* context node, but also has a weaker association with the *cat* node, reflecting the fact that the two words are semantically related). The strengths of these semantic associations were determined by Latent Semantic Analysis (LSA; Landauer & Dumais, 1997). LSA allows one to measure the semantic relationship between two words as the cosine of the angle between the words' representations in a multidimensional model of semantic space. The raw LSA values are then scaled by two model parameters (one for feature–to–context associations and one for context–to–feature associations) that control the strength of pre–experimental semantic associations. Finally, in the simulations presented below the actual layers and associative matrices are much larger than those in the 4–item example; they include nodes for each item presented (30 lists × 10–items per list = 300 items for the first study considered here) plus nodes for an equal number of items that were not presented. These un–presented items simulate the fact that participants must distinguish list items from a much larger lexicon of words that includes close semantic associates of the list items, which often occur as intrusions (Zaromb et al., 2006).

Simulating Age–Related Memory Change

The ADH, IDH, and CSH provide clear starting points in our attempt to develop a model that can simultaneously account for the full pattern of age differences on free recall and recognition. Each of the theories implicate certain cognitive processes as the locus of age–related memory impairments. Our goal is to determine if deficits in these processes can account for the effects in Figure 2. To do so we will identify mechanisms in the CMR2 model that correspond to the cognitive processes implicated by the theories. We will then "lesion" those model mechanisms and determine if the lesioned model produces simulated data that duplicates the pattern of spared and impaired performance shown by older adults.

Our approach assumes that the memory mechanisms that comprise CMR2 *can* be impaired in a way that captures the age effects. To ensure that the CMR2 framework can simulate the age effects, we began by fitting the model independently to both the older and younger adult data, allowing all parameters to vary. Each parameter corresponds to a process simulated by the model. The parameters can be thought of as dials that control their corresponding process. For example, the parameter β_{enc} controls context drift at encoding. If β_{enc} is set to 1 (i.e., the dial is turned all the way up) context drifts very quickly, with each new item replacing the previous state of context with its own associated context state. If β_{enc} is set to 0, context does not drift at all and newly presented items do not change the state of context. Fitting the model involves tuning all of the parameters to find the set of values under which the model's simulated behavior most closely matches participants' actual behavior. As described in more detail below, we use a genetic algorithm to achieve this tuning. As illustrated in Figure 6, the model provided accurate fits to both the older and younger adult data.

We can now use the parameter values that produced the simulated *younger* adult data as our model of healthy memory. We will determine if this healthy model can simulate older adult data after lesioning the mechanisms implicated by the aging theories. Next, we describe how we implement each theory within the retrieved context framework.

Implementing a theory that was originally specified at a verbal level within a formal model requires making many choices. For example, does an associative deficit occur due to a failure to encode new associations or due to difficulty retrieving them (or both)? Some of these issues have been discussed in the aging literature (Naveh-Benjamin, 2000; Wingfield, Lindfield, & Kahana, 1998), but generally not at the level of detail required to implement a formal model. The result is that the modeler has a great degree of flexibility in how to implement the theories, and it is impractical to explore all possible implementations. Rather than pick a single arbitrary implementation, we have chosen to implement each of the theories in three different ways that are consistent with existing descriptions of the theories.

Associative Deficit Hypothesis

The associative deficit hypothesis (ADH) attributes older adults' memory impairment to a reduced ability to form associations (Naveh-Benjamin, 2000). In CMR2, the formation of new episodic associations is governed by two parameters, γ_{FC} and γ_{CF}, that respectively control the extent to which the strength of feature–to–context and context–to–feature associations change with learning. If these parameters are set to zero, no new associations are formed and recall is dominated by pre-existing semantic associations. As the values of these parameters are increased above zero, the influence of new associations learned during list presentation increases. Therefore, a natural way to implement an associative deficit is to reduce the value of γ_{FC} or γ_{CF}. Impairing the ability of the model to form feature–context associations would lead to smaller changes in the association between list items and the state of context that prevailed when they were presented. During retrieval these weak associations would prevent the model from using each recalled item to retrieve a new context that serves as a good cue for neighboring list items. In contrast, impaired context–feature associations would prevent the model from using current contextual states as retrieval cues. The fact that there are separate parameters controlling feature-to-context and context-to-feature associations raises the question of whether only one or both types of associations are affected by aging (and if both, whether they are equally affected). To give this implementation of the ADH the greatest chance of simulating the age effects, we allow both γ_{FC} and γ_{CF} to vary independently. Concretely, we implemented this version of the ADH by taking the best–fitting *younger* adult parameters (i.e., our model of healthy memory in Figure 6) and attempted to fit the older adult data while holding all parameters at the younger adult values except γ_{FC} and γ_{CF}, which were allowed to vary.

An associative deficit could arise if newly formed associations are *weak* relative to pre–existing associations. But it could also arise from *noisy* association formation. In the standard version of CMR2, association formation is noise–free in that when an item is presented its feature node becomes associated with only those context nodes that are currently active. That is, the associations encode a noise–free representation of which item and context nodes were co-active and the degree to which they were co-active. We can introduce noise into the association formation process by adding a random vector to the contextual states that retrieved items reinstate during retrieval so that items are no longer retrieving a noise free representation of the contexts to which they were associated. Specifically, $\mathbf{c}^{\mathbf{IN}}_i = M^{FC}\mathbf{f}_i + \zeta\mathbf{n}_r$, where $\mathbf{n}_r$ is a normal random vector and ζ is a parameter controlling the weighting of the noise vector. To test this version of the ADH we allowed ζ, which controls the amount of associative noise, to vary.

Of course, weak and noisy associations are not mutually exclusive. Therefore we included a third implementation of the ADH in which we allow both the parameters controlling association strength (γ_{FC} and γ_{CF}) and the parameter controlling associative noise (ζ) to vary.

Inhibitory Deficit Hypothesis

Under the inhibitory deficit hypothesis (IDH) older adults have difficulty preventing irrelevant information from interfering with retrieval of relevant information (Hasher & Zacks, 1988). Hasher, Zacks, and May (1999) argued that inhibition is not a unitary construct (see also Kramer, Humphrey, Larish, Logan, & Strayer, 1994; Nigg, 2000). Rather, they proposed three inhibitory processes: access, deletion, and restraint. The access function serves to prevent irrelevant information from accessing the focus of attention. Irrelevant information can either be irrelevant stimuli or irrelevant memories, therefore in a recall task the access function would limit retrieval to task relevant memories. The deletion function serves to suppress (i.e., delete) information from the focus of attention when it is no longer relevant to the task (e.g., prior lists in free recall) [1]. Together, access and deletion control the relative activation of relevant versus irrelevant memories; their efficient functioning ensures that retrieval is limited to task relevant memories (e.g., items from the current list in free recall). The restraint function allows us to withhold, or restrain response tendencies that are strong but inappropriate given the task at hand. In the case of memory search, restraint may allow participants to reject intrusions when they come to mind (e.g., if a semantic associate of a genuine list item wins a retrieval competition, restrain may allow the participant to withhold the response and avoid making an intrusion).

Within CMR2, the access function can be modeled as a mechanism that modulates the rate at which mental context drifts during encoding and retrieval. When a item is presented for encoding it triggers a cascade of associated thoughts and memories, many of which are not relevant to the task at hand. Younger adults may use the access function of inhibition to prune this cascade (Healey et al., 2013; Healey, Ngo, & Hasher, 2014). For older adults, an inhibitory deficit may allow these irrelevant thoughts and memories to remain active and become bound to relevant information in memory (e.g., Hamm & Hasher, 1992). As a result, older adults' memories may be densely interconnected, setting the stage for massive interference at retrieval (Radvansky, Zacks, & Hasher, 2005). Moreover, if during retrieval each recalled item triggers retrieval of many associated, but irrelevant, memories it may hamper the ability to target list items. Such deficits can be simulated in CMR2 by varying how rapidly mental context drifts. For both younger and older adults, when an item is presented it will activate its pre-experimental context—its semantic associates. Older adults may integrate *too much* of this newly retrieved context into their context representations (but see Howard, Kahana, & Wingfield, 2006). Such rapid drift can be simulated with the β_{enc} parameter, which controls the extent to which newly retrieved context replaces existing context in the context layer. Similarly, rapid context drift during retrieval can be simulated with a higher value of β_{rec}. Therefore, we implemented this version of the IDH by allowing both β_{enc} and β_{rec} to vary freely.

[1]The three inhibitory functions were originally framed within a dual-store model of memory (Hasher et al., 1999) with access and deletion being described as processes that regulate the contexts of working memory. Later work has not assumed a dual-store model (Healey et al., 2013; Healey, Ngo, & Hasher, 2014). Under a single-store framework the functions map naturally to processes that regulate the relative accessibility of relevant and irrelevant memories.

The deletion function can be modeled by using contextual drift to make irrelevant information inaccessible, effectively "deleting" it by isolating it. In the case of free recall such contextual isolation could be accomplished by allowing the internal context representation to drift by a large amount during the period between successive lists. Such drift will decrease the similarity between the context that prevailed during the previous list and the context of the current list. As a consequence, items in successive lists will be associated to largely uncorrelated contextual states. Thus, during retrieval prior list items will be relatively less accessible because context cues that provide a good match to the current list will be a poor match to previous lists. Perhaps older adults do not allow context to drift enough to form such contextual boundaries. To test this possibility we allowed the β_{post}^{recall} parameter, which governs the rate of context drift between lists, to vary.

Finally the restraint function can be modeled as difficultly rejecting intrusions. Such a deficit maps directly onto the post-retrieval decision mechanism in CMR2 and can be implemented by lowering the threshold, c_{thresh}, of similarity between an item's retrieved context and the current state of context needed for an item to be recalled. We tested this implementation by allowing c_{thresh} to vary.

The Cognitive Slowing Hypothesis

The cognitive slowing hypothesis (Salthouse, 1996) suggests that older adults suffer a general slowing of cognitive processes. During encoding, slowed processing would result in weaker associations being formed per unit of processing time. For example, older adults may not have time to finish processing the current word before the next appears. Therefore, we can simulate slowed encoding processes by simply down-weighting the influence of newly learned episodic associations in the same way we did for the ADH.

Slowed retrieval can also be readily simulated in CMR2. The recall period is limited in time (30 seconds in the first study considered here), and any slowing of retrieval processes will reduce the maximum number of items that can be recalled in that time. Each retrieval attempt takes a finite amount of time, and that time is governed by a model parameter. By varying that parameter we can simulate slowed retrieval—the consequence of which is to reduce the number of retrieval attempts that can be made within the recall period time limit.

As a final implementation of the CSH, we can assume that older adults suffer from both slowed encoding and slowed retrieval processes. To do so we simply allow the parameters governing both association formation and the speed of retrieval to vary.

Simulating Effects in Isolation

Ultimately, we want to arrive at a set of lesioned mechanisms that can simulate all of the age effects simultaneously using a single set of parameter values. Simulating the effects simultaneously, however, involves complex trade-offs, which make it difficult to understand why a particular mechanism fails or succeeds in capturing the age effects. We therefore start by considering the ability of each lesioned mechanism to capture the effects in isolation. We did this by taking the SPC, PFR, lag-CRP, intrusion rates, and the PLI-Recency data one at a time and determining if each of the theory implementations was able to fit that aspect of the data while ignoring the other aspects. In our first set of simulations we used the Kahana et al. (2002) shown in Figure 2.

As described above, each theory implementation consists of a set of parameters that may differ from the optimal younger adult values, thereby simulating a deficit in the mechanisms governed by those parameters. To determine if a theory can describe older adults' behavior for a particular effect we must systematically vary the implicated parameters, and for each variation run the model to produce simulated data. We can measure how well each of these simulations fits the older adult data by computing the root-mean-square deviation (RMSD) between the model simulations and the older adult data. The goal is then to find the parameter set that minimizes the RMSD. If we want to claim that a given theory *cannot* account for an effect, it is critical that we have first thoroughly explored the parameter space to ensure that there is no set of parameter values that would allow the model to simulate the effect.

Thus, to search for the best parameter set for each theory we began by simulating 100 parameter sets evenly distributed across the parameter space. We then conducted a focused grid search centered on the best–fitting of those initial 100 parameter sets. Specifically, we created a multi–dimensional grid by selecting 20 evenly spaced values for each parameter and evaluating the parameter sets defined by the intersections of the grid, for a total of 20^p parameter sets where p is the number of parameters in the theory implementation. Take, for example, CSH implemented as slowed encoding and retrieval processes, which implicates 3 parameters (the two association formation parameters and the parameter governing how long a retrieval attempts takes). For this theory, the grid search would encompass $20^3 = 8000$ parameter sets. The best–fitting parameter values for all simulations are reported in Appendix B.

Serial Position Curve

Figure 7 shows the simulated serial position curves produced by the best–fitting version of each model. Rows one to three show the ADH, IDH, and CSH respectively. Each column corresponds to a different implementation. Each panel shows both the younger and older adult data along with simulated data from the parameter sets that provided the best fit to the data. Simulations from the healthy young adult model are shown in Figure 6 and are not reproduced here to improve the readability of the graphs.

The ADH, as implemented by weak associations, provides a very good fit to the data. Noisy associations provide a good fit to most of the curve, but under–predict the age deficit on primacy items. The combination of weak *and* noisy associations provides a very good fit.

The IDH as implemented by allowing the contextual drift parameters to vary was able to fit all aspects of the SPC with the exception of overestimating the age deficit at serial position 1. Reducing contextual drift between lists fits the age deficit for most of the serial position curve but underestimates the deficit on the primacy portion. Finally, changing the post-retrieval editing threshold impairs primacy items but has little impact of rest of the SPC.

For the CSH, the slowed encoding implementation provides a good fit to older adults' SCPs. The slowed retrieval implementation also provides a good fit, though it does overestimate the age difference at the first serial position. When combined, slowed encoding and slowed retrieval provide a very good fit at all serial positions.

Probability of First Recall

Figure 8 shows the results of fitting the lesioned models to PFR curves, ignoring all other effects. A quick examination of the simulations speaks for itself: all of the models provided excellent fits to older adults' performance and are indistinguishable from the actual data. This is not surprising given that the PFR exhibits no age differences. This lack of an age difference essentially guarantees that the lesioned models will be able to simulate the older adult data.

Lag Conditional Response Probability

Figure 9 shows simulated Lag-CRP curves generated by the best–fitting lesioned models. The weak association implementation of the ADH underestimated older adults' deficit on +1 transitions. Adding noise to new associations does allow the model to fit the older adult data. The combination of weak and noisy associations provides a very good fit to the older adult data.

For the IDH, both the list isolation and the retrieval editing implementations were unable to capture older adults' deficit. The unregulated contextual drift implementation, however, provides good fit to the lag-CRP data. To understand how the model achieves this fit, we can compare the values of drift rate parameters that produced the simulated data with the corresponding values in the younger adult model (See Table A1 and A2 for all parameter values). This comparison reveals that the drift rate during encoding was somewhat lower than for the healthy model ($\beta_{enc}^{younger}=0.561$ and $\beta_{enc}^{lesioned}=0.532$). The drift rate during retrieval was substantially lower in the lesioned model ($\beta_{rec}^{younger}=0.375$ and $\beta_{rec}^{lesioned}=0.252$). A lower value of β_{rec} means that each newly retrieved item injects less of its associated context into the context layer, and therefore has less of an influence on the state of context that will be used to cue the next recall. As a consequence, the updated context representation will be a weaker cue for items that were presented in temporal proximity to the just–recalled item than it would be with higher values of β_{rec}. Note, however, that this is actually the opposite of the intuition that drove this implementation of the IDH. IDH seems to suggest that older adults should integrate *more* contextual information into mental context than younger adults, not less as the simulation suggests.

For the CSH, slowed encoding processes underestimated the extent of the age effect. Similarly, slowed retrieval processes over–predict older adults' performance on the critical +1 lag. The combination of slowed encoding and retrieval, however, provides a near perfect fit.

Prior-List and Extra-List Intrusions

Figure 10 shows the simulated PLI and ELI data for each lesioned model. For the ADH, the weak association implementation provided a reasonable fit. The noisy association implementation was able to capture older adults' PLI rate but not their ELI rate. However, the combination of both weak and noisy associations allowed the model to capture PLIs and ELIs.

For the IDH, both the list isolation and the retrieval editing implementations failed to capture the older adult ELI rate. The context drift implementation of the IDH—the version that allowed older and younger adults to have different rates of context drift both during encoding and during retrieval was able to capture both the PLI and the ELI rates.

For the CSH, the slowed encoding implementation provided a reasonable fit. The slowed retrieval implementation failed to fit either the PLI or the ELI rates. Slowing both encoding and retrieval process provided a very good fit to both the PLI and ELI data.

Prior-List Intrusion Recency

The results of simulating the PLI Recency effect are shown in Figure 11. As with the PFR, there were no behavioral age differences in PLI Recency, and as such, all of the lesioned models provided excellent fits to the data.

Simulating All Effects

Examining the effects in isolation has demonstrated that all three theories have at least one implementation that can capture each age effect. That is, each of the theories is able to pass the first hurdle of accounting for the individual effects with a reasonable level of quantitative precision. This finding represents an important step in building aging theories as most previous treatments of these theories have accounted for broad patterns at a qualitative level, and those that have used models to test quantitative predictions have mainly considered effects that the theory was specifically designed to account for (e.g., Li et al., 2005).

Before any of the theories can be considered to provide an adequate description of age differences in free recall, a second, more challenging hurdle must be crossed. Specifically, the theories must be able to account for the above effects *simultaneously* rather than in isolation. Meeting this challenge is critical because a complete aging theory must account for all of the effects with a common set of parameter values. In other words, because the aging brain uses the same configuration to produce SPCs, PFRs, lag-CRPs, and intrusion effects, a model of aging must also use a single configuration (i.e., the same parameter values) to produce all of the effects. It is, however, possible that the parameter values that allow a theory to fit one effect provide a very poor fit to another effect. Therefore, we tested the ability of each lesioned model to simulate all effects with a single set of parameters. To find the best–fitting parameter set for each theory we used the same grid search approach employed with the isolated effects (the best–fitting parameter values for all simulations are reported in Appendix B).

Figure 12 shows the performance of the various ADH implementations. Weak associations (top row) capture the age effect in the primacy portion of the SPC, but not the recency portion. Unsurprisingly the model does capture both the PFR and the PLI recency effect, which show no age differences. However, it fails to account for the CRP and for PLIs and

ELIs. The noisy association implementation also performs poorly, failing to capture any effects save the lack of age differences on PFR and PLI recency. The combined weak and noisy associations implementation is able to capture the PLI and ELI data but does so at the cost of severely mispredicting the SPC, lag-CRP and PLI recency data.

Figure 13 shows the IDH simulations. Unlike when each effect was considered in isolation, none of the three implementations were able to capture all of the effects simultaneously. The top row of Figure 13 reveals that allowing context drift rates during encoding and during retrieval to differ between older and younger adults fails to capture several effects. For the SPC, it accurately captures the age deficit in the recency portion but overestimates the effect in the primacy portion. It is also unable to simulate older adults impaired temporal contiguity. The model does, however, accurately simulate older adults' PFR, intrusion, and PLI recency data. The second row shows that the impaired list–isolation editing model was able to simulate older adults' PLI rate, but was unable to capture either the SPC, lag-CRP, or ELI deficits. Similarly, the impaired post–retrieval editing model (Figure 13 bottom row), fails to capture either the SPC, the lag-CRP, or the intrusion effects.

The performance of the CSH is shown in Figure 14. Slowed encoding processes (top row) fail to account for the SPC, lag-CRP and intrusion data, as does slowed retrieval processes. Slowing both encoding and retrieval provides a better fit, especially to the SPC, but still misses several key aspects of the data. The model accurately simulates older adults' temporal contiguity deficit for positive lags, but at the cost of underestimating their deficit at negative lags. The slowed encoding and retrieval implementation also fails to account for the PLI data. Surprisingly, this implementation also distorts older adults' PLI recency data, predicting a non–monotonic function (the dip at serial position 2).

In summary, while each theory is able to accurately simulate some aspects of the full pattern of age effects, in the process of doing so it severely mispredicts other aspects.

Each implementation's fit to the full set of effects is summarized in Table 1, which shows Bayesian Information Criterion (BIC; Kahana, Zhou, Geller, & Sekuler, 2007; Schwarz, 1978) values for each model. Lower BIC values indicate a better fit and take in to account the number of free parameters in a model. Table 1 reveals that none of the models performs as well as the full model (i.e., the model with all parameters free to vary), even though the full model is heavily penalized for having many parameters. When viewed in light of the fact that CMR2 can easily capture the age pattern when all parameters are allowed to vary (Figure 6), the failure of the aging theories to capture the effects by varying parameters associated with a single process suggests that age-related changes in multiple memory processes underlie the age pattern. This finding is consistent with other work suggesting that aging involves changes in multiple memory processes (e.g., K. J. Mitchell & Johnson, 2009). The challenge is to determine which processes can account for the pattern of age differences on free recall.

Multi-Component Theories

A sensible starting point is to ask if some combination of the IDH, ADH, and CSH theories can account for the full pattern of age effects. We explored this possibility by allowing all of the parameters implicated by the theories to vary freely. Specifically, γ_{FC}, γ_{CF}, ζ, β_{enc}, β_{rec}, β_{post}^{recall}, c_{thresh}, and τ were allowed to vary. The multi-component model includes 8 free parameters, which makes the parameter space too large to efficiently explore with a grid search. We therefore used a genetic algorithm, which past work has shown to be quite effective at finding optimal parameter sets for the CMR family of models (Healey & Kahana, 2014; Lohnas et al., 2015; Polyn et al., 2009; Sederberg et al., 2008). The genetic algorithm started by simulating 1000 parameter sets evenly distributed across the parameter space. The algorithm was then allowed to run for 10 generations, with each generation simulating 500 parameter sets. A further 10 generations each simulated 100 sets but with each simulation producing 5 times the true number of trials, allowing the algorithm to distinguish subtle differences between parameter sets. The mutation rate, which determines the standard deviation of the random changes (drawn from a normal distribution with mean zero) introduced to the parameters in each set, was systematically lowered across generations so that the algorithm would make large changes to explore the full parameter space in early generations and make progressively smaller changes to fine tune good–fitting parameter sets in later generations. Figure 15 shows the simulated data from the best–fitting parameter set (see Appendix B for the best–fitting parameter values).

This combined model provided a very good fit to the SPC, the PFR, both PLIs and ELIs, and PLI-recency. However, the model provided a poor fit to older adults' contiguity deficit, over–predicting the extent of deficit for +1 lags. Older adult's temporal contiguity deficit is perhaps the single most important aspect of the pattern of age effects on free recall for several reasons. First, as we have argued above, the ability to reproduce temporal order represents a key feature of episodic memory—the ability to place events on an autobiographical timeline. A reduced ability to situate events in their temporal context, a blurring of the autobiographical timeline, is likely a large part of what makes memory difficulties so troubling for older adults. Second, among younger adults, the extent to which an individual shows temporal contiguity is correlated with general cognitive ability, as measured by WAIS IQ (Healey, Crutchley, & Kahana, 2014). Therefore, a reduced temporal contiguity effect may be closely related to deficits on various non–memory tasks. In this light, the failure to accurately predict older adults' temporal contiguity deficit poses a serious challenge to a theory of aging. The BIC values in table 1 confirms that the full model, allowing all parameters to vary, provided a superior fit to the data than did the combined aging theory model.

As we discussed above, there are many decision points when implementing a verbal theory in a model. Because it is impractical to explore all possible implementations, we choose to consider three versions of each theory. Given that none of these implementations was able to capture the data it is natural to wonder if implementations we did not consider would account for the effects. We can make this question tractable by first finding the key model mechanisms that allow the full model to simulate age differences and then asking whether these mechanisms map onto the verbal theories.

What are those key mechanisms? A simple approach would be to examine the parameter values used to fit the younger versus older adult data shown in Figure 6, and determine which parameters differ substantially between age groups. But there is no way to test if the parameters are *significantly* different between age groups. We can overcome this challenge by moving from fitting average data to fitting the data of each individual participant. Doing so provides us with an estimate of each of the model parameters from each participant. We can use these individual parameter estimates to form a distribution of values for each parameter, which can be broken down by age group. We can then determine for which parameters the older and younger adult distributions differ significantly.

We fit individual participant data using the same genetic algorithm procedure described for average data. To ensure stable parameter estimates, we repeated the fitting procedure 3 times for each participant and averaged the estimates for each parameter and then used these average estimates to generate simulated data.

To visualize age differences on the parameter distributions, we placed all of the parameters on a common scale by converting them to z-score units (ignoring age group). Then for each parameter we calculated a young minus older adult difference score. These difference scores are plotted in Figure 16 along with 95% confidence intervals on the age differences (see Appendix B for the raw parameter values). The figure reveals age differences on the magnitude and decay rate of the primacy effect (ϕ_s and ϕ_d), on the context drift rate during retrieval (β_{rec}), the editing threshold (c_{thresh}), and the lateral inhibition and noise parameters in the decision accumulator (λ and η).

To confirm that these 6 parameters are indeed capable of accounting for the pattern of age differences, we used the same lesioning procedure we used to test the aging theories. Specifically, we took the same younger adult model we lesioned when testing the verbal theories and attempted to simulate the older adult data allowing only these 6 parameters to vary. As can be seen in Figure 17, we achieved an excellent fit to the older adult data. The BIC values in Table 1 reveal that this 6–parameter model outperformed all of the aging theories. That is, a reduced model consisting of only 6 parameters provided a better fit than the 8 parameter combined aging theory model. Although promising, we must be aware of several potential pitfalls.

First, even though the 6–parameter model has fewer free parameters than the combined aging theory model, it may be that the 6–parameter model is more flexible (i.e., it may be able to fit a broader range of possible datasets). The contribution of flexibility can be reduced by fitting a model to one subset of a dataset and testing the fitted model's ability to simulate an independent subset. An overly flexible model may (over)fit the first subset, but at the expense of mis-predicting the independent subset. Therefore, to compare the two models while controlling for flexibility, each model was fit to a random half of the participants in the Kahana et al. (2002) dataset and the simulated data was compared to the real data from the held-out half of participants by computing the root mean square deviation

(RMSD) of the simulated data from the real data. This procedure was repeated 20 times for each model (fitting a different random half of participants each time) producing distributions of RMSD values for each model; the best model should have a lower mean RMSD value. The mean RMSD value for the 6–parameter model ($M = 0.039$, $SD = 0.006$) was indeed lower than the mean for the combined aging theory model ($M = 0.053$, $SD = 0.006$). Although a t-test comparing these distributions is highly significant, we do not report a p-value as it can be made arbitrarily small by increasing the number iterations of the split-half procedure and does not account for the fact that each iteration draws from the same dataset.

Second, we need to ensure that the correlations between parameter values are not overly high. If particular parameters were highly intercorrelated (i.e., a participant's estimate for one parameter predicted their estimate for another), it would suggest that the two parameters do not represent independent processes and should be considered as a single parameter. Two parameters, lateral inhibition and the standard deviation of noise in the decision accumulator, were correlated at $r = .93$. In other words, $.93^2 = 86\%$ of the variance in the noise parameter is explained by the lateral inhibition parameter suggesting the parameters are not capturing unique sources of information. Therefore, when discussing these two parameters, we avoid interpreting them as reflecting distinct age deficits (though they were allowed to vary independently in fitting all models). No other parameters had a correlation above $|r| = .63$ (i.e., less than 40% shared variance; see Appendix B for the full correlation matrix).

A third potential pitfall we must avoid is the possibility that the reduced 6–parameter model we have found is but one of many reduced models that are capable of fitting the data. If it were the case that many different reduced models provided simulations of roughly equivalent accuracy (i.e., the model is underdetermined), then it would simply be telling us that there is not enough data to distinguish between different reduced models. We took three steps to test the validity of the 6–parameter model.

First, we attempted to determine if any of the six parameters we identified are necessary to simulate the age pattern. We did this by attempting to fit the pattern using all of the model parameters *except* those in the reduced model. This model consisted of 9 parameters and failed to fit the data ($BIC = -163$), suggesting that at least some of the parameters in the reduced model are critical.

Second, to determine if any of the six parameters are unnecessary, we ran a step–wise leave-one-out procedure. Starting with the six parameters we tested whether any of those parameters could be removed by testing the ability of all possible 5 parameter models to fit the data (i.e., allowing five of the original six to vary and fixing the 6^{th} to the younger adult value). We compared the six–parameter model with each of the five–parameter models: The best–fitting five–parameter model excluded the editing threshold parameter (i.e., c_{thresh} was held at the younger adult value). However, this model had a lower BIC value than the 6–parameter model ($BIC = -230$ for the six–parameter model and $BIC = -219$ for the five–parameter model). The remaining 5–parameter models had BIC values ranging from −157 to −173. That is, the parameters excluded from the six–parameter model are unable to simulate the older adult data and excluding any of the six parameters from the reduced model degrades the quality of its fit.

Finally, we tested if we could substantially improve the fit of the six–parameter model by allowing a 7^{th} parameter to vary. We tested this possibility by fitting all possible seven–parameter models (i.e., those in which the parameters from the original six–parameter model, plus one of the remaining parameters were allowed to vary). The best–fitting seven–parameter model allowed ζ, which controls the amount of associative noise, to vary and resulted in only a small improvement in fit ($BIC = -230$ for the six–parameter model and $BIC = -231$ for the seven–parameter model). We now turn to interpreting the six–parameter model.

The parameters in the reduced model map directly on to psychological constructs. The age differences in the primacy parameters, ϕ_s and ϕ_d, suggest that older adults give a strong attentional boost to early items but that their attention drops off more rapidly than younger adults. This pattern is evident in Figure 18, which plots the age difference in the primacy gradient (i.e., the initial boost and how it decays across items) derived from the model parameters. The figure shows that older adults have a stronger initial boost than younger adults (due to a higher value of ϕ_s), but their decay rate (ϕ_d) is so high that by the second item younger adults are receiving a greater primacy boost. One interpretation of this pattern is that older adults initially allocate too much attention to encoding and then suffer from rapid camatosis (i.e., decreasing efficiency of encoding processes across a trial; Serruya et al., 2014; Tulving & Rosenbaum, 2006). The idea that older adults have difficulty allocating attention is consistent with evidence that older adults have difficulty with attention tasks (Darowski et al., 2008), and also with the idea that frontal executive processes are impaired with age (West, 1996). We note that the fact that the IDH, ADH, and CSH do not suggest any age differences in the allocation of attention across time may explain part of their difficulty in fitting all of the effects.

The next mechanism showing an age deficit is the rate of context drift during retrieval. This drift rate, β_{rec}, governs the extent to which the just–recalled item reinstates its associated contextual states and was lower for older adults. Contextual reinstatement lies at the core of the retrieved context framework, and is a key component in producing the contiguity effect: Because nearby list items become associated with similar contexts during encoding, if a retrieved item reactivates its context, it will provide a strong cue for its list neighbors. In contrast, if a retrieved item reactivates little of its associated context, it will provide a relatively weak cue to its neighbors, thus reducing the temporal contiguity effect. The fact that context drift differs between older and younger adults only at retrieval helps explain the fact that there are no age differences on the PFR—if older adults had lower context drift rates during encoding, CMR2 would predict they would have shallower PFR curves than younger adults, as the end–of–list context would provide a strong match to a wider range of items Howard et al. (2006). Insofar as contextual reinstatement during retrieval represents a deliberate attempt by participants to create an effective retrieval cue, age differences in this drift rate are consistent with the view that older adults suffer at self–initiated processing (Craik, 1977, 1983) or the ability to use well–specified cues to constrain retrieval (Jacoby, Bishara, Hessels, & Toth, 2005). It is also consistent with the suggestion that regulation of contextual representation allows for the selective retrieval of relevant memories, which we have termed contextual gating (Healey, Crutchley, & Kahana, 2014).

The age difference in the post–retrieval editing threshold suggests that older adults are more willing to accept intrusions as true list items. As discussed above, one interpretation of the IDH is that older adults have difficulty rejecting intrusions due to deficits in the restraint function of inhibition (Hasher et al., 1999) that can be modeled as a lower editing threshold. Although our simulations of the IDH showed that this deficit cannot provide a complete account of age differences in free recall, our 6–parameter model suggests that an inhibitory deficit is a critical component. The suggestion that older adults use a lower editing threshold is consistent with the finding that older adults tend to endorse more intrusions in externalized free recall (in which participants say all words that come to mind, and reject those that they know are errors; Kahana, Dolan, Sauder, & Wingfield, 2005). An impaired ability to monitor for intrusions is consistent with evidence that older adults have difficulty monitoring retrievals in other tasks such as autobiographical recall (McDonough & Gallo, 2013). An age difference in retrieval editing may seem inconsistent with our simulation of the retrieval editing implementation of the IDH, which showed that the editing threshold parameter, by itself, was not able to simulate older adults' intrusion rates (Figure 7). However, for intrusions to reach the editing stage, they must win a retrieval competition, and evidence from externalized free recall suggests that older adults produce more potential intrusions than do younger adults (Kahana et al., 2005). The retrieval editing mechanism has no influence on how many non–list items win retrieval competitions; some other deficit must provide the editing mechanism with a supply of potential intrusions. The next impaired mechanism does just that.

Older adults had higher estimates of the lateral inhibition and noise parameters of the decision accumulator process (λ and η). We stress that these two parameters are so highly correlated that they cannot be treated as refelecting separate age deficits. Instead the fact that older adults had high values for both parameters suggests that the evidence accumulation process is noisy for older adults, which must be balanced by stronger lateral inhibition. A noisy accumulation process would allow items that would seldom win a retrieval competition in the younger adult model to occasionally accumulate enough evidence to win. The likelihood of an item accumulating enough evidence to be retrieved is codetermined by the strength of genuine evidence for the item (given by the match between the context cue and the item) and the noise. Therefore, a higher variance in the noise parameter means that an item with little genuine support will occasionally have enough spurious support from noise to win the retrieval competition. Conversely, given that the noise parameter can take on negative values, items with high genuine support will occasionally lose due to noise. Thus, a higher value of the noise parameter will lead to fewer accurate recalls and more false recalls being passed on to the post–retrieval editing phase. Moreover, noise will tend to disrupt retrieval dynamics. Normally items that were near the just–recalled item in the presentation list will receive the most support in the retrieval competition (because they are associated to states of context similar to that reinstated by the recalled item). With a high degree of noise, however, distant items will more often accumulate sufficient evidence to be recalled, which will tend to lower the extent of temporal contiguity.

Together, the age differences in the CMR2 parameter estimates suggest four distinct age–related impairments: Rapid waning of attention across the list; difficulty retrieving contextual representations during the recall period; a lower threshold for endorsing candidate retrievals; and a noisy evidence accumulation process.

Validating the four–component model

The simulations presented so far have shown that whereas several existing theories are unable to account for the details of age differences on the free recall task when implemented in a computational model, a novel four–component theory is able to account for these differences. Fitting a single data set, however, is not enough to validate a theory: it is possible that our method for identifying which parameters differ between older and younger adults produced a model that is tailored to unique aspects of the Kahana et al. (2002) dataset and the four–component model would fail to capture age differences on an independent dataset. For example, variables such as list length, item presentation time, and participants' familiarity with the task are known to influence memory performance (Brodie & Murdock, 1977; Dallett, 1963; Goodwin, 1976; Grenfell-Essam & Ward, 2012; Hasher, 1973; Huang, 1986), and may also influence the ability of a model to account for age differences. We therefore test the generalizability of the model by evaluating the ability of the four–component model identified using the Kahana et al. (2002) dataset to account for age–differences on a novel dataset collected using different task parameters.

Just as a valid model of aging should be able to account for the pattern of age differences on independent data sets, it should be able to generalize beyond the set of benchmark effects that were used to develop it. Therefore, we will add an additional effect to our set of free recall benchmarks: the semantic contiguity effect, which shows participants tendency to cluster recalls based on long-standing semantic associations (Howard, Addis, Jing, & Kahana, 2007; Howard & Kahana, 2002b; Sederberg, Miller, Howard, & Kahana, 2010). Older adults show no deficit on this effect, and thus the four–component model will be challenged to simultaneously account for older adults' impaired use of new temporal associations and their intact use of semantic associations. Finally, we will extend the model beyond free recall and show that it can, using parameter values optimized on a free recall task, simultaneously account for age differences in recognition memory performance.

For this purpose, we used data from Experiment 1 of the Penn Electrophysiology of Encoding and Retrieval Study (PEERS). PEERS aims to assemble a large database on the electrophysiological correlates of memory encoding and retrieval. Previous work has reported data from younger adults (Healey, Crutchley, & Kahana, 2014; Healey & Kahana, 2014; Lohnas & Kahana, 2013, 2014a; Miller, Kahana, & Weidemann, 2012); here we report data from a new sample of older adults.

PEERS Methods

Participants—The present analyses are based on the participants who had completed Experiment 1 of PEERS as of December 2013. The sample included 156 college students (age range: 18–30 years, $M = 22.39$, SD = 3.00) and 38 older adults (age range: 61–85 years, $M = 69.26$, $SD = 6.48$). The older adults had an average of 17.82 ($SD = 2.51$) years of education and an average score of 48.93 (SD = 6.61) on the WAIS IV (Wechsler, 2008) vocabulary test; the younger adults had an average of 14.94 (SD = 1.98) years of education and an average vocabulary score of 50.08 (SD = 4.78). The age-related differences in education were statistically significant, but those in vocabulary were not. The vocabulary test was administered during a separate session that was not completed by all participants, therefore vocabulary data was missing for 9 older adults and 37 younger adults. As described in detail elsewhere (Healey & Kahana, 2014; Lohnas & Kahana, 2013), PEERS Participants were right-handed native English speakers who, during an introductory session, did not make an excess of eye movements during item presentation epochs and had a probability of recall less than 0.8. To ensure we do not confound effects of healthy aging with pathological deficits, older adults were extensively pre-screened for signs of pathology using a detailed medical history and the Short Blessed Test (Katzman et al., 1983).

PEERS Experiment 1—For full details on the design of PEERS see (Healey & Kahana, 2014; Lohnas & Kahana, 2013). Here we focus on the immediate free recall and recognition data. Participants performed a free recall experiment consisting of 1 practice session and 6 subsequent experimental sessions. Each session consisted of 16 lists of 16 words presented one at a time on a computer screen. Each study list was followed by an immediate free recall test. At the end of each session there was a recognition test and, for a subset of sessions, a final free recall test.

Words were either presented concurrently with a task cue, indicating the judgment that the participant should make for that word, or with no encoding task. The two encoding tasks were a size judgment ("Will this item fit into a shoebox?") and an animacy judgment ("Does this word refer to something living or not living?"), and the current task was indicated by the color and typeface of the presented item. There were three conditions: no-task lists (participants did not have to perform judgments with the presented items), single-task lists (all items were presented with the same task), and task-shift lists (items were presented with either task). List and task order were counterbalanced across sessions and participants.

Each word was drawn from a pool of 1638 words. Lists were constructed such that varying degrees of semantic relatedness occurred at both adjacent and distant serial positions. Semantic relatedness was determined using the Word Association Space (WAS) model described by Steyvers, Shiffrin, and Nelson (2004). WAS similarity values were used to group words into four similarity bins (high similarity: $\cos\theta$ between words > 0.7; medium–high similarity, $0.4 < \cos\theta < 0.7$; medium-low similarity, $0.14 < \cos\theta < 0.4$; low similarity, $\cos\theta < 0.14$). Two pairs of items from each of the four groups were arranged such that one pair occurred at adjacent serial positions and the other pair was separated by at least two other items.

For each list, there was a 1500 ms delay before the first word appeared on the screen. Each item was on the screen for 3000 ms, followed by jittered (i.e., variable) inter-stimulus interval of 800–1200 ms (uniform distribution). If the word was associated with a task, participants indicated their response via a keypress. After the last item in the list, there was a jittered delay of 1200–1400 ms, after which a tone sounded, a row of asterisks appeared, and the participant was given 75 seconds to attempt to recall aloud any of the just-presented items. If a session was randomly selected for final free recall, following the immediate free recall test from the last list, participants were shown an instruction screen for final free recall, telling them to recall all the items from the preceding lists. After a 5 second delay, a tone sounded and a row of asterisks appeared. Participants had 5 min to recall any item from the preceding lists.

A recognition test was administered after either final free recall or the last list's immediate recall test. In this final recognition test, lures were selected from the items remaining in the original 1,638-word pool that not presented for study during the free recall phase. The target/lure ratio varied with session, with targets making up 80%, 75%, 62.5%, or 50% of the total items. In total, 320 words were presented one at a time on the computer screen. When a word was presented on the screen, participants were instructed to indicate whether the test word had been presented previously. Participants were told to verbally respond "pess" for old items and "po" for new items and to confirm their response by pressing the space bar. These responses ("pess" and "po") were chosen so that both response types would initiate with the same stop consonant (or plosive), thus assisting in automated detection of word onset times. Following the old-new judgment, participants made a confidence rating on a scale of 1 to 5, with 5 being the most confident. Recognition was self-paced, although participants were encouraged to respond as quickly as possible without sacrificing accuracy. Participants were given feedback on accuracy and reaction time.

PEERS Free Recall Data

Behavioral results—Figure 19 shows age differences on the five key effects we considered in the previous simulations. This new dataset replicates all of the effects. Older adults show impaired recall throughout most of the SPC. One notable difference between the PEERS data in Figure 19 and the Kahana et al. (2002) data in Figure 2 is the lack of an age difference at the final serial position in the former. The PEERS data is consistent with studies that have found that age differences in the SPC are smaller for recency items than for mid–list and primacy items (for a review see, Craik & Jennings, 1992). We return to this issue in the discussion. The PEERS data also replicate the lack of age differences in the PFR, and the robust deficit in temporal contiguity. which is largest for close temporal transitions (lags of +1 and −1). The intrusion data replicate the findings that older adults show more PLIs and ELIs than younger adults but have an intact PLI-recency effect.

Our simulations of the verbal aging theories have shown that testing a theory requires evaluating its ability to account for multiple effects simultaneously. Therefore, rather than conduct a simple cross–validation in which we test the ability of our four–component model to account for exactly the same set of effects in a new dataset, we wanted to include additional data points that were not used in developing the model. To maximize the value of the cross–validation, we wanted to consider an effect that has played an important role in the aging literature. As we discussed above, the extent to which aging spares long-standing, or crystallized, semantic knowledge while impairing new episodic associations has been an important topic in the aging literature. Therefore, we test the ability of the four–component model to simulate the effect of existing semantic knowledge on recall. Typically this is done with fluency tasks (e.g., Loonstra et al., 2001; Troyer et al., 1997), but again we take the approach of using the free recall task to do so. We can examine the influence of long–standing semantic associations on transition probabilities (Bousfield, 1953; Romney, Brewer, & Batchelder, 1993) using Latent Semantic Analysis (LSA; Landauer & Dumais, 1997), which measures the proximity of words in a multidimensional model of semantic space. Using LSA values to create a semantic-CRP curve (Figure 19F) reveals a strong semantic contiguity effect (Howard & Kahana, 2002b). Consistent with work on fluency, the semantic-CRP reveals that older adults show a preserved ability to use semantic associations to guide recalls[2].

Simulations—When participants are first exposed to a new task, they must learn how to best use their memory system to meet the demands of the task. We have previously argued that a large part of such learning may consist of tuning parameters of the memory system (Healey, Crutchley, & Kahana, 2014; Healey & Kahana, 2014). For example, participants may be able to tune the extent two which they rely on preexisting semantic information. A free recall task that uses truly random word lists, such as Kahana et al. (2002), may require participants to tune the parameter to a fairly low setting. For a task that includes some moderately related words in each list, such as PEERS, the value may be set to a somewhat higher level. Other differences between tasks might affect other parameters of the system. Therefore, it is unlikely that exactly the same model parameters that allowed us to simulate the Kahana et al. (2002) data will provide a good fit to the PEERS data. Instead, we want to capture these differences in parameter tuning levels by first finding the optimal set of parameter values for simulating *younger* adult data from PEERS and then determine if we can simulate older adult PEERS data by fixing all parameters at the younger adult values except for those implicated in the four–component model.

We have recently shown that almost all individual younger adults have SPCs, PFRs, Lag-CRPs, and semantic-CRPs that are well described by the average curves (Healey & Kahana, 2014). This fact makes it possible to reduce computational demands by fitting the model to a subsample of younger adults. Specifically, we randomly selected 38 younger adults (i.e., to match the older adult sample size) with the constraint that the subsample did not differ significantly from the full sample on any of the data points in Figure 19 according to two–tailed τ–tests with α set at .15 to minimize false negatives. We then fit the model to the average data of this subsample using the same genetic algorithm approach used above. Figure 20 presents simulated data produced by the best fitting younger adult model. Just as we did with the Kahana et al. (2002) dataset, we can use these younger adult parameter values as a model of healthy memory. We can then take that healthy model and lesion the mechanisms implicated by the four–component theory of aging. We did so by allowing the 6 parameters implicated in the theory (ϕ_s, ϕ_d, β_{rec}, c_{thresh}, λ, and η) to vary while keeping all others fixed at the younger adult values. Figure 20 shows the simulated data from the best fitting parameter set. Replicating our simulations of the Kahana et al. (2002) data, the four–component model was able to capture all of the age effects. Moreover, the four–component

[2]For each list, a limited number of words fall into each semantic similarity bin. Thus participants who recall more words will naturally tend to have a higher probability of making transitions to any given bin, which will tend to increase the intercept (but not the slope) of the curve even in the absence of true differences in the use of semantic associations (this is less of an issue for the lag-CRP as there are many ways to make transitions at each lag). To control for this confound, we considered only the first 10 output positions to roughly equate age groups on the number of words recalled.

model provided an excellent fit to the new semantic-CRP data, capturing older adults' preserved ability to use long–standing semantic associations to guide recalls.

Finally, we conducted a similar cross-validation to confirm that the failure of the ADH, IDH, and CSH generalize to the PEERS dataset. To give the theories the best chance, we assumed that older adults suffer from all three deficits and allowed all 9 of the implicated parameters (γ_{FC}, γ_{CF}, ζ, β_{enc}, β_{rec}, β_{post}^{recall}, c_{thresh}, and τ) to vary from the optimal younger adult values. As can be seen in Figure 21 the combined aging theory model, although able to fit many of the data points, again underestimated older adults' temporal contiguity deficit. Comparing the BIC values for the four–component and the combined aging theory models (Table 1), confirms that the four–component model provided a superior fit to the data.

PEERS Recognition Data

Next we tested the ability of the 4–component model to generalize beyond free recall and account for age differences in recognition. In a standard item recognition task, a list of items are presented followed by a test phase in which the participant must indicate if a probe item was, or was not, a member of the list. These tasks typically show a modest age deficit in accuracy (we discuss age differences in reaction time below), which can be difficult to detect in studies with small sample sizes (Verhaeghen et al., 1993). The recognition task in PEERS has two features which make it well suited to detecting any age differences in accuracy. First, because participants completed multiple sessions, each of which included over 300 recognition probes, the study is sufficiently powered to detect modest age differences. Second, whereas in most recognition tasks the lag between studying an item and being tested on that item (in terms of both the amount of time and the number of intervening items) is relatively short, in PEERS the lag is on the order of dozens of items and minutes, which will tend to increase difficulty and minimize any ceiling effects that could mask age differences (For other examples of the use of post-recall recognition see Lohnas & Kahana, 2013; Merkow, Burke, & Kahana, Submitted).

Behavioral results—Because the recognition task followed a free recall test (and on some sessions, a final free recall test), probe items fall into three categories: 1) old items that were successfully recalled (either during free or final free recall), 2) old items that were not recalled, and 3) new (i.e., lure) items. Our analyses and simulations focus on old items and new items that were not recalled. We exclude old items that were recalled for two reasons. First, performance for these items is very near ceiling for both younger and older adults. Second, restricting analyses to items that were not recalled helps control for any age differences in output encoding (i.e., encoding information about items as they are recalled, as opposed to studied).

Figure 22 shows younger and older adults' hit rates for items that were not recalled along with their false alarm rates for lures. Replicating many previous studies (Craik, 1971; Jacoby, 1999; Ratcliff, Thapar, & McKoon, 2004; Schonfield & Robertson, 1966; Spaniol et al., 2006), there were no age differences in hit rate ($\tau(217) = 0.79$, $p = 0.43$). Older adults did, however, make reliably more false alarms than did younger adults ($\tau(217) = 4.82$, $p = 2.74 \times 10^{-6}$). Confirming this interpretation, there was a significant age × item type interaction ($\phi(1, 217) = 13.70$, $p = 0.0003$). That is, older adults show no impairment in endorsing old items, but are more willing to endorse lures as old items. This mirrors older adults' increased intrusion rate in free recall (Zaromb et al., 2006). Can our four–factor model account for the interaction using the same set of impairments that allow it to capture age differences on free recall?

Simulations—To date, the retrieved context framework has not been used to formally model recognition tasks. The CMR2 model, however, includes several mechanisms that would allow it to perform recognition. The challenge of recognition is to distinguish between old and new items both of which exist in the model's memory system but only one of which was studied during the experiment. This challenge is very similar to the challenge of distinguishing intrusions from genuine list items during free recall (K. J. Mitchell & Johnson, 2009). To screen for intrusions, the model allows each candidate retrieval to activate its associated contextual representation, which is then compared with the current state of context: if the two are similar (i.e., similarity passes a threshold, c_{thresh}) it is endorsed as a list item.

In exactly the same way, we can simulate recognition by allowing each probe item to activate its associated context state and comparing that to the existing state of context; if the similarity exceeds a threshold, c_{recog}, it is endorsed as an old item, if not it is rejected as a lure. This comparison can distinguish old from new probes because context drifts slowly and the state of context during the recognition phase of the experiment is expected to be somewhat similar to the state of context that prevailed during the study phase. Old probes were studied during the recall phase and should have formed an association with the study phase context. New probes were never studied and therefore have no association with the study phase context. Therefore, if a probe item is allowed to activate its associated context when it is presented during the recognition task, old probes would be expected to activate a state of context that is more similar to the current state of context than will new probes.

This thresholded context reinstatement mechanism corresponds closely with the process of recollection in dual process theories of recognition (Yonelinas, 2002), in that it depends on the probe item reactivating contextual details of past experiences with the item via feature-to-context associations (in other words, recollecting details of previous study episodes). CMR2 could also be used to simulate the familiarity process by simply comparing the probe's item representation to the current state of context, which would not involve the activation of any associative information.

Rather than use both the context-based and familiarity-based mechanisms, we have chosen to start with a minimal model of recognition based on the retrieved context comparison mechanism for two reasons. First, past work suggests that older adults have relatively intact familiarity processes (Jacoby, 1999). Second, attempting to fit the recognition data using exactly the same set of model mechanisms *and* parameter values used to simulate free recall provides a strong test of our four–component model. Adding a familiarity mechanism would require fitting new parameters.

To simulate the recognition phase of PEERS, we started with the model fit to the PEERS free recall data in the previous section (see table B4 for model parameters). Using these parameters, we simulated the free recall and final free recall phases. During encoding, we presented the model with the actual word lists studied by the participants. During retrieval (both the free recall period following each list, and the final free recall period), we constrained the model to recall the same sequence of items that the participant recalled. Thus, at the beginning of the recognition phase the model's associative matrices and context layer represent the expected state of the participant's memory system. As described above, when simulating free recall CMR2 allows context to drift between each list to simulate the disruption to context thought to occur when changing tasks (Sahakyan & Kelley, 2002). Following this logic, context was allowed to drift between the free recall phase and the final free recall period and again between the final free recall phase and the recognition phase. To allow us to simulate recognition without introducing any free parameters, the same context drift rate parameter, β_{post}^{recall}, used to update context between lists was used to update context between phases.

During the recognition phase, we allowed each probe item to activate its associated context state in exactly the same way items activate their associated context states during the study phase of free recall lists (see the appendix for a formal description). This reinstated context was then compared to the current state of context ($\mathbf{c}_{t+1}^{\mathbf{IN}} \cdot \mathbf{c}_t$), and if the similarity exceeded a threshold, c_{recog}, it was endorsed as an old item, if not it was rejected as a lure. Once an item was endorsed as an old item its associated context was integrated into the current state of context using the same context updating mechanism that is used during free recall. A parameter, β_{recog}, controls the drift rate of this updating process. No context updating occurs if the probe is rejected as a lure. This simple model of recognition is controlled by two parameters, c_{recog} and β_{recog}, that are directly analogous to c_{thresh} and β_{rec} which control post-recall editing and context drift during free recall. Under the four–factor model, older adults have lower values of both of these parameters indicating lower thresholds for endorsing items as being previously studied (c_{thresh}) and less integration of a retrieved item's context into the current context state (β_{rec}). To determine if the four–factor model can simultaneously account both for free recall and for the age × item type interaction in recognition accuracy without fitting any additional parameters, we took the best fitting parameters of the four–factor model and set $c_{recog} = c_{thresh}$ and $\beta_{recog} = \beta_{rec}$.

Figure 22 shows the results of these simulations: the four–factor model, using the same parameter settings used to simulate free recall, provided an excellent fit to the data, successfully capturing the age × item type interaction. A lower threshold, c_{recog}, on the context similarity comparison will make it more likely that any probe, be it an old item or a lure, will cross the threshold and thus will contribute to intact hit rates at the expense of increased false alarm rates. A reduced context drift rate, β_{recog}, also likely contributes to the model's ability to capture the age difference. Large values of β_{recog} mean that when an item is endorsed as previously studied, its context representation becomes strongly active on the context layer, largely replacing the previously activated context representation. If the probe item was indeed an old item, this newly activated representation will be similar to the context representations of other old items as they became associated to similar contextual states during study. Therefore, the higher the value of β_{recog}, the higher the expected match between the current state of context and the retrieved context of old probes, and, conversely, the lower the expected match with the context states of lure probes (i.e., old and new items become more distinguishable). By contrast, lower values of β_{recog} would lead to a smaller difference in the expected similarity for old and new items. Therefore, if the threshold, c_{recog}, is set at a level that ensures a high hit rate, a lower value of β_{recog} will tend to ensure that more lures also pass the threshold, increasing the false alarm rate.

The fact that we were able to simulate age differences in recognition using the same set of model mechanisms *and* the same parameter values used to simulate recall strongly suggests that participants do indeed use a common set of processes to screen for intrusions in recall and to distinguish old items from lures in recognition. The model makes a novel and testable prediction: participants who make more free recall intrusions should also make more false alarms in recognition. Confirming this prediction, there was a significant positive correlation between intrusion rates in free recall and false alarm rates in recognition (Figure 23). The correlation was significant when ignoring age group, $r(192) = .33$, $p < .001$, and the size of the correlation did not differ between age groups, $z = 1.35$, $p = .18$. The correlation was significant within the younger group ($r(154) = .33$, $p = 2.21 \times 10^{-5}$); within older adult group the sample size was much smaller and, as such, the confidence interval on the correlation encompassed both zero and the younger adult value, $-.23 \leq r(36) \leq .40$.

Discussion

Developing a detailed understanding of why episodic memory changes with age is amongst the most important tasks facing cognitive aging researchers. This understanding will form the foundation for interventions designed to prevent, slow, or remediate age–related memory declines. The pattern of age differences across tasks and within tasks is complex, with some aspects of performance showing stability, others showing modest decline, and others showing substantial impairment. Many theories have been proposed to account for age-related memory change, few however, have directly engaged with the multivariate nature of the pattern. Here we used a set of benchmark findings from free recall and recognition to show that theories which are able to account for many effects when each effect is considered in isolation, are unable to account for the effects simultaneously. To develop a theory that can simultaneously account for the full multivariate pattern, we simulated the performance of individual older and younger adults and compared parameter distributions across the age groups. We discovered four critical cognitive mechanisms that allowed the model to capture the full pattern of age differences: 1) the ability to sustain attention across an encoding episode, 2) the ability to reinstate contextual representations for use as retrieval cues, 3) the ability to monitor retrievals and reject intrusions, and 4) the level of noise in accumulating evidence during retrieval competitions. This four–component model captured age differences on an independent dataset. Moreover, the four–component model was able to capture age differences on a recognition task using the same parameter settings that allowed it to simulate free recall.

The Importance of Fitting Many Effects

In an initial set of simulations we considered each of the age effects in isolation and tested whether several existing theories, implemented as lesions to the healthy younger adult model, were able to simulate older adults' deficit on that effect (or lack of deficit in the case of recall initiation and the prior–list intrusion recency effect). Specifically we evaluated the Associative Deficit Hypothesis (ADH; Naveh-Benjamin, 2000), The Inhibitory Deficit Hypothesis (IDH; Hasher & Zacks, 1988), and the Cognitive Slowing Hypothesis (CSH; Salthouse, 1996). We found that all of the theories were able to account for each of the effects.

But accounting for several effects in isolation is not sufficient to establish the validity of a theory, because considering the effects in isolation allows the model to use different sets of parameter values when fitting each effect. The aging brain produces each of the effects with the same configuration, and we must expect a truly adequate theory of cognitive aging to do the same. Therefore we tested the ability of the theories to simultaneously fit all of the age effects with a single set of parameter values. No theory was able to account for the full pattern simultaneously. Both the ADH and the IDH were unable to capture the SPC and lag-CRP data simultaneously. The CSH, implemented as slowed encoding and retrieval processes, faired somewhat better, and was able to accurately simulate older adults SPCs. It was, however, unable to fully capture their temporal contiguity deficit, predicting an age deficit only for forward, but not backward, transitions. It also failed to account for older adults' PLI rates. Examining Table A2 shows why these models had difficulty: The parameter values that allowed these models to capture the effects when considered in isolation varied considerably from effect to effect and were in general quite different than the parameter values that allowed the model to make the best compromise when fit to the effects simultaneously. In other words, there was no single model configuration that was able to simulate all aspects of older adults' performance.

The ADH, IDH, and CSH each implicate a specific memory process as the locus of age differences. The fact that none of these theories were able to simultaneously account for the full pattern of age differences suggests that a multi–component theory is needed. A logical first step in developing such a theory is to combine single–component theories. Surprisingly, even when all of the parameters implicated by the ADH, IDH, and CSH were allowed to vary, we were unable to achieve a satisfactory simulation of the age differences. In particular, the combined theories were unable to capture older adults' temporal contiguity deficit. Again, the reason for this is apparent from inspecting the parameter values in Table A1: The values for the combined theories model all differ substantially from the parameter values that allowed individual implementations to simulate the temporal contiguity deficit when considered in isolation, which are in turn quite different than the values required to allow the implementations to fit the other age effects.

It is noteworthy that simulating older adults' temporal contiguity deficit posed the greatest challenge to the theories. As we have argued above, temporal contiguity reflects the ability to place events in their temporal context, and a deficit in this ability may blur the autobiographical time-line. Moreover, in light of the fact that strong temporal contiguity effects predict recall success (Healey, Crutchley, & Kahana, 2014; Sederberg et al., 2010;

Spillers & Unsworth, 2011) and higher IQ (Healey, Crutchley, & Kahana, 2014), a deficit in temporal contiguity may be related to age deficits on non–memory tasks. Therefore, the temporal contiguity deficit is perhaps the single most important aspect of the pattern of age effects.

A Four–component theory of cognitive aging

To develop a theory that can account for all of the effects, including the critical temporal contiguity deficit, we allowed the data speak for themselves and tell us which model mechanisms are essential in capturing the aging pattern. We simulated each individual participant's data and examined the resulting across-participant distributions of parameter values and found that 6 model parameters showed showed significant age differences. Next, we showed that it is possible to fit the full pattern of age differences by allowing these 6 critical parameters to vary and fixing all other parameters at the younger adult values. Moreover, the fit of the 6–parameter model could not be improved by allowing any additional parameters to vary, but was hurt by fixing any of the 6 parameters to the younger adult values. This 6–parameter model involves two fewer free parameters than did the combined ADH, IDH, and CSH model (8 total parameters). Moreover, in a split-half cross-validation test, the 6–parameter model provided a significantly better fit than did the combined aging theory model. Thus, the fact that 6–parameter model provided a superior fit is not simply a consequence of greater model flexibility.

The 6 parameters in our model of aging correspond to 4 distinct memory processes: 1) the ability to sustain attention across an encoding episode, 2) the ability to retrieve contextual representations for use as retrieval cues, 3) the ability to monitor retrievals and reject intrusions, and 4) the level of noise in evidence accumulation during retrieval competitions.

The first component needed to simulate the age differences was the primacy gradient. Two parameters (ϕ_s and ϕ_d) control the extent of the primacy boost and how rapidly it decays. Older adults tended to show a somewhat stronger initial primacy boost than young adults, but one that decays rapidly (Figure 18). This pattern suggests that older adults have difficulty sustaining attention over the course of a trial. The notion that older adults have difficulty sustaining the efficiency of encoding processes across a trial is broadly consistent with theories that implicate age differences in attentional (Craik, 1977; Hasher & Zacks, 1988) or frontal executive processes (West, 1996). Our simulations allow a refinement of the general claim that older adults have impaired attentional processes by suggesting that the attentional deficit is specifically related to the ability to *sustain* attention.

The next component was an impaired ability to reinstate contextual states during recall. One parameter, (β_{rec}), governs the extent to which contextual representations are reactivated during retrieval of items. The ability to reactivate contextual states is at the heart of CMR2's ability to simulate retrieval dynamics as each item reactivates its associated context, which in turn serves as a retrieval cue for items near it in the presentation list. A reduced ability to retrieve contextual states that provide effective cues suggests a mechanistic basis for the self–initiated processing deficits from which Craik and colleagues have argued that older adults suffer (Craik, 1977; Craik et al., 2010). Some early modeling work suggested that older adults have a more slowly drifting context representation (Balota, Duchek, & Paullin, 1989). However, as noted by Howard et al. (2006), a specific impairment in the ability to retrieve contextual states, as opposed to a general difference in drift rate, provides an elegant account of the fact that older adults show no impairment in initiating recall but do show a deficit in subsequent temporally–mediated transitions: Under the retrieved context framework, when the recall period begins, the existing state of mental context is used to cue recall with no need to reactivate previous states, only subsequent recalls depend on context reinstatement. Here we build on work of Howard et al. (2006) by showing that although impaired contextual retrieval is critical, is not the full story—other mechanisms are necessary to account for the full aging pattern. As we discussed above, the notion that an item can be successfully recalled but fail to reinstate is associated contextual state is conceptually related to the distinction between familiarity and recollection Yonelinas (2002) and resonates with the suggestion that older adults are selectively impaired on recollection (Jacoby, 1999).

The next component needed to simulate older adults' performance was a lower threshold for accepting intrusions (the c_{thresh} parameter). In CMR2, intrusions are controlled (in part) during a post–retrieval editing phase in which the context representation associated with candidate retrievals is compared to the current context to ensure that retrieved items are associated with context similar to that which prevailed during list presentation. The higher the threshold on this similarity comparison, the less likely an intrusion will be made. Older adults had a lower threshold, indicating they were more likely to endorse intrusions (for behavioral evidence from externalized free recall tasks see Kahana et al., 2005). This finding is consistent with the spirit of the IDH (indeed, lowering the editing threshold was one way we implemented the IDH), and suggests older adults do have a specific deficit in the ability to reject intrusions of non–relevant items. The notion that older adults have difficulty monitoring the output of retrieval processes and editing out those that come from inappropriate sources has a long history in the study of aging (Hashtroudi, Johnson, & Chrosniak, 1989; McDonough & Gallo, 2013; K. J. Mitchell & Johnson, 2009). CMR2's mechanism of computing the similarity of reinstated context to the current state of context suggests a computational basis of impaired source monitoring. Indeed, the context reinstatement processes of CMR2 are conceptually similar to the idea that an item's source is defined by a variety of internal (e.g., emotions, thoughts) and external (e.g., spatial location, visual stimuli) features and that reactivation of these source features give a memory its episodic character (Johnson, Hashtroudi, & Lindsay, 1993; K. J. Mitchell & Johnson, 2009).

We used this same source monitoring mechanism to simulate recognition in the PEERS dataset, which showed no age differences in hit rate but increased false alarms for older adults. Specifically, when a probe item was presented, it was allowed to reinstate its associated context representation which was then compared to the current state of context. If the similarity was above a threshold (constrained to match the threshold used to simulate free recall) the probe was endorsed as a studied item, otherwise it was rejected as an intrusion. This lower threshold, together with reduced context drift following endorsing an item, allowed the four–component model to simultaneously account for age differences both on free recall and recognition. Critically, the model developed exclusively with free recall data was able to account for age differences on recognition without fitting any new free parameters. Because this model uses the same mechanism both to screen for intrusions in free recall and to distinguish studied from lure items in recognition, it makes the clear prediction that intrusion rates in free recall should predict false alarm rates in recognition. Confirming this prediction, we found a positive correlation between intrusion and false alarm rates.

In addition to difficulties sustaining attention, impaired context retrieval, and lower editing thresholds, we found that it is necessary to assume an additional deficit to fully account for the aging pattern. Specifically, older adults also tended to have higher values for the parameters controlling noise in the retrieval–phase decision accumulators (λ and η). A consequence of noisy retrieval competitions is that items that would ordinarily stand little chance of being recalled will occasionally accumulate enough spurious evidence in the form of noise to win the competition. Such noise–induced recalls will have multiple consequences. First, they reduce the probability that true list items will be recalled, lowering recall accuracy. Second, they increase the likelihood of distant temporal transitions and may contribute to older adults' reduced temporal contiguity. Finally, they will allow more non–list items to win retrieval competitions and be passed on to the post–retrieval editing phase which, in combination with a reduced editing threshold, paves the way for PLIs and ELIs. In essence, noisy retrieval competitions make it difficult to distinguish target memories (list items) from competitors. This difficulty may help explain recent findings that whereas young adults suppress competitors during the process of interference resolution (Healey et al., 2010; Healey, Ngo, & Hasher, 2014), older adults do not (Healey et al., 2013; Healey, Ngo, & Hasher, 2014). Effectively applying suppression may require a signal–to–noise ratio that is high enough to allow target memories to gain an initial advantage that can then be amplified by suppressing competitors; without that initial advantage it would be difficult for the memory system to "know" which items are likely to be competitors and therefore should be suppressed.

In summary, our multi–component model of age differences in free recall suggests that age deficits cannot be localized to a particular process or even to a particular stage of processing. Instead, older adults suffer deficits at multiple stages: attentional deficits during encoding, difficulty generating cues during retrieval, difficulty selecting among competitors during retrieval competition, and difficulty monitoring and editing responses. The need for multiple deficits is consistent the finding that aging has broad influences on the brain (Raz, 2005).

It is important to note that although our results suggest memory deficits cannot be attributed to a single deficient memory process, they leave open the possibility that a single underlying cause produces deficits in multiple processes. For example, some biological cause, such as accumulation of plaques or loss of white matter, may impair multiple brain networks, and thus multiple memory processes. Moreover, the rate and degree of change need not be the same in different networks, leaving room for divergent trajectories of impairments in different processes. Such a common cause may help explain the apparent tension between our finding that impairments in multiple processes are needed to capture the aging pattern and correlational studies that seem to suggest a single statistical factor accounts for age–related variation (Baltes & Lindenberger, 1997; Salthouse, 1996).

Relation to other paradigms and theories

For most of the paper we have focused on detailed measures of episodic memory derived from the dynamics of memory search in free recall and recognition. We began the paper, however, by discussing the broader empirical landscape of the aging and memory literature. We suggested that the broadest generalization that one can make about how memory changes with age is that the size of age deficits tracks the specificity of retrieval cues. Tasks that provide extremely strong cues, such as repetition priming in which the cue is the item itself show essentially no deficit (D. B. Mitchell & Bruss, 2003). As retrieval cues become less specific, age deficits increase, ranging from moderate on priming tasks with ambiguous cues (Ikier et al., 2008) to large on tasks like free recall that provide non-specific cues (Craik, 1968; Hultsch, 1969; Schonfield & Robertson, 1966).

Our results suggest a refinement and a caveat to this generalization. First, the refinement: the key dimension along which tasks vary may not be cue specificity but the extent to which they require the reinstatement of past contextual states. Simple tasks like priming do not require any reinstatement of context but simply rely on existing states of context. More complex tasks like free recall are critically dependent on reinstating contexts. Highly complex tasks such as reasoning may rely on the ability, not just to reinstate contextual states, but to control which states are reinstated and thereby provide access to task relevant memories (Healey, Crutchley, & Kahana, 2014). The caveat is that this generalization is a serious oversimplification in that it ignores the fact that differences in attention, evidence accumulation, and retrieval editing are also critical in determining whether a task will show age deficits and the exact nature of those deficits. Nonetheless the need for context reinstatement may prove to be a useful framework for thinking about age differences across many different types of tasks.

Next we provide a sketch of how the four–component model might be applied to age differences on a variety of tasks and discuss how our model relates to existing theoretical accounts. This sketch is not intended as a substitute for the sort of formal modeling we applied to the free recall and recognition tasks, but rather as a road map for future modeling work.

Priming—Repetition priming in tasks such as word naming or lexical decision involves a more rapid response to a stimuli the second time it is presented. The first time an item is presented it is likely to automatically activate its associated context representation as in the encoding phase of free recall. Therefore, when the item is presented a second time the context representation will be in a state that makes it easy to recognize the item. One way to think of this, is that in a naming task the context layer continually feeds activation to the feature layer via context–to–feature associations. Because the context layer tracks the history of recent events, the item representations of recently presented items will have more activation than the representations of items that were not presented recently. Therefore, when an item is presented a second time its item representation will already be partially active, reducing the amount of time required for it to fully activate and trigger a response. Note that this process does not require the reinstatement of any newly learned associations as in the retrieval phase of free recall. Moreover, priming effects are thought to arise, in part, from an automatic process (Balota, Black, & Cheney, 1992) and therefore older adults' difficulty with sustained attention would not be a detriment. Similarly, there is little need to edit out intrusions and the cues are so strong that evidence for the correct response is likely to accumulate rapidly even in the presence of noise.

The situation is more complex, however, with associative priming in which an initial item primes a related item. As discussed in the introduction, older adults do show deficits if a cue elicits several potential responses. For example, Ikier et al. (2008) had participants count the vowels in a list of words that included orthographically similar pairs.

When a word fragment (e.g., a _ l_ _ gy) resembled several words that had been seen in an earlier task but only one of which was a correct solution (e.g., ALLERGY and ANALOGY), older adults tended to show less priming for the correct solution than did younger adults. Several of the mechanisms in our four–component theory could contribute to such an effect. For example, younger adults may be able to reject the competitor because contextual retrieval allows them to determine that it is familiar because it appeared on an earlier task and not because it is a correct solution. For older adults, an inability to reinstate contextual states may make it difficult to determine that a candidate response was seen earlier in the task (we return to this idea when discussing source memory below). Noisy retrieval competitions may also contribute, increasing the chance that competitors will accumulate enough evidence to win. Finally, reduced editing thresholds would prevent older adults from editing such intrusions.

Implicit transfer—Implicit transfer tasks are those on which older adults use information from a previous, ostensibly irrelevant task, on the current task (for a review see Healey et al., 2008). For example, Biss et al. (2013) had participants study a list of words for free recall but gave them a surprise second recall test 15min after the first. During the 15min delay, participants completed an unrelated task in which some of the words from the free recall list were presented as distractors. Presenting list items as distractors had little influence on the performance of younger adults, but older adults showed better recall for the repeated words than the non–repeated words.

Findings such as this (for related examples see Campbell et al., 2010; Campbell, Trelle, & Hasher, 2014; Gopie, Craik, & Hasher, 2011; Thomas & Hasher, 2012) have been interpreted within the framework of the IDH as evidence that older adults fail to inhibit information from previous tasks. Although there is strong evidence that older adults do indeed have difficulty inhibiting irrelevant information during retrieval (Healey et al., 2013; Healey, Ngo, & Hasher, 2014), our four–component theory suggests that other factors may also contribute. It may be that when context is used to cue retrievals in these tasks, items from the prior task receive equal degrees of evidence for both younger and older adults, but because the evidence accumulation process is noisier for older adults than for younger adults, the prior–task items win the competition more often for older adults. Older adults' lower threshold for endorsing candidate retrievals would amplify this effect. Adjudicating between the inhibitory and noisy competition accounts would require detailed simulations.

Recognition—Our simulations show that the four–factor model is able to account for age differences in recognition accuracy using the same set of impaired mechanisms (and the same parameter values) used to simulate age differences in free recall. In addition to age differences in accuracy, recognition tasks show robust differences in reaction times. These differences have been modeled by Ratcliff, Thapar, and McKoon (2004) using the diffusion model. They found that the differences in reaction time were largely accounted for by differences in non–decision time rather than by differences in the rate of evidence accumulation. The lack of age differences in the rate of evidence accumulation during recognition may seem at odds with our finding of a noisy evidence accumulation process in free recall (note that although our model simulates evidence accumulation during free recall, we did not simulate evidence accumulation during *recognition*). But differences between recognition and free recall make it difficult to interpret this discrepancy. A possible explanation, for example, is that evidence accumulation in recognition is driven by the match between the probe and memory (the probe will tend to be a very strong and specific cue for a particular item) whereas in free recall it is driven between a match between a contextual state and memory (any given contextual state will provide a relatively weak and non-specific cue for many list, and non–list items). Consistent with this interpretation, Spaniol et al. (2006) did find lower diffusion model drift rates on a version of the recognition task that placed greater demands on episodic retrieval (e.g., longer retention intervals).

Associative Recognition—As we have reviewed above, Naveh-Benjamin and colleagues have extensively investigated tasks that require memory for associations among items or items and aspects of the presentation event (Old & Naveh-Benjamin, 2008). Older adults' difficulty with such associative recognition tasks has been attributed to difficulty forming or retrieving new associations (Naveh-Benjamin, 2000). We have shown that such a deficit is not sufficient to account for age differences in free recall, but can our four–component theory account for differences in associative recognition? Under the retrieved context framework, associations between items are not direct but mediated by mental context. For example, while studying the item pair *cat–table* both words would form associations to very similar states of mental context. Recognizing the pair at test could be modeled in much the same way we have modeled item recognition here: by allowing both words to retrieve their associated contexts and then computing the similarity between the two retrieved contexts, which is then be compared to some response threshold (the higher the similarity, the more likely the words were studied together). A reduced ability to reactivate past contextual states (i.e., the reduced β_{rec} parameter in our four–component model) would be expected to impair such a process, just as we have shown for item recognition.

Closely related to the associative deficit view is the notion of source memory: the ability to remember both that a particular item was seen and to remember specific details of the episode in which it occurred (Johnson et al., 1993; K. J. Mitchell & Johnson, 2009). There is considerable evidence that older adults are selectively impaired in the ability to recall the sources of memories (Spencer & Raz, 1995). Difficulty reinstating past contextual states may contribute to these source memory deficits.

Recall tasks—Age related–deficits on cued recall (paired associates) tasks tend to be larger than those on recognition but not as large as those for free recall. This pattern makes a great deal of sense in light of our four–component theory. Under the retrieved context framework, producing the associate of a probe item on a cued recall task would require allowing the probe item to reinstate its context and then using that context to cue recall of the associate. This need for context reinstatement suggests older adults would have difficulty with the task and will often fail to reinstate the required context. The consequence of a single failure to reinstate context is quite different in cued and free recall, however. In cued recall, if one probe fails to reinstate its context, the failure is unlikely to have much of an impact on the probability that the next probe will reinstate its context. By contrast, in free recall there are no probe items, rather the stream of recalls depends on each item reinstating a context that provides an effective cue for another list item. Therefore, a failure of a single item to reinstate its context could halt the entire recall process. The dependence of sequential recalls in free recall may be a large part of the reason it shows larger age deficits than cued recall.

The influence of emotionally valenced stimuli on the magnitude of age–related memory impairments has recently become a topic of much interest. There is evidence that age deficits are smaller for positively valenced material (Mather & Knight, 2005; Thomas & Hasher, 2006). It is possible that emotional stimuli capture attention (Mather & Knight, 2005), which helps to offset older adults' difficulty with sustained attention. It may also be that emotional stimuli introduce large shifts in mental context by injecting "more" of their context into the context layer than do non–emotional stimuli (i.e., emotional stimuli may temporarily increase the drift rate parameter). Such large context shifts may serve to make them particularly accessible: All else being equal if emotional stimuli cause their associated contextual states to be strongly represented in the context vector, the change will take a long time to dissipate (i.e., to be erased due to drift), and therefore many subsequent events will be linked to the emotional stimuli, providing many cues, essentially producing a encoding variability effect (Lohnas & Kahana, 2014b).

Moving away from laboratory tasks, there is evidence that older adults have difficulty retrieving the details of their autobiographical memories (St Jacques, Rubin, & Cabeza, 2012). When asked to recall specific life episodes, younger adults tend to provide many details. For example, if asked to recall a birthday party, they may report which people attended, what kind of cake they had, what presents they received. By contrast, older adults tend to provide "gist" level summaries that rely on semantic details rather than episodic details tied to a specific spatio–temporal context (Levine et al., 2002). Under the retrieved context framework, recalling specific details of an episode in response to a question about birthday parties would require using the question as a probe to reinstate a particular contextual state. If older adults have difficulty with such reinstatement, as our simulations suggest, one would expect that they would have impaired access to the details of the episode.

There are, of course, other perspectives on why older adults have difficulty with recall tasks. It has been suggested that one source of age differences on recall tasks may be that older and younger adults engage in different types of processing during encoding. For example, there is evidence that older adults are less likely than young adults to engage in deep, elaborative processing (Craik, 1977, 2002). Variations in the type of processing carried out during encoding are not directly simulated by CMR2. In future work it will be important to determine if age differences in the effect of processing type can be captured by the model. There are several ways variations in processing could be modeled in the retrieved context framework. Deep processing may take the form of allowing an item to activate contextual states related to its deep semantic meaning (e.g., allowing the item *allergy* to activate contextual states related to plants, pollen, the immune system, etc.) whereas shallower processing might involve activation limited to lower level orthographic representations (e.g., allowing the item *allergy* to activate contextual states related to words with similar spellings such as *analogy*). To the extent that attentional demands increase with depth of processing (Craik & Byrd, 1982), age differences in levels of processing may be captured by the attentional parameters (ϕ_s and ϕ_d) in our four–component model. Alternatively, older adult's tendency to engage in shallower processing might be due to a reduced ability to reinstate context—perhaps they are able to reinstate the "shallow" aspects of context but not the "deeper" aspects related to meaning.

It is also noteworthy that age differences can be minimized for items that are made particularly salient. For example Castel et al. (2002) assigned items arbitrary point values and asked participants to try to remember the high-value items; age differences were reduced for the highest value items. Similarly, May et al. (2005) found reduced age differences if tasks are framed in a way that emphasizes information of importance to older adult's (e.g., remembering which food is safe to eat and which is not). These data point to the need to model factors that modulate memory processes such as emotional valence and an individual's priorities.

An alternative to our context–based account of age differences on recall tasks suggests that older adults have relatively spared short–term memory, but impaired long–term memory processes. On such an account, the lack of age differences in recall initiation (i.e., the PFR curve) would be due to a preserved ability to retrieve the last few items of a list from short–term memory. This view is supported by the finding that age differences are generally small on simple span tasks (Craik 1977; but see Bopp and Verhaeghen 2005 for evidence that, although small, the differences are reliable). It is also consistent with studies that have found that age differences in the SPC are smaller for recency items than for mid–list and primacy items (for a review see, Craik & Jennings, 1992). Many other studies, however, show roughly equal age differences across the SPC (Capitani et al., 1992; Foos et al., 1987; Kahana et al., 2002; Parkinson et al., 1982; Poitrenaud et al., 1989; Rissenberg & Glanzer, 1987; Ward & Maylor, 2005). These inconsistencies are partially resolved by examining probability first recall curves, which provide a more direct measure of which items are most accessible as recall begins and generally reveal a clearly preserved recency effect among older adults (Kahana et al., 2002). We also note that many of the studies that have found intact recency effects that extend over several serial positions have instructed participants to recall the last few items first to minimize age differences in strategy (Craik & Jennings, 1992); this methodological difference likely accounts for some of the variation across studies.

Another likely source of inconsistent age differences in the SPC is the fact that most existing studies involve a single session with a small number of lists. In such cases, learning-to-learn effects can lead to fairly rapid changes in the size of the recency effect (Dallett, 1963; Goodwin, 1976; Hasher, 1973; Huang, 1986) and there may be age differences in how rapidly these effects occur. The new study reported here helps address this issue by having participants complete 7 sessions of free recall, thus providing ample time for participants to learn the task (Healey & Kahana, 2014). In these data, we do indeed see preserved recency effect, but only for the very last serial position (Figure 19A).

We argue that the balance of the evidence supports a context–based account of preserved recency for several reasons. First, the finding that preserved recency is limited to a single serial position is highly consistent with our four–component model's suggestion that older adults are able to use the state of context at the end of the list as a cue, and therefore can easily recall the final word, but have diffculty reinstating previous context states to guide further recalls, which predicts that age differences will be absent for the very first output by emerge quickly thereafter. This view also helps explain why the Kahana et al. (2002) data do not show preserved recency in the SPC. Comparing the PFR curves of the two studies (Figure 2B versus Figure 19B) shows that participants in Kahana et al. (2002) had elevated initiation probabilities for several items near the end of the list whereas participants in PEERS had elevated probabilities for only approximately 2 items. Therefore, the spared recency effect would be spread out over more serial positions in the SPC of Kahana et al. (2002) making it diffcult to detect. In contrast to this context–based account, a single item sparing of recency seems inconsistent with a dual–store account which would predict that older adults have easy access to several recent items in STM and therefore age differences should be absent from several of the final serial positions, and not just the final position.

A second reason to favor a context account over a preserved STM account is that that recency emerges at multiple time scales. For example, when a distracting task intervenes between presentation of a list and recall (i.e., delayed free recall), both younger and older adults tend to initiate recall with primacy items rather than recency items (Kahana et al., 2002). If primacy items are retrieved from LTM and older adults have a selective deficit in LTM, the prediction would seem to be that older adult's should show a deficit in initiating recall on such a task, but the do not (Kahana et al., 2002). By contrast, if, as our model suggests, older adults have an intact ability to use existing states of context to cue recall, there should be no age differences in recall initiation on either immediate or delayed free recall. Similarly, the Prior List Intrusions (PLI) recency effect (Figure 2E) in which intrusions tend to come from the most recent prior lists, shows no age differences. It is difficult to argue that this recency effect, which extends over many minutes and dozens of intervening items, reflects retention in short-term memory. Our model provides a natural account of recency effects at multiple timescales, as well as of the lack of age differences in these effects.

Finally, evidence from other tasks also challenges the view that short–term memory is spared with aging. For example, although age differences are generally small on simple span tasks, there are clear age differences on complex "working memory" span tasks (Bopp & Verhaeghen, 2005; May et al., 1999; Rowe et al., 2008; Salthouse, 1993).

Non-memory tasks—Older adults show deficits even on tasks that do not directly test memory (Baltes & Lindenberger, 1997; MacDonald et al., 2004; Park et al., 2002; Zelinski & Burnight, 1997). Although we certainly make no claim that our four–component theory provides a comprehensive account of all of these differences, it is worth considering a few examples of how it could be extended beyond memory tasks.

Tasks that measure processing speed have arguably been the most influential non-memory tasks in the cognitive aging literature. We suggest that noisy accumulation of evidence could lead to slowing on many different tasks. For example, string comparison tasks are often used to measure processing speed. In these tasks participants must compare two strings of numbers or letters and determine if the two are identical or not. Older adults are slower than younger adults to make such comparisons (Salthouse & Babcock, 1991). Their slow responding may be due to a general slowing of processing, or it may be due to a specific slowing of the evidence accumulation processes caused by increased noise. Because studies that measure processing speed tend not to include tasks specifically designed to measure the evidence accumulation process, it is conceivable that many of the findings that have been attributed to age differences in speed of processing are, on a computational level, due to age differences in selecting among competing response alternatives. Indirect support for this notion comes from the finding that older adults’ deficit in processing speed is reduced in comparison tasks that present a single item on screen at a time, rather than a full page of items as is more typical (Lustig et al., 2006). Although this finding has been interpreted as evidence that older adults have trouble inhibiting interference from items they are not currently working on, it may be that these items add noise to the accumulation process.

Processing speed tasks are quite simple. At the other end of the spectrum older adults show deficits on complex cognitive activities such as reasoning tasks. We have suggested that an ability to regulate the drift of the mental context representation so that it provides effective cues to memories relevant to the current task is a critical component in general intelligence. In support of this claim we have shown that, among young adults, those individuals who’s recall dynamics is biased toward the use of newly formed temporal associations (rather than long–standing semantic associations) show higher IQ scores than those who show a weaker influence of temporal associations (Healey, Crutchley, & Kahana, 2014). It may be that older adults’ reduced ability to reinstate contextual states at retrieval contributes both to their memory deficits and to impairments on complex cognitive tasks such as IQ tests.

A New Approach to Theory Development

The above is merely a sketch of how one could begin to apply the four–component theory beyond free recall and recognition; a first–approximation verbal fit of the theory to the data. In the spirit of the approach we have taken to analyzing the free recall and recognition tasks we want to stress that this sort of verbal theory fitting is insufficient. Much more work will be needed to rigorously assess the limits of the explanatory power of our theory. Indeed, rather than endorsing the specifics of the four–component model as a complete theory of aging, our main goal is to advocate for a new approach to theory development in the study of aging and memory. Specifically, theories of aging should be specified in sufficient detail that model–based quantitative predictions can be derived and validated against a large volume of data.

Model–based predictions are necessary because, as illustrated in our simulations, it can be quite difficult to predict how an impairment in a particular cognitive process will alter the behavior of the memory system. These difficulties are only compounded when the goal is to adjudicate between competing theories. An even more fundamental limitation in verbal theories lies in the flexibility of their predictions. We implemented the ADH, IDH, and CSH theories, each in three different ways. Each implementation has some free parameters, which grant the implementation a certain amount of flexibility; setting the parameters at different values allows the model to generate different patterns of predicted behavior. The more flexibility a theory has, the more likely it will be able to fit a new data point. But if a theory has too much flexibility—if it can contort itself to fit many different data patterns—it runs the risk of being unfalsifiable. As it turned out, the flexibility of our implementations was not enough to allow the theories to fit the older adults' data. But one could imagine many other ways to implement these theories, each with its own degree of flexibility. That is, because it admits of being specified and implemented in many different ways, a verbal theory has a tremendous amount of "latent flexibility". Because a model reflects one out of many ways a given verbal theory can be implemented, the model has, almost by necessity, less flexibility than the verbal theory that it implements. In our view, the latent flexibility of verbal theories is a hidden reason that the aging literature has not converged toward a common theoretical account.

But model–based predictions are not, by themselves, sufficient. Theories must also be evaluated against a large volume of data because, as our simulations show, even when translated into models, single effects were not sufficient to distinguish among theories. However, by considering a complex, multi–variate pattern we were able to show that none of the theories, not even the combination of the theories, can fully account for the pattern of age differences. To put it more starkly, if we had attempted to validate the theories at a verbal level against isolated effects, we would have concluded that each of the theories provided a viable account, but quantitatively validating them against multiple effects simultaneously, revealed that none of the theories were able to account for the data.

Therefore, we suggest that a profitable avenue for future work is to document and model the details of age differences on paradigms that have been foundational in the young adult episodic memory literature. This will include recognition, cued recall, free recall, and serial recall tasks. These tasks are ideal because detailed measures of task performance are well–documented with younger adults, allowing for easy age comparisons. Moreover, there are existing well–specified and validated models of these tasks (e.g., Farrell, 2012; Kahana & Sekuler, 2002; Murdock, 1993; Nosofsky et al., 2011; Ratcliff, Thapar, & McKoon, 2004; Shiffrin & Steyvers, 1997), which will admit quite easily of the same approach we used here—use the models as a healthy memory system and lesion various processes to see which are critical in producing age deficits. Researchers have already made much progress applying drift diffusion models to recognition data (Ratcliff, Thapar, & McKoon, 2004; Spaniol et al., 2006) and various forced alternative choice tasks like lexical decision and brightness discrimination (Ratcliff, Gomez, & McKoon, 2004; Starns & Ratcliff, 2010). Surprenant et al. (2006) have used the SIMPLE model to show the influence of stimuli similarity on age differences in serial recall. Modeling efforts should also focus on key tasks that have been used to test existing aging theories such as those used to assess associative deficits, inhibitory deficits, and speed deficits. Researchers working on the associative deficits (Li et al., 2005) and source memory deficits (Benjamin, 2010) have already begun pursuing this path.

A potential complication of our proposal is that different researchers will use different models. For example, we have used the CMR2 model, Ratcliff, Thapar, and McKoon (2004) and Spaniol et al. (2006) have used the diffusion model, Surprenant et al. (2006) have used the SIMPLE model, and both Benjamin (2010) and Li et al. (2005) have used novel models specifically designed to address age differences. The use of so many different models creates the danger that the landscape of this emerging modeling literature will become fragmented. It will therefore be critical to directly pit model–based theories against each other within a common framework, much as we have done here for verbal theories.

Our model-based approach also holds the promise of improving the assessment, and eventually treatment, of memory disorders. By fitting a model to the free recall performance of individuals it may be possible to identify impairments that are common to most older adults, impairments that characterize subgroups of older adults, and impairments that are specific to individuals. Such individualized profiles of cognitive aging may allow for the development of measures to detect early signs of decline, to distinguish healthy aging from profiles that predict transition to Alzheimer’s disease and other forms of dementia, and to tailor interventions to the needs of individual older adults.

Conclusions

Theory development in cognitive aging has reached a level of sophistication at which it is no longer possible to fully evaluate a given theory against univariate measures: a theory that accounts for multiple single effects, when considered in isolation, may fail to account for the same effects when considered simultaneously. Instead theories must be evaluated against the details of multivariate datasets such as the ones presented here. Such evaluations will require implementing aging theories within computationally explicit models of healthy memory.

Using this model–based approach, we developed a four–component model of age differences on the free recall and recognition tasks. By insisting that candidate theories account for multiple data points *simultaneously* we were able to quickly eliminate theories that cannot account for the data and develop a novel theory that implicates four components: 1) the ability to sustain attention across an encoding episode, 2) the ability to retrieve contextual representations for use as retrieval cues, 3) the ability to monitor retrievals and reject intrusions, and 4) the level of noise in evidence accumulation during retrieval competitions. This approach can be extended to develop a model of age differences on a broad range of tasks. A tight coupling between experimentation and model–based theory evaluation promises rapid progress in building a comprehensive theory of the aging memory system.

Acknowledgments

This research was funded by National Institutes of Health Grant MH55687 and National Science Foundation Grant NSF1058886. We thank Jonathan Miller for assistance with designing and programming the experiment and Kylie Hower, Joel Kuhn, and Elizabeth Crutchley for help with data collection. We thank Patrick Crutchley, Lynn Lohnas, Nicole Long, Logan O'Sullivan, Youssef Ezzyat, for helpful comments on this manuscript.

References

Anderson JR, Bower GH. Recognition and retrieval processes in free recall. Psycholgical Review. 1972; 79:97–123. DOI: 10.1037/h0033773

Anderson ND, Craik FIM, Naveh-Benjamin M. The attentional demands of encoding and retrieval in younger and older adults: I. evidence from divided attention costs. Psychology and Aging. 1998; 13:405–423. [PubMed: 9793117]

Atkinson, RC.; Shiffrin, RM. Human memory: A proposed system and its control processes. In: Spence, KW.; Spence, JT., editors. The psychology of learning and motivation. Vol. 2. New York: Academic Press; 1968. p. 89-105.

Balota DA, Black SR, Cheney M. Automatic and attentional priming in young and older adults: Reevaluation of the two-process model. Journal of Experimental Psychology: Human Perception and Performance. 1992; 18:485–502. [PubMed: 1593231]

Balota DA, Duchek JM, Paullin R. Age-related differences in the impact of spacing, lag, and retention interval. Psychology and Aging. 1989; 4:3–9. [PubMed: 2803609]

Baltes PB, Lindenberger U. Emergence of a powerful connection between sensory and cognitive functions across the adult life span: a new window to the study of cognitive aging? Psychology and Aging. 1997; 12:12–21. [PubMed: 9100264]

Baltes PB, Staudinger UM. Wisdom: A metaheuristic (pragmatic) to orchestrate mind and virtue toward excellence. The American psychologist. 2000; 55:122–136. [PubMed: 11392856]

Bender A, Naveh-Benjamin M, Raz N. Associative deficit in recognition memory in a lifespan sample of healthy adults. Psychology and Aging. 2010; 25:940–948. [PubMed: 20822256]

Benjamin A. Representational explanations of "process" dissociations in recognition: The dryad theory of aging and memory judgments. Psychological Review. 2010; 117:1055–1079. [PubMed: 20822289]

Biss RK, Ngo KWJ, Hasher L, Campbell KL, Rowe G. Distraction can reduce age-related forgetting. Psychological Science. 2013; 24:448–455. [PubMed: 23426890]

Bopp KL, Verhaeghen P. Aging and verbal memory span: A meta-analysis. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences. 2005; 60B:P223–P233.

Bousfield WA. The occurrence of clustering in the recall of randomly arranged associates. Journal of General Psychology. 1953; 49:229–240.

Bower, GH. A multicomponent theory of the memory trace. In: Spence, KW.; Spence, JT., editors. The psychology of learning and motivation : Advances in research and theory. Vol. 1. New York: Academic Press; 1967. p. 229-325.

Brodie DA, Murdock BB. Effects of presentation time on nominal and functional serial position curves in free recall. Journal of Verbal Learning and Verbal Behavior. 1977; 16:185–200. DOI: 10.1016/S0022-5371(77)80046-7

Campbell KL, Hasher L, Thomas RC. Hyper-binding: A unique age effect. Psychological Science. 2010; 21:399–405. [PubMed: 20424077]

Campbell KL, Trelle A, Hasher L. Hyper-binding across time: Age differences in the effect of temporal proximity on paired-associate learning. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2014; 40:293–299.

Campbell KL, Zimerman S, Healey MK, Lee MMS, Hasher L. Age differences in visual statistical learning. Psychology and Aging. 2012; 27:650–656. [PubMed: 22251380]

Capitani E, Della Sala S, Logie RH, Spinnler H. Recency, primacy, and memory: Reappraising and standardising the serial position curve. Cortex. 1992; 28:315–342. [PubMed: 1395637]

Carstensen LL, Isaacowitz DM, Charles ST. Taking time seriously. A theory of socioemotional selectivity. The American psychologist. 1999 Mar.54:165–181. [PubMed: 10199217]

Castel AD. Memory for grocery prices in younger and older adults: The role of schematic support. Psychology and Aging. 2005; 20:718–721. [PubMed: 16420146]

Castel AD, Benjamin AS, Craik FIM, Watkins MJ. The effects of aging on selectivity and control in short-term recall. Memory and Cognition. 2002; 30:1078–1085. [PubMed: 12507372]

Ceci SJ, Tabor L. Flexibility and memory: Are the elderly really less flexible? Experimental Aging Research. 1981; 7:147–158. [PubMed: 7274321]

Chalfonte BL, Johnson MK. Feature memory and binding in young and older adults. Memory & Cognition. 1996; 24:403–416. [PubMed: 8757490]

Copeland DE, Radvansky GA. Aging and integrating spatial mental models. Psychology and Aging. 2007; 22:569–579. [PubMed: 17874955]

Craik FIM. Two components in free recall. Journal of Verbal Learning and Verbal Behavior. 1968; 7:996–1004.

Craik FIM. Age differences in recognition memory. The Quarterly Journal of Experimental Psychology. 1971; 23:316–323.

Craik, FIM. Age differences in human memory. In: Birren, JE.; Schaie, KW., editors. Handbook of the psychology of aging. New York: Van Nostrand Reinhold; 1977. p. 384-420.

Craik FIM. On the transfer of information from temporary to permanent memory. Philosophical Transactions of the Royal Society B: Biological Sciences. 1983; 302:341–359.

Craik FIM. Levels of processing: Past, present... and future? Memory. 2002 Sep.10:305–318. [PubMed: 12396643]

Craik, FIM.; Byrd, M. Aging and cognitive deficits: The role of attentional resources. In: Craik, FIM.; Trehub, S., editors. Aging & cognitive processes: Advances in the study of communication and affect. Vol. 8. New York: Plenum Press; 1982. p. 191-211.

Craik, FIM.; Jennings, JM. Human memory. In: Craik, FIM.; Salthouse, TA., editors. Handbook of aging and cognition. Hillsdale, NJ: Earlbaum; 1992. p. 586

Craik FIM, Luo L, Sakuta Y. Effects of aging and divided attention on memory for items and their contexts. Psychology and Aging. 2010; 25:968–979. [PubMed: 20973605]

Craik FIM, Rose NS. Memory encoding and aging: A neurocognitive perspective. Neuroscience and Biobehavioral Reviews. 2012; 36:1729–1739. [PubMed: 22155274]

Dallett KM. Practice effects in free and ordered recall. J Exp Psychol. 1963; 66:65–71. [PubMed: 14024751]

Daneman M, Carpenter PA. Individual difference in working memory and reading. Journal of Verbal Learning and Verbal Behavior. 1980; 19:450–466.

Darowski ES, Helder E, Zacks RT, Hasher L, Hambrick DZ. Age-related differences in cognition: The role of distraction control. Neuropsychology. 2008 Jan.22:638–644. [PubMed: 18763883]

Davelaar EJ, Goshen-Gottstein Y, Ashkenazi A, Haarmann HJ, Usher M. The demise of short-term memory revisited: Empirical and computational investigations of recency effects. Psychological Review. 2005; 112:3–42. DOI: 10.1037/0033-295X.112.1.3 [PubMed: 15631586]

Dennis S, Humphreys MS. A context noise model of episodic word recognition. Psychological Review. 2001; 108:452–478. [PubMed: 11381837]

Dunlosky J, Hertzog C. Aging and deficits in associative memory: what is the role of strategy production? Psychology and Aging. 1998 Dec.13:597–607. [PubMed: 9883460]

Estes WK. Statistical theory of distributional phenomena in learning. Psychological Review. 1955; 62:369–377. DOI: 10.1037/h0046888 [PubMed: 13254976]

Farrell S. Temporal clustering and sequencing in short-term memory and episodic memory. Psychological Review. 2012; 119:223–271. DOI: 10.1037/a0027371 [PubMed: 22506678]

Foos PW, Sabol MA, Corral G, Mobley L. Age differences in primary and secondary memory. Bulletin of the psychonomic society. 1987; 25:159–160.

Gallistel, C. Dead reckoning, cognitive maps, animal navigation and the representation of space: An introduction. Springer; 2008.

Gazzaley A, Clapp W, Kelley J, McEvoy K, Knight R, D'Esposito M. Age-related top-down suppression deficit in the early stages of cortical visual memory processing. Proceedings of the National Academy of Sciences, USA. 2008; 105:13122.

Goodwin J. Changes in primacy and recency with practice in single-trial free recall. Journal of Verbal Learning & Verbal Behavior. 1976; 15:119–132.

Gopie N, Craik F, Hasher L. A double dissociation of implicit and explicit memory in younger and older adults. Psychological Science. 2011; 22:634–640. [PubMed: 21421935]

Grenfell-Essam R, Ward G. Examining the relationship between free recall and immediate serial recall: The role of list length, strategy use, and test expectancy. Journal Of Memory And Language. 2012; 67:106–148.

Hamm VP, Hasher L. Age and the availability of inferences. Psychology and Aging. 1992; 7:56–64. [PubMed: 1558706]

Hasher L. Position effects in free recall. American Journal Of Psychology. 1973 Jan.86:389–397.

Hasher, L.; Lustig, C.; Zacks, RT. Inhibitory mechanisms and the control of attention. In: Conway, A.; Jarrold, C.; Kane, M.; Miyake, A.; Towse, J., editors. Variation in working memory. New York: Oxford University Press; 2007. p. 227-249.

Hasher, L.; Zacks, RT. Working memory, comprehension, and aging: A review and a new view. In: Bower, GH., editor. The psychology of learning and motivation: Advances in research and theory. San Diego: Academic Press; 1988. p. 193-225.

Hasher, L.; Zacks, RT.; May, CP. Inhibitory control, circadian arousal, and age. In: GA; Koriat, A., editors. Attention and performance xvii, cognitive regulation of performance: Interaction of theory and application. Vol. 17. Cambridge, MA: MIT Press; 1999. p. 653-675.

Hashtroudi S, Johnson MK, Chrosniak LD. Aging and source monitoring. Psychology and Aging. 1989; 4:106–112. [PubMed: 2803603]

Healey, MK. Interference. In: WSK, editor. The encyclopedia of adulthood and aging. New York: Wiley-Blackwell; in press

Healey, MK.; Campbell, KL.; Hasher, L. Cognitive aging and increased distractibility: Costs and potential benefits. In: Sossin, WS.; Lacaille, JC.; Castellucci, VF.; Belleville, S., editors. Progress in brain research. Vol. 169. Elsevier; Amsterdam: 2008. p. 353-363.

Healey MK, Campbell KL, Hasher L, Ossher L. Direct evidence for the role of inhibition in resolving interference in memory. Psychological science. 2010; 21:1464–1470. [PubMed: 20807896]

Healey MK, Crutchley P, Kahana MJ. Individual differences in memory search and their relation to intelligence. Journal of Experimental Psychology: General. 2014; 143:1553–1569. DOI: 10.1037/a0036306 [PubMed: 24730719]

Healey MK, Hasher L, Campbell KL. The role of suppression in resolving interference: Evidence for an age-related deficit. Psychology and Aging. 2013; 28:721–728. [PubMed: 23957222]

Healey MK, Kahana MJ. Is memory search governed by universal principles or idiosyncratic strategies? Journal of Experimental Psychology: General. 2014; 143:575–596. DOI: 10.1037/a0033715 [PubMed: 23957279]

Healey MK, Ngo KWJ, Hasher L. Below–baseline suppression of competitors during interference resolution by younger but not older adults. Psychological Science. 2014; 25:145–151. [PubMed: 24214245]

Hertzog C, Dixon RA, Hultsch DF, MacDonald SWS. Latent change models of adult cognition: are changes in processing speed and working memory associated with changes in episodic memory? Psychology and Aging. 2003; 18:755–769. [PubMed: 14692862]

Hertzog C, Dunlosky J. Metacognition in later adulthood: Spared monitoring can benefit older adults' self-regulation. Current Directions in Psychological Science. 2011 May.20:167–173. [PubMed: 24478539]

Hogan RM. Interitem encoding and directed search in free recall. Memory & Cognition. 1975; 3:197–209. [PubMed: 21287060]

Howard, MW.; Addis, KA.; Jing, B.; Kahana, MJ. Semantic structure and episodic memory. In: Landauer, TK.; McNamara, DS.; Dennis, S.; Kintsch, W., editors. Handbook of latent semantic analysis. Mahwah, NJ: Laurence Erlbaum and Associates; 2007. p. 121-141.

Howard MW, Kahana MJ. Contextual variability and serial position effects in free recall. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1999; 25:923–941. DOI: 10.1037/0278-7393.25.4.923

Howard MW, Kahana MJ. A distributed representation of temporal context. Journal of Mathematical Psychology. 2002a; 46:269–299. DOI: 10.1006/jmps.2001.1388

Howard MW, Kahana MJ. When does semantic similarity help episodic retrieval? Journal of Memory and Language. 2002b; 46:85–98.

Howard MW, Kahana MJ, Wingfield A. Aging and contextual binding: Modeling recency and lag-recency effects with the temporal context model. Psychonomic Bulletin & Review. 2006; 13:439–445. DOI: 10.3758/BF03193867 [PubMed: 17048728]

Huang I. Transitory changes of primacy and recency in successive single-trial free recall. Journal of General Psychology. 1986; 113:5–21.

Hultsch DF. Adult age differences in the organization of free recall. Developmental psychology. 1969; 1:673–678.

Ikier S, Yang L, Hasher L. Implicit proactive interference, age, and automatic versus controlled retrieval strategies. Psychological Science. 2008; 19:456–461. [PubMed: 18466406]

Jacoby LL. Ironic effects of repetition: Measuring age-related differences in memory. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1999; 25:3–22.

Jacoby LL, Bishara AJ, Hessels S, Toth JP. Aging, subjective experience, and cognitive control: dramatic false remembering by older adults. Journal of Experimental Psychology: General. 2005 May.134:131–148. [PubMed: 15869342]

Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. Psychological Bulletin. 1993; 114:3–28. [PubMed: 8346328]

Kahana MJ. Associative retrieval processes in free recall. Memory & Cognition. 1996; 24:103–109. DOI: 10.3758/BF03197276 [PubMed: 8822162]

Kahana MJ, Dolan ED, Sauder CL, Wingfield A. Intrusions in episodic recall: Age differences in editing of overt responses. Journal of Gerontology: Psychological Sciences. 2005; 60:92–97. DOI: 10.1093/geronb/60.2.P92

Kahana MJ, Howard MW, Zaromb F, Wingfield A. Age dissociates recency and lag recency effects in free recall. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2002; 28:530–540. DOI: 10.1037/0278-7393.28.3.530

Kahana MJ, Sekuler R. Recognizing spatial patterns: A noisy exemplar approach. Vision Research. 2002; 42:2177–2192. [PubMed: 12207978]

Kahana MJ, Zhou F, Geller AS, Sekuler R. Lure-similarity affects visual episodic recognition: Detailed tests of a noisy examplar model. Memory & Cognition. 2007; 35:1222–1232. [PubMed: 18035623]

Katzman R, Brown T, Fuld P, Peck A, Schechter R, Schimmel H. Validation of a short orientation-memory-concentration test of cognitive impairment. The American journal of psychiatry. 1983; 140:734–739. [PubMed: 6846631]

Kausler, DH. Learning and memory in normal aging. San Diego, CA: Academic Press; 1994.

Keppel, G. Verbal behavior and general behavior theory. Dixon, TR.; Horton, DL., editors. Prentice-Hall; 1968. p. 172-213.

Kilb A, Naveh-Benjamin M. The effects of pure pair repetition on younger and older adults' associative memory. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2011; 37:706–719.

Kramer AF, Humphrey DG, Larish JF, Logan GD, Strayer DL. Aging and inhibition: beyond a unitary view of inhibitory processing in attention. Psychology and Aging. 1994; 9:491–512. [PubMed: 7893421]

Laming D. Testing the idea of distinct storage mechanisms in memory. International Journal of Psychology. 1999; 34:419–426.

Laming D. An improved algorithm for predicting free recalls. Cognitive Psychology. 2008; 57:179–219. DOI: 10.1016/j.cogpsych.2008.01.001 [PubMed: 18329010]

Laming DL. Predicting free recalls. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2006; 32:1146–1163.

Landauer TK, Dumais ST. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review. 1997; 104:211–240.

Laurence M. Memory loss with age: A test of two strategies for its retardation. Psychonomic Society. 1967; 9:209–210.

Levine B, Svoboda E, Hay JF, Winocur G, Moscovitch M. Aging and autobiographical memory: Dissociating episodic from semantic retrieval. Psychology and Aging. 2002; 17:677–689. [PubMed: 12507363]

Li SC, Naveh-Benjamin M, Lindenberger U. Aging neuromodulation impairs associative binding: A neurocomputational account. Psychological Science. 2005; 16:445–450. [PubMed: 15943670]

Light L. Memory and again: Four hypotheses in search of data. Annual Review of Psychology. 1991; 42:333–376.

Lindenberger U, Ghisletta P. Cognitive and sensory declines in old age: Gauging the evidence for a common cause. Psychology and Aging. 2009; 24:1–16. [PubMed: 19290733]

Lohnas LJ, Kahana MJ. Parametric effects of word frequency effect in memory for mixed frequency lists. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2013; 39:1943–1946. DOI: 10.1037/a0033669

Lohnas LJ, Kahana MJ. Compound cuing in free recall. Journal of Experimental Psychology: Learning, Memory and Cogntion. 2014a; 40:12–24. DOI: 10.1037/a0033698

Lohnas LJ, Kahana MJ. A retrieved context account of spacing and repetition effects in free recall. Journal of Experimental Psychology: Learning Memory and Cognition. 2014b; 40:755–764. DOI: 10.1037/a0035585

Lohnas LJ, Polyn SM, Kahana MJ. Expanding the scope of memory search: Intralist and interlist effects in free recall. Psychological Review. 2015; 122:337–363. [PubMed: 25844876]

Loonstra AS, Tarlow AR, Sellers AH. COWAT metanorms across age, education, and gender. Applied Neuropsychology. 2001; 8:161–166. [PubMed: 11686651]

Lustig C, Hasher L, Tonev ST. Distraction as a determinant of processing speed. Psychonomic Bulletin and Review. 2006; 13:619–625. [PubMed: 17201361]

MacDonald SWS, Dixon RA, Cohen AL, Hazlitt JE. Biological age and 12-year cognitive change in older adults: findings from the victoria longitudinal study. Gerontology. 2004; 50:64–81. [PubMed: 14963373]

Marshall PH, Werder PR. The effects of the elimination of rehearsal on primacy and recency. Journal of Verbal Learning and Verbal Behavior. 1972; 11:649–653.

Mather M, Knight M. Goal-directed memory: The role of cognitive control in older adults' emotional memory. Psychology and Aging. 2005; 20:554–570. [PubMed: 16420131]

May CP, Hasher L. Synchrony effects in inhibitory control over thought and action. Journal of Experimental Psychology: Human Perception and Performance. 1998; 24:363–379. [PubMed: 9554091]

May CP, Hasher L, Kane MJ. The role of interference in memory span. Memory & Cognition. 1999; 27:759–767. [PubMed: 10540805]

May CP, Rahhal T, Berry EM, Leighton EA. Aging, source memory, and emotion. Psychology and aging. 2005; 20:571. [PubMed: 16420132]

McDonough IM, Gallo DA. Impaired retrieval monitoring for past and future autobiographical events in older adults. Psychology and Aging. 2013; 28:457–466. [PubMed: 23795764]

McGeoch JA. Forgetting and the law of disuse. Psychological Review. 1932; 39:352–70. DOI: 10.1037/h0069819

Mensink GJM, Raaijmakers JGW. A model for interference and forgetting. Psychological Review. 1988; 95:434–455. DOI: 10.1037/0033-295X.95.4.434

Merkow MB, Burke JF, Kahana MJ. The human hippocampus contributes to both recollection and familiarity components of recognition memory. Submitted. Submitted.

Miller JF, Kahana MJ, Weidemann CT. Recall termination in free recall. Memory & Cognition. 2012; 40:540–550. DOI: 10.3758/s13421-011-0178-9

Mitchell DB, Bruss PJ. Age differences in implicit memory: Conceptual, perceptual, or methodological? Psychology and Aging. 2003; 18:807–822. [PubMed: 14692866]

Mitchell KJ, Johnson MK. Source monitoring 15 years later: what have we learned from fmri about the neural mechanisms of source memory? Psychological bulletin. 2009; 135:638. [PubMed: 19586165]

Modigliani V, Hedges DG. Distributed rehearsals and the primacy effect in single-trialfree recall. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1987; 13:426–436.

Mogle JA, Lovett BJ, Stawski RS, Sliwinski MJ. What's so special about working memory? An examination of the relationships among working memory, secondary memory, and fluid intelligence. Psychological Science. 2008; 19:1071–1077. [PubMed: 19076475]

Moscovitch, M.; Winocur, G. The neuropsychology of memory and aging. In: Craik, FIM.; Salthouse, TA., editors. Handbook of aging and cognition. Hillsdale, NJ: Earlbaum; 1992. p. 315-372.

Murdock BB. TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. Psychological Review. 1993; 100:183–203. [PubMed: 8483981]

Murdock BB. Context and mediators in a theory of distributed associative memory (TODAM2). Psychological Review. 1997; 104:839–862. DOI: 10.1037/0033-295X.104.4.839

Naveh-Benjamin M. Adult-age differences in memory performance: Tests of an associative deficit hypothesis. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2000; 26:1170–1187.

Naveh-Benjamin M, Brav TK, Levy O. The associative memory deficit of older adults: The role of strategy utilization. Psychology and Aging. 2007; 22:202–208. [PubMed: 17385995]

Naveh-Benjamin M, Guez J, Kilb A, Reedy S. The associative memory deficit of older adults: further support using face-name associations. Psychology and Aging. 2004 Sep.19:541–546. [PubMed: 15383004]

Neath I, Crowder RG. Schedules of presentation and temporal distinctiveness in human memory. Journal of Experimental Psychology: Learning, Memory and Cognition. 1990; 16:316–327.

Newson RS, Kemps EB. The nature of subjective cognitive complaints of older adults. International Journal of Aging Human Development. 2006; 63:139–151. [PubMed: 17137031]

Nigg JT. On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. Psychological Bulletin. 2000; 127:220–246.

Nosofsky RM, Little DR, Donkin C, Fific M. Short-term memory scanning viewed as exemplar-based categorization. Psychological Review. 2011; 118:280–315. [PubMed: 21355662]

Old S, Naveh-Benjamin M. Differential effects of age on item and associative measures of memory: A meta-analysis. Psychology and Aging. 2008; 23:104–118. [PubMed: 18361660]

Overman AA, Becker JT. The associative deficit in older adult memory: Recognition of pairs is not improved by repetition. Psychology and Aging. 2009; 24:501–506. [PubMed: 19485666]

Park DC, Lautenschlager G, Hedden T, Davidson NS, Smith AD, Smith PK. Models of visuospatial and verbal memory across the adult life span. Psychology and Aging. 2002; 17:299–320. [PubMed: 12061414]

Park DC, Smith AD, Lautenschlager G, Earles JL, Frieske D, Zwahr M, Gaines CL. Mediators of long-term memory performance across the life span. Psychology and Aging. 1996; 11:621–637. [PubMed: 9000294]

Parkinson SR, Lindholm JM, Inman VW. An analysis of age differences in immediate recall. Journal of Gerontology. 1982; 37:425–431. [PubMed: 7086078]

Pichora-Fuller MK, Schneider BA, Daneman M. How young and old adults listen to and remember speech in noise. The Journal of the Acoustical Society of America. 1995; 97:593–608. [PubMed: 7860836]

Poitrenaud J, Moy F, Girousse A, Wolmark Y, Piette F. Psychometric procedures for analysis of memory losses in the elderly. Archives of Gerontology and Geriatrics. 1989; 1:173–183. [PubMed: 2757730]

Polyn SM, Norman KA, Kahana MJ. A context maintenance and retrieval model of organizational processes in free recall. Psychological Review. 2009; 116:129–156. DOI: 10.1037/a0014420 [PubMed: 19159151]

Provyn JP, Sliwinski MJ, Howard MW. Effects of age on contextually mediated associations in paired associate learning. Psychology and Aging. 2007; 22:846–857. [PubMed: 18179302]

Radvansky G, Zacks R, Hasher L. Age and inhibition: The retrieval of situation models. Journals Of Gerontology. 2005; 60B:P276–P278.

Ratcliff R, Gomez P, McKoon G. A diffusion model account of the lexical decision task. Psychological Review. 2004; 111:159–182. [PubMed: 14756592]

Ratcliff R, Thapar A, McKoon G. A diffusion model analysis of the effects of aging on recognition memory. Journal of Memory and Language. 2004; 50:408–424.

Raymond BJ. Free recall among the aged. Psychological Reports. 1971; 29:1179–1182. [PubMed: 5139357]

Raz N. Regional brain changes in aging healthy adults: General trends, individual differences and modifiers. Cerebral Cortex. 2005; 15:1676–1689. [PubMed: 15703252]

Rissenberg M, Glanzer M. Free-recall and word finding ability in normal aging and senile dementia of the Alzheimer's type- the effect of item concreteness. Journal of Gerontology. 1987; 42:318–322. [PubMed: 3571868]

Romney AK, Brewer DD, Batchelder WH. Predicting clustering from semantic structure. Psychological Science. 1993; 4:28–34.

Rowe G, Hasher L, Turcotte J. Age differences in visuospatial working memory. Psychology and Aging. 2008; 23:79–84. [PubMed: 18361657]

Rundus D. Maintenance rehearsal and long-term recency. Memory & Cognition. 1980; 8:226–230. [PubMed: 7392949]

Sahakyan L, Kelley CM. A contextual change account of the directed forgetting effect. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2002; 28:1064–1072. DOI: 10.1037/0278-7393.28.6.1064

Salthouse, TA. Theoretical perspectives on cognitive aging. Hillsdale, NJ: Lawrence Erlbaum and Associates; 1991.

Salthouse TA. Influence of working memory on adult age differences in matrix reasoning. British Journal Of Psychology. 1993; 84:171–199. [PubMed: 8319054]

Salthouse TA. The processing-speed theory of adult age differences in cognition. Psychological Review. 1996; 103:403–428. [PubMed: 8759042]

Salthouse TA, Babcock RL. Decomposing adult age differences in working memory. Developmental psychology. 1991; 27:763–776.

Schonfield D, Robertson BA. Memory storage and aging. Canadian Journal of Psychology. 1966; 20:228–236. [PubMed: 5942327]

Schwarz G. Estimating the dimension of a model. Annals of Statistics (6). 1978; (2):461–464.

Sederberg PB, Gauthier LV, Terushkin V, Miller JF, Barnathan JA, Kahana MJ. Oscillatory correlates of the primacy effect in episodic memory. NeuroImage. 2006; 32:1422–1431. DOI: 10.1016/j.neuroimage.2006.04.223 [PubMed: 16814568]

Sederberg PB, Howard MW, Kahana MJ. A context-based theory of recency and contiguity in free recall. Psychological Review. 2008; 115:893–912. DOI: 10.1037/a0013396 [PubMed: 18954208]

Sederberg PB, Miller JF, Howard WH, Kahana MJ. The temporal contiguity effect predicts episodic memory performance. Memory & Cognition. 2010; 38:689–699. DOI: 10.3758/MC.38.6.689 [PubMed: 20852233]

Serruya MD, Sederberg PB, Kahana MJ. Power shifts track serial position and modulate encoding in human episodic memory. Cerebral Cortex. 2014; 24:403–413. DOI: 10.1093/cercor/bhs318 [PubMed: 23081881]

Shiffrin RM, Steyvers M. A model for recognition memory: REM—retrieving effectively from memory. Psychonomic Bulletin and Review. 1997; 4:145.doi: 10.3758/BF03209391 [PubMed: 21331823]

Shipley, WC. Shipley institute of living scale. Los Angeles, CA: Western Psychological Services; 1946.

Smith AD. Adult age differences in cued recall. Developmental psychology. 1977; 13:326–331.

Spaniol J, Madden DJ, Voss A. A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. Journal Experimental Psychology: Learning, Memory and Cogntion. 2006; 32:101–117.

Spencer WD, Raz N. Differential effects of aging on memory for content and context: A meta-analysis. Psychology and Aging. 1995; 10:527–539. [PubMed: 8749580]

Spillers GJ, Unsworth N. Variation in working memory capacity and temporal–contextual retrieval from episodic memory. Journal Experimental Psychology: Learning, Memory and Cogntion. 2011; 37:1532–1539.

St Jacques PL, Rubin DC, Cabeza R. Age-related effects on the neural correlates of autobiographical memory retrieval. Neurobiology of Aging. 2012; 33:1298–1310. [PubMed: 21190759]

Stark SM, Yassa MA, Stark CEL. Individual differences in spatial pattern separation performance associated with healthy aging in humans. Learning & Memory. 2010; 17:284–288. [PubMed: 20495062]

Starns J, Ratcliff R. The effects of aging on the speed-accuracy compromise: Boundary optimality in the diffusion model. Psychology and Aging. 2010; 25:377–390. [PubMed: 20545422]

Steyvers, M.; Shiffrin, RM.; Nelson, DL. Word association spaces for predicting semantic similarity effects in episodic memory. In: Healy, AF., editor. Cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer. Washington, DC: American Psychological Association; 2004.

Surprenant AM, Neath I, Brown GDA. Modeling age-related differences in immediate memory using SIMPLE. Journal Of Memory And Language. 2006; 55:572–586. [PubMed: 18172514]

Tan L, Ward G. A recency-based account of the primacy effect in free recall. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2000; 26:1589–1626. DOI: 10.1037/0278-7393.26.6.1589

Thomas RC, Hasher L. The influence of emotional valence on age differences in early processing and memory. Psychology and Aging. 2006; 21:821–825. [PubMed: 17201502]

Thomas RC, Hasher L. Reflections of distraction in memory: Transfer of previous distraction improves recall in younger and older adults. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2012; 38:30–39.

Troyer AK, Moscovitch M, Winocur G. Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. Neuropsychology. 1997; 11:138–146. [PubMed: 9055277]

Tulving, E. Episodic and semantic memory. In: Tulving, E.; Donaldson, W., editors. Organization of memory. New York: Academic Press; 1972. p. 381-403.

Tulving, E.; Rosenbaum, RS. Distinctiveness and memory. Hunt, RR.; Worthen, JB., editors. New York, NY: Oxford University Press; 2006. p. 407-423.

Turner ML, Engle RW. Is working memory capacity task dependent? Journal Of Memory And Language. 1989; 28:127–154.

Underwood BJ. The effect of successive interpolations and proactive inhibition. Psychological Monographs. 1945; 59:v + 33.doi: 10.1037/h0093547

Unsworth N. Variation in working memory capacity, fluid intelligence, and episodic recall: A latent variable examination of differences in the dynamics of free recall. Memory and Cognition. 2009; 37:837–849. [PubMed: 19679863]

Usher M, McClelland JL. The time course of perceptual choice: The leaky, competing accumulator model. Psychological Review. 2001; 108:550–592. [PubMed: 11488378]

Verhaeghen P. Aging and vocabulary scores: A meta-analysis. Psychology and Aging. 2003; 18:332–339. [PubMed: 12825780]

Verhaeghen P. Aging and executive control: Reports of a demise greatly exaggerated. Current Directions in Psychological Science. 2011; 20:174–180. [PubMed: 25866452]

Verhaeghen P, Marcoen A, Goossens L. Facts and fiction about memory aging: A quantitative integration of research findings. Journal of Gerontology: Psychological Sciences. 1993; 48:157–171.

Ward G. A recency-based account of the list length effect in free recall. Memory & Cognition. 2002; 30:885–892. [PubMed: 12450092]

Ward G, Maylor E. Age-related deficits in free recall: The role of rehearsal. Quarterly Journal Of Experimental Psychology. 2005; 58A:98–119.

Wechsler, D. Wechsler Adult Intelligence Scale. 4. San Antonio, TX: Pearson; 2008.

Werkle-Bergner M, Freunberger R, Sander MC, Lindenberger U, Klimesch W. Inter-individual performance differences in younger and older adults differentially relate to amplitude modulations and phase stability of oscillations controlling working memory contents. NeuroImage. 2012; 60:71–82. [PubMed: 22178810]

West R. An application of prefrontal cortex function theory to cognitive aging. Psychological Bulletin. 1996; 120:272–292. [PubMed: 8831298]

Wingfield A, Lindfield KC, Kahana MJ. Adult age differences in the temporal characteristics of category free recall. Psychology and Aging. 1998; 13:256–266. [PubMed: 9640586]

Wixted JT, Rohrer D. Analyzing the dynamics of free recall: An integrative review of the empirical literature. Psychonomic Bulletin and Review. 1994; 1:89–106. DOI: 10.3758/BF03200763 [PubMed: 24203416]

Yassa MA, Mattfeld A, Stark SM, Stark CEL. Age-related memory deficits linked to circuit-specific disruptions in the hippocampus. Proceedings of the National Academy of Sciences, USA. 2011; 108:8873–8878.

Yonelinas AP. The nature of recollection and familiarity: A review of 30 years of research. Journal of Memory and Language. 2002; 46:441–517.

Zacks, R.; Hasher, L.; Li, KZH. Human memory. In: Craik, FIM.; Salthouse, TA., editors. The handbook of aging and cognition. Mahwah, NJ: Lawrence Erlbaum and Associates; 2000. p. 293-358.

Zaromb FM, Howard MW, Dolan ED, Sirotin YB, Tully M, Wingfield A, Kahana MJ. Temporal associations and prior-list intrusions in free recall. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2006; 32:792–804. DOI: 10.1037/0278-7393.32.4.792

Zelinski EM, Burnight KP. Sixteen-year longitudinal and time lag changes in memory and cognition in older adults. Psychology and aging. 1997; 12:503–513. [PubMed: 9308097]

Zimmerman, S.; Hasher, L.; Goldstein, D. Cognitive ageing: a positive perspective. In: Kapur, N., editor. The paradoxical brain. Cambridge, UK: Cambridge University Press; 2011. p. 130-150.

Appendix A. Model Details

In CMR2 (Lohnas et al., 2015) two types of cognitive representations interact: the feature representation (F), in which the features of the current list item are activated, and the context representation (C), in which the current state of context is activated. Each of these representational layers is defined as a vector. Hebbian associative matrices connect these representations, one connecting features to context (M^{FC}), and one connecting context to features (M^{CF}).

Each association matrix is a weighted sum of a pre-experimental component (M_{pre}^{FC} and M_{pre}^{CF}) that reflects longstanding semantic relationships and an experimental component (M_{exp}^{FC} and M_{exp}^{CF}) that reflects new learning that occurs during the experiment. The semantic associations in M_{pre}^{FC} and M_{pre}^{CF} are defined using Latent Semantic Analysis (LSA; Landauer & Dumais, 1997). LSA measures the semantic relationship between two words as the cosine of the angle between the words' representations in a multidimensional model of semantic space.

Studying an item activates the corresponding features, $\mathbf{f}_i$, which in turn retrieve the context states to which those features have previously been associated:

$$\mathbf{c}_i^{\mathbf{IN}}=\frac{M^{FC}\mathbf{f}_i}{\|M^{FC}\mathbf{f}_i\|} \quad (1)$$

This retrieved context, $\mathbf{c}_i^{\mathbf{IN}}$, which is normalized to have a length of 1 so that it can be directly compared to other states of context (e.g., see the section on post–retrieval editing below), is incorporated into the context representation by adding it to the current context vector $\mathbf{c}_{i-1}$. The context vector is continuously maintained at unit length. Therefore, when a new state of context is added to the existing state, the two vectors, $\mathbf{c}_{i-1}$ and $\mathbf{c}_i^{\mathbf{IN}}$ must be scaled so their sum has a length of one:

$$\mathbf{c}_i=\rho_i\mathbf{c}_{i-1}+\beta\mathbf{c}_i^{\mathbf{IN}} \quad (2)$$

Where β is a model parameter governing how quickly context changes, and ρ_i is chosen such that $\|\mathbf{c}_i\| = 1$:

$$\rho_i=\sqrt{1+\beta^2[(\mathbf{c}_{i-1}\cdot\mathbf{c}_i^{\mathbf{IN}})^2-1]}-\beta(\mathbf{c}_{i-1}\cdot\mathbf{c}_i^{\mathbf{IN}}). \quad (3)$$

Because context is always of unit length it can be thought of as point on the surface of a (hyper) sphere, with β determining how far along the surface of the sphere it travels with each newly presented item and $\mathbf{c}_i^{\mathbf{IN}}$ determining the direction of travel.

At the start of an experimental session, the experimental associations are initialized to zero. As each new item is presented, new experimental associations are formed, both between the item's feature representation and the current state of context (stored in M_{exp}^{FC}) and between the current state of context and the item's feature representation (stored in M_{exp}^{CF}). These associations are formed according to a Hebbian outer-product learning rule:

$$\begin{aligned}\Delta M_{exp}^{FC}&=\mathbf{c}_{i-1}\mathbf{f}_i^{\top}\\ \Delta M_{exp}^{CF}&=\mathbf{f}_i\mathbf{c}_{i-1}^{\top}\end{aligned} \quad (4)$$

These newly formed experimental associations are incorporated with pre-experimental semantic associations. The balance between new and existing associations is controlled by parameters γ_{FC} and γ_{CF}:

$$M^{FC}=(1-\gamma_{FC})(I+s_{FC}M_{pre}^{FC})+\gamma_{FC}M_{exp}^{FC}$$
$$M^{CF}=(1-\gamma_{CF})(I+s_{CF}M_{pre}^{CF})+\gamma_{CF}\phi_i M_{exp}^{CF} \quad (5)$$

s_{FC} and s_{CF} are scaling parameters that control the influence of pre-experimental semantic associations. Note that the Lohnas et al. (2015) implementation of CMR2 used only s_{CF}, effectively setting s_{FC} to zero and preventing features from activating the contexts of semantically associated features. In line with the observation that most extra–list intrusions are semantically related to list items (Zaromb et al., 2006), we found that simulating the extra–list intrusion rates of older adults required allowing features to activate semantically related contexts. I is an identity matrix the same size as M_{pre}^{CF} and M_{pre}^{FC} (The two associative matrices are of the same size for all simulations in this paper). Effectively this means that the on-diagonal terms are not multiplied by the s parameter. This allows the s parameter to scale semantic relations between pairs of different items while having no effect on auto-associations.

ϕ_i simulates increased attention to beginning-of-list items, producing a primacy effect, by scaling the magnitude of context-to-feature associations across the list:

$$\phi_i=\phi_s e^{-\phi_d(i-1)}+1, \quad (6)$$

where ϕ_s and ϕ_d are model parameters. See Sederberg et al. (2008) for a more complete discussion.

In previous implementations of the retrieved context framework the experimental component of the associative matrices were re-initialized for each simulated list, erasing the learning that occurred on previous lists. By contrast, in CMR2 the matrices are initialized once at the beginning of the simulated experiment and allowed to accumulate associations across all lists. This development allows CMR2 to simulate prior list intrusions. In a departure from the Lohnas et al. (2015) implementation of CMR2, we include an extended vocabulary that includes a strong associate of each presented item. These items are included on the assumption that ELIs are driven by strong semantic associations, and can be most naturally modeled by providing the model with a set of rich semantic associates.

During recall, the current contextual state is used to cue retrieval via the M^{CF} associations:

$$\mathbf{f}_t^{\mathbf{IN}}=M^{CF}\mathbf{c}_t, \quad (7)$$

The resulting $\mathbf{f}_t^{\mathbf{IN}}$ gives the degree of support, or activation, for each item in the model's vocabulary. Items with low activation values are unlikely to be recalled by the retrieval process described below and considering them as candidates for retrieval is extremely computationally expensive. Therefore, only the 40 items with the highest activation values are assigned to a vector **a** of retrieval candidates. The values in **a** are then used as the initial input for a set of competitive accumulators, one for each candidates, according to the leaky competitive accumulator model of Usher and McClelland (2001):

$$\begin{aligned}\mathbf{x}_n &= (1-\tau\kappa-\tau\lambda\mathbf{N})\,\mathbf{x}_{n-1}+\tau\mathbf{a}+\varepsilon,\\ \mathbf{x}_n &\rightarrow max(\mathbf{x}_n,\mathbf{0})\end{aligned} \quad (8)$$

$\mathbf{x}_n$ is a vector with one element for each retrieval candidate in **a**. When the retrieval competition starts, all elements are set to zero (i.e., $\mathbf{x}_0 = \mathbf{0}$) and the activation for each item given in **a** is used as its starting line in the race to threshold. τ is a fixed time constant, κ is a parameter that determines the decay rate for item activations, and λ is the lateral inhibition parameter, scaling the strength of an inhibitory matrix **N** that subtracts each item's activations from all of the others except itself. ε is a random vector whose elements are drawn from a random normal distribution with mean zero and standard deviation η. The second line of Equation 8 means that the accumulating elements cannot take on negative values. $\mathbf{x}_n$ continues to be updated until one of the activation values exceeds its threshold or until the recall period ends.

CMR2 dynamically sets the retrieval threshold of each item as the recall period progresses to allow items that were recalled earlier in the period to participate in, but not dominate, current retrieval competitions. Specifically, at the beginning of the recall period, each item, i, has a threshold of $\Theta_i = 1$. If item i is retrieved, its threshold is incremented by a value ω and then gradually returns to 1 with subsequent recalls:

$$\Theta_i=1+\omega\alpha^j \quad (9)$$

Where j is the number of subsequent retreivals, α is a parameter between 0 and 1; the larger the value of α, the more intervening retrievals are needed before an already-recalled item is likely to be retrieved again.

The first word to accumulate enough activation to cross its threshold wins the retrieval competition. The winner's representation is reactivated on ϕ, allowing the model to retrieve the contextual state associated with the item. Context is updated using the same mechanism used during the study period (separate parameters, β_{enc} and β_{rec} control the rate of context drift during encoding and retrieval). Before the item is actually output by the model, however, it undergoes a post-retrieval editing phase, consistent with the observation that people often report thinking of items that they do not overtly recall during free recall experiments (Keppel, 1968; Wixted & Rohrer, 1994). Editing is accomplished by comparing the context representation retrieved by the candidate item with the currently active context representation (J. R. Anderson & Bower, 1972; Dennis & Humphreys, 2001):

$$\mathbf{c}_{t+1}^{\mathbf{IN}}\cdot\mathbf{c}_t. \quad (10)$$

Because associations are formed between items and the context that prevailed when they were originally presented, true list items will tend to retrieve a context that is similar to the context that prevails during retrieval. The match between retrieved context and the current context will depend on how much contextual drift has occurred between original presentation and the recall event. Relatively little drift will have occurred for items that were actually presented on the current list, whereas considerable drift will have occurred for items that were presented on earlier lists. Thus, on average, accurate recalls will produce a higher value than either PLIs or ELIs. If the comparison returns a value that is beneath a threshold parameter, c_{thresh}, the item is rejected as an intrusion.

Once an item is either recalled or rejected, another recall competition begins. This series of competitions continues until the end of the recall period is reached, at which point the next trial begins. It is thought that the transition from one trial to the next is accompanied by a change in temporal context (Sahakyan & Kelley, 2002). CMR2 simulates this shift in temporal context by activating a unique "disruption" item on the feature layer and allowing this item to update context using Equation 2, with a post-recall drift rate parameter, β_{post}^{recall}.

Appendix B. Details of Simulations

The model parameters obtained by fitting the full model (i.e., allowing all parameters to vary) independently to the average younger adult and average older adult data from Kahana et al. (2002) are shown in Table B1. These are the parameters used to generate the model simulations shown in Figure 6. Table B1 also shows the average younger and older adult parameter values obtained by fitting the model to individual Kahana et al. (2002) participants' data. Table B2 shows the correlations among the individual subject parameter estimates across all subjects in the Kahana et al. (2002) dataset (i.e., for each parameter we have a vector of the individual participants' estimated values for that parameter; Table B2 shows the correlations among those vectors).

Table B3 shows the parameter values obtained by fitting each of the verbal theory implementations, both independently to each effect (i.e., the values used to generate Figures 7—11) and simultaneously to all the effects (i.e., the values used to generate Figures 12—14). The Table B3 also shows the Root Mean Squared Deviation (RMSD) values for each model.

Table B4 shows the best fitting parameter values obtained by fitting the full model, the four–competent model, and the combined aging theories to data from the Penn Electrophysiology of Encoding and Retrieval Study. The full model was fit to younger adult data, whereas the four–component model and the combined aging theories were fit to older adult data.

Table B1

The optimized parameter values for simulations of average data and individual participant data from Kahana et al. (2002).

Parameter	Younger (average data)	Older (average data)	Younger (individual data)	Older (individual data)
ϕ_s	1.283	1.553	1.483	1.655
ϕ_d	0.960	1.145	0.884	1.034
s_{CF}	6.147	3.398	2.109	2.177
s_{FC}	0.006	0.005	0.004	0.001
γ_{CF}	0.984	0.969	0.918	0.898
γ_{FC}	0.540	0.583	0.411	0.403
β_{enc}	0.561	0.552	0.514	0.522
β_{rec}	0.375	0.266	0.421	0.329
κ	0.108	0.104	0.274	0.289
λ	0.178	0.275	0.187	0.263
η	0.427	0.492	0.431	0.496
c_{thresh}	0.000	0.000	0.052	0.009
α	0.617	0.591	0.810	0.809
ω	13.658	4.726	12.515	12.201
β_{post}^{recall}	0.961	0.922	0.924	0.905
τ	10.000	10.000	10.000	10.000
ε	0.000	0.000	0.000	0.000

Table B2

Pairwise correlations between parameters for the individual subject fits to the Kahana et al. (2002) data.

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. ϕ_s	–														
2. ϕ_d	0.24	–													
3. s_{CF}	**0.26**	−0.05	–												
4. s_{FC}	−0.05	0.13	0.04	–											
5. γ_{CF}	−0.06	0.02	0.18	−0.04	–										
6. γ_{FC}	−0.11	0.16	−0.02	0.19	−0.02	–									
7. β_{enc}	0.20	−0.09	0.11	−0.01	−0.10	−0.05	–								
8. β_{rec}	−0.16	−0.14	0.09	0.08	−0.09	−0.11	−0.00	–							
9. κ	−0.08	0.17	−0.02	−0.11	−0.10	−0.22	0.15	0.11	–						
10. λ	0.19	−0.04	0.09	**−0.38**	−0.01	**−0.31**	0.01	**−0.30**	−0.05	–					
11. η	0.17	−0.01	−0.02	**−0.38**	−0.15	**−0.38**	0.02	−0.24	0.11	**0.93**	–				
12. c_{thresh}	**−0.30**	−0.08	−0.14	**0.39**	0.21	**0.28**	−0.13	0.07	**−0.30**	**−0.54**	**−0.63**	–			
13. α	0.02	−0.09	0.05	0.10	−0.03	−0.16	0.00	−0.02	−0.24	−0.21	−0.23	0.25	–		
14. ω	0.01	0.01	0.18	0.15	0.11	0.22	**0.34**	0.14	−0.02	−0.06	−0.10	0.07	−0.04	–	

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
15. β_{post}^{recall}	0.11	−0.08	**0.34**	**0.27**	0.01	0.06	−0.00	0.16	−0.25	−0.05	−0.17	0.25	0.20	0.10	–

Correlations in **bold** are significant at α = .05

Table B3

The optimized parameter values for each verbal theory implementation's simulation of all data simultaneously and each effect independently.

Theory	Implementation	Parameter	All Data	SPC	PFR	Lag-CRP	Intrusions	PLI-Recency
ADH	Weak Associations	γ_{CF}	0.986	0.926	0.956	0.899	0.964	0.983
		γ_{FC}	0.436	0.510	0.614	0.477	0.966	0.874
		RMSD	0.065	0.038	0.023	0.025	0.044	0.007
	Noisy Associations	ε	0.001	0.061	0.186	0.052	0.000	0.000
		RMSD	0.082	0.058	0.026	0.023	0.091	0.074
	Weak & Noisy	γ_{CF}	0.977	0.943	0.961	0.895	0.982	0.982
		γ_{FC}	0.389	0.389	0.316	0.359	0.357	0.716
		ε	0.004	0.006	0.002	0.007	0.000	0.000
		RMSD	0.070	0.030	0.020	0.020	0.020	0.019
IDH	Drift Rate	β_{enc}	0.610	0.596	0.489	0.532	0.630	0.458
		β_{rec}	0.551	0.846	0.887	0.252	0.514	0.823
		RMSD	0.065	0.046	0.019	0.032	0.001	0.006
	List Isolation	β_{post}^{recall}	0.807	0.362	0.095	0.711	0.923	0.746
		RMSD	0.071	0.056	0.024	0.043	0.093	0.006
	Retrieval Editing	c_{thresh}	0.006	0.264	0.001	0.088	0.004	0.019
		RMSD	0.073	0.103	0.030	0.045	0.084	0.026
CSH	Slowed Encoding	γ_{CF}	0.986	0.926	0.956	0.899	0.964	0.983
		γ_{FC}	0.436	0.510	0.614	0.477	0.966	0.874
		RMSD	0.065	0.038	0.023	0.025	0.044	0.007
	Slowed Retrieval	τ	18.286	14.524	9.190	18.524	39.619	9.857
		RMSD	0.069	0.040	0.035	0.029	0.271	0.034
	Both Slowed	τ	17.818	14.545	53.182	37.727	9.182	11.818
		γ_{CF}	0.959	0.943	0.975	0.927	0.986	0.973
		γ_{FC}	0.877	0.552	0.318	0.761	0.320	0.789
		RMSD	0.049	0.026	0.020	0.022	0.003	0.008
All Theories Combined	—	γ_{CF}	0.962					
		γ_{FC}	0.595					
		ε	0.002					
		β_{enc}	0.519					
		β_{rec}	0.181					
		β_{post}^{recall}	0.749					
		c_{thresh}	0.006					

Theory	Implementation	Parameter	All Data	SPC	PFR	Lag-CRP	Intrusions	PLI-Recency
		τ	18.339					
		RMSD	0.028					
Four–Component Model	—	β_{rec}	0.251					
		ϕ_s	1.716					
		ϕ_d	1.011					
		c_{thresh}	0.009					
		λ	0.273					
		η	0.493					
		RMSD	0.028					

RMSD = Root Mean Square Deviation, which was minimized by the genetic algorithm.

Table B4

The optimized parameter values for simulations of data from the Penn Electrophysiology of Encoding and Retrieval Study.

Parameter	Full Model fit to younger data	Four–Component Model fit to older data	Combined Aging Theories fit to older data
ϕ_s	1.700	2.229	
ϕ_d	0.306	0.426	
s_{CF}	8.277		
s_{FC}	0.005		
γ_{CF}	0.925		0.928
γ_{FC}	0.480		0.313
β_{enc}	0.466		0.462
β_{rec}	0.443	0.265	0.450
κ	0.539		
λ	0.133	0.278	
η	0.360	0.475	
c_{thresh}	0.001	0.000	0.000
α	3.765		
ω	8.907		
β_{post}^{recall}	0.940		0.940
τ	10.000		20.000
ε	0.000		0.034

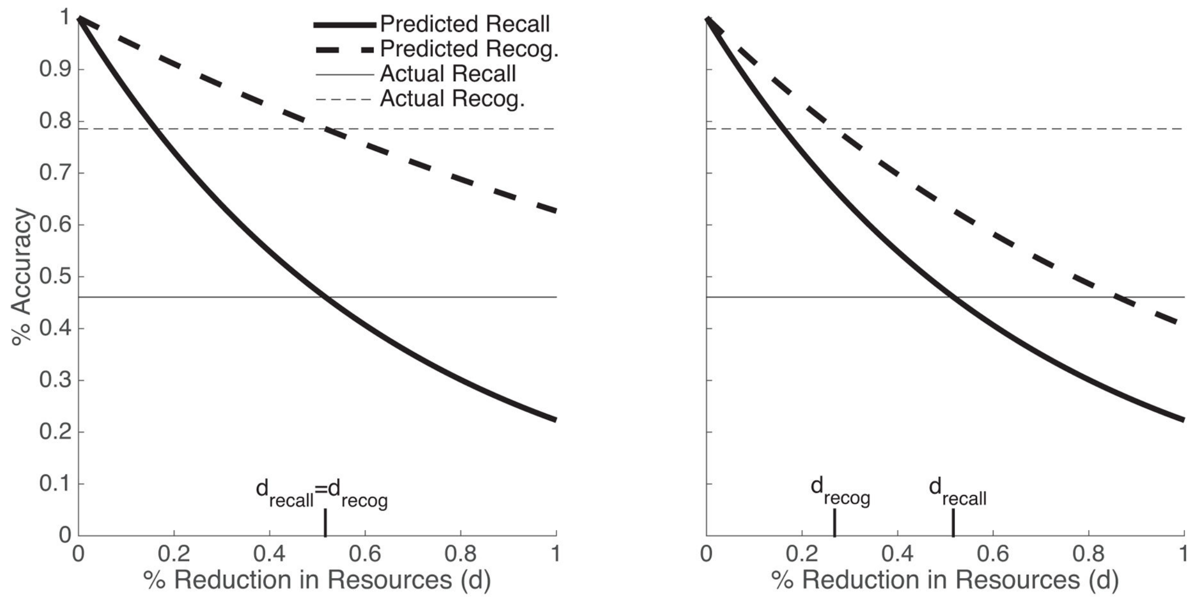

Figure 1.

Two hypothetical sets of functions for the relationships between reduced processing resources and deficits on the free recall and recognition tasks. In each panel the thick solid line shows predicted memory accuracy for different levels of reduced processing resources, the thick dotted line shows the predicted recognition performance (hit rate). The thiner horizontal lines show the actual performance level. For the first set of functions (A) there is a single value of impaired resources at which the model simultaneously predicts correct performance on both recall and recognition. For the second set of functions (B) there is no single value of impairment that allows the model to predict performance on both tasks.

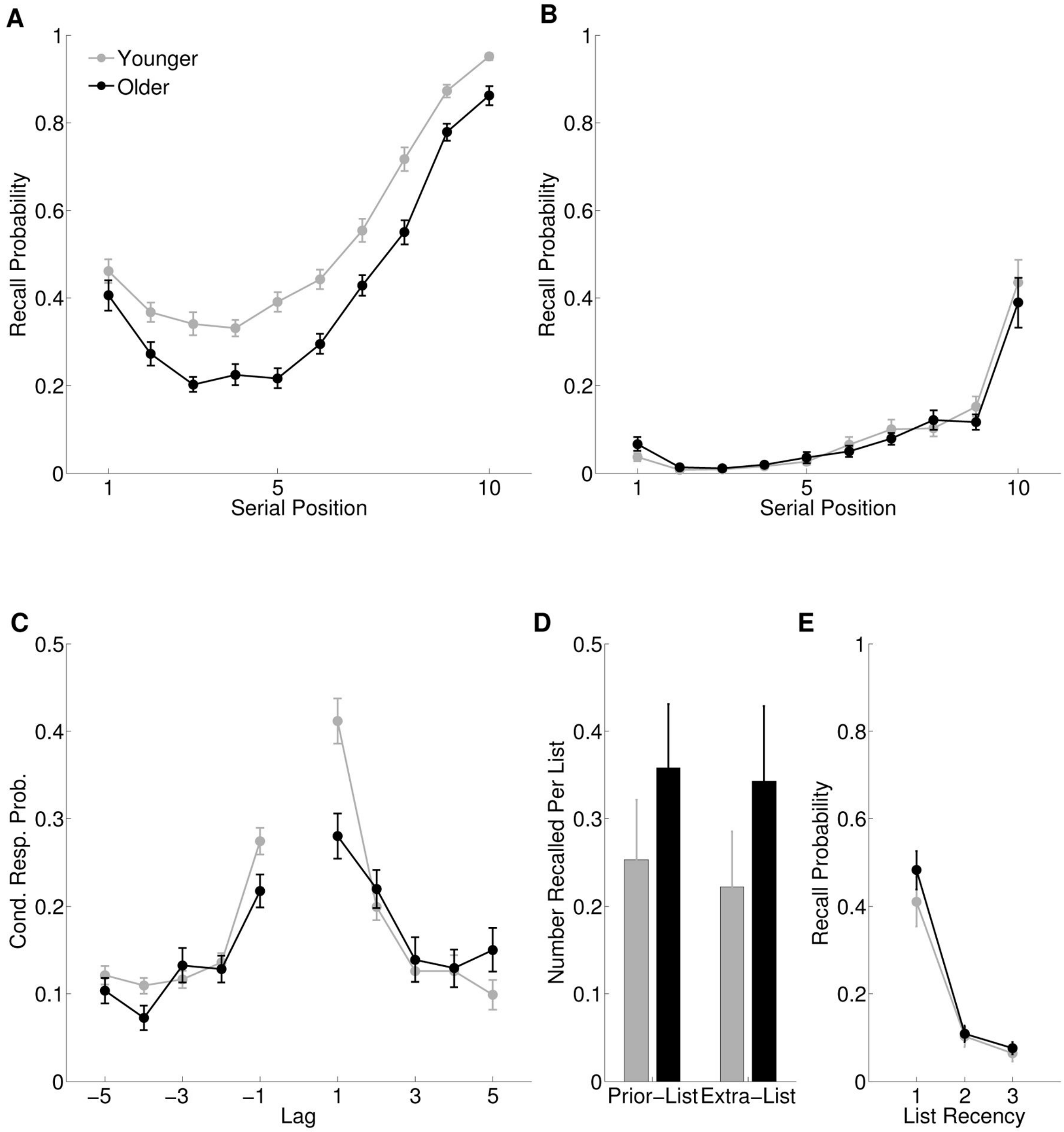

Figure 2.
The Free Recall Aging Pattern: Serial position curve (A), probability first recall function (B), lag-conditional response probability function (C), Prior–list and Extra–list intrusions (D), and Prior–list intrusion recency effect (E). Error bars represent one standard error of the mean. Data from Kahana et al. (2002).

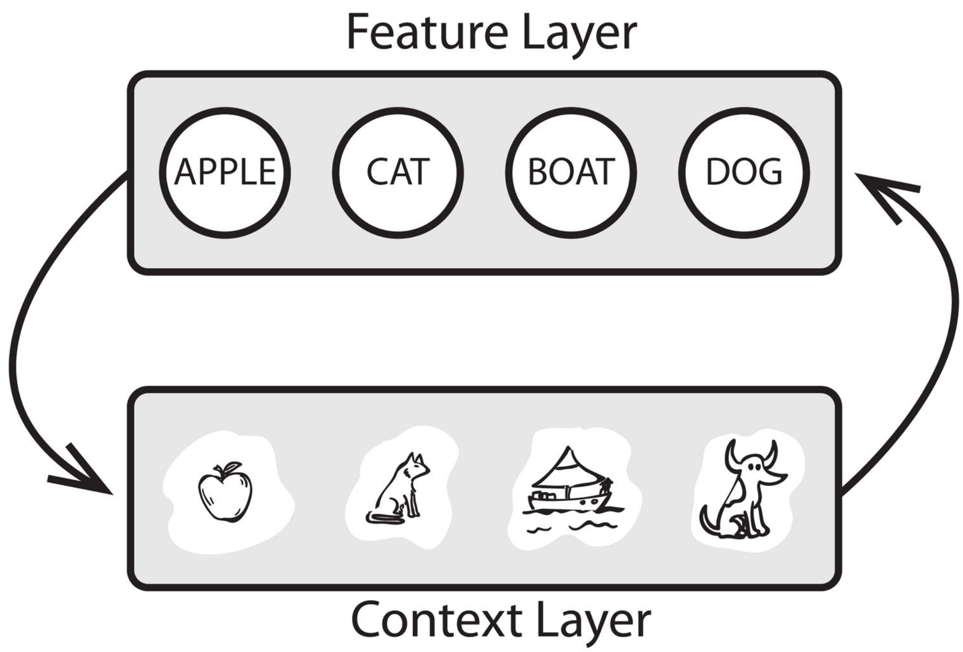

Figure 3.

Schematic of CMR2. The feature layer represents the identity of list items, with one node for each item. The context layer represents the ensemble of contextual associates that are activated when an item is presented; each item has a corresponding context node. The two layers are connected by two associative matrices; one encoding feature–to–context associations and one encoding context–to–feature associations.

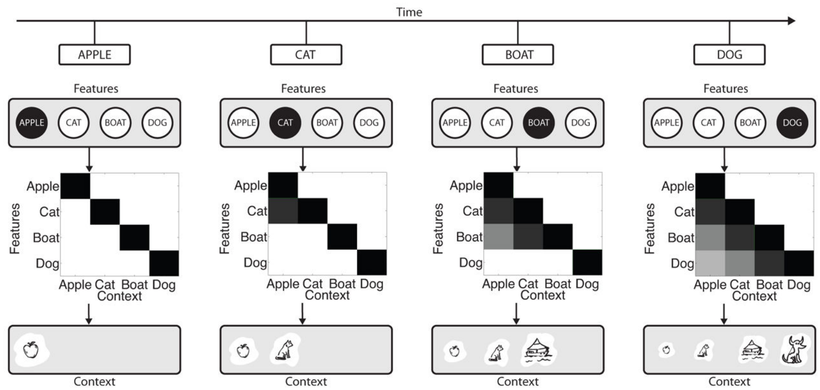

Figure 4.

Sequence of events during the encoding period. When the first item, *apple*, is presented its feature node is activated, which in turn activates its context representation by projecting through the feature-to-context associative weight matrix. When the next item, *cat*, is presented, its context representation is activated, and added to the existing context representation so that both *apple* and *cat* are active on the context layer. As each item is presented Hebbian learning creates links between co-active items and context nodes. At the end of the trial the context layer provides a recency–weighted history of item presentation and the associative matrices encode newly formed associations between items and contexts that were co-active.

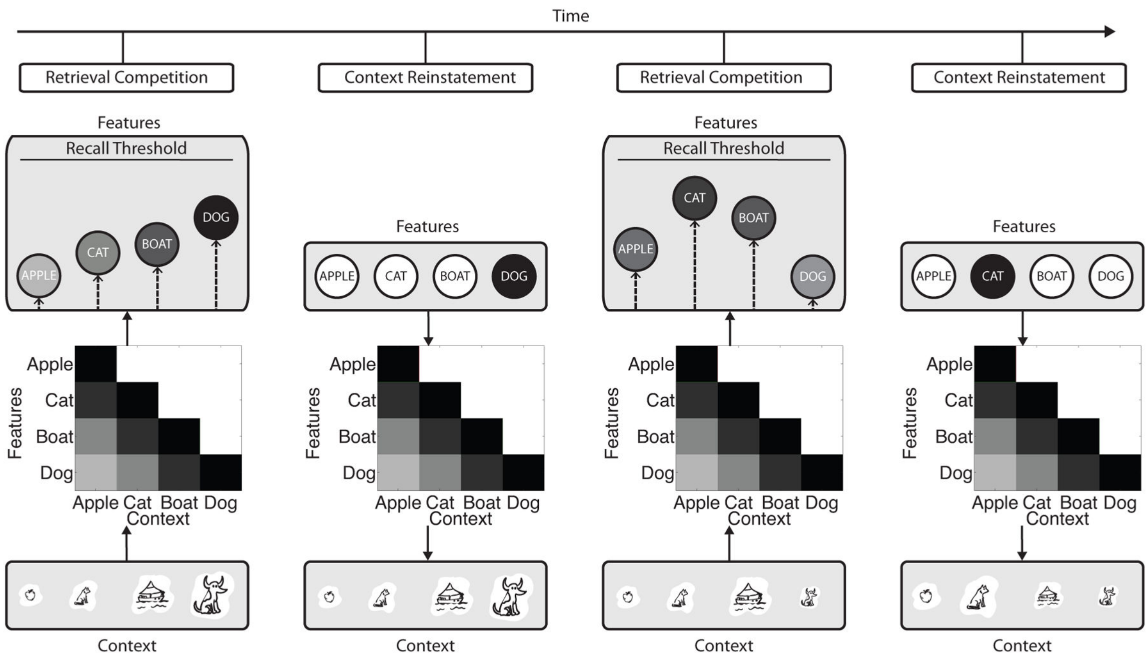

Figure 5.
Sequence of events during the retrieval period. The current state of context is used as a retrieval cue by allowing it to project through the context–to–feature associative matrix, activating nodes on the feature layer. This activation is used as the starting point for an noisy diffusion process in which the first item to cross an activation threshold is recalled. The recalled item is reactivated on the feature layer and activates its associated context representation, which is then added to the previous state of context. This updated context is used as a cue to initiate a second retrieval competition. Retrieval competitions continue until the recall period ends.

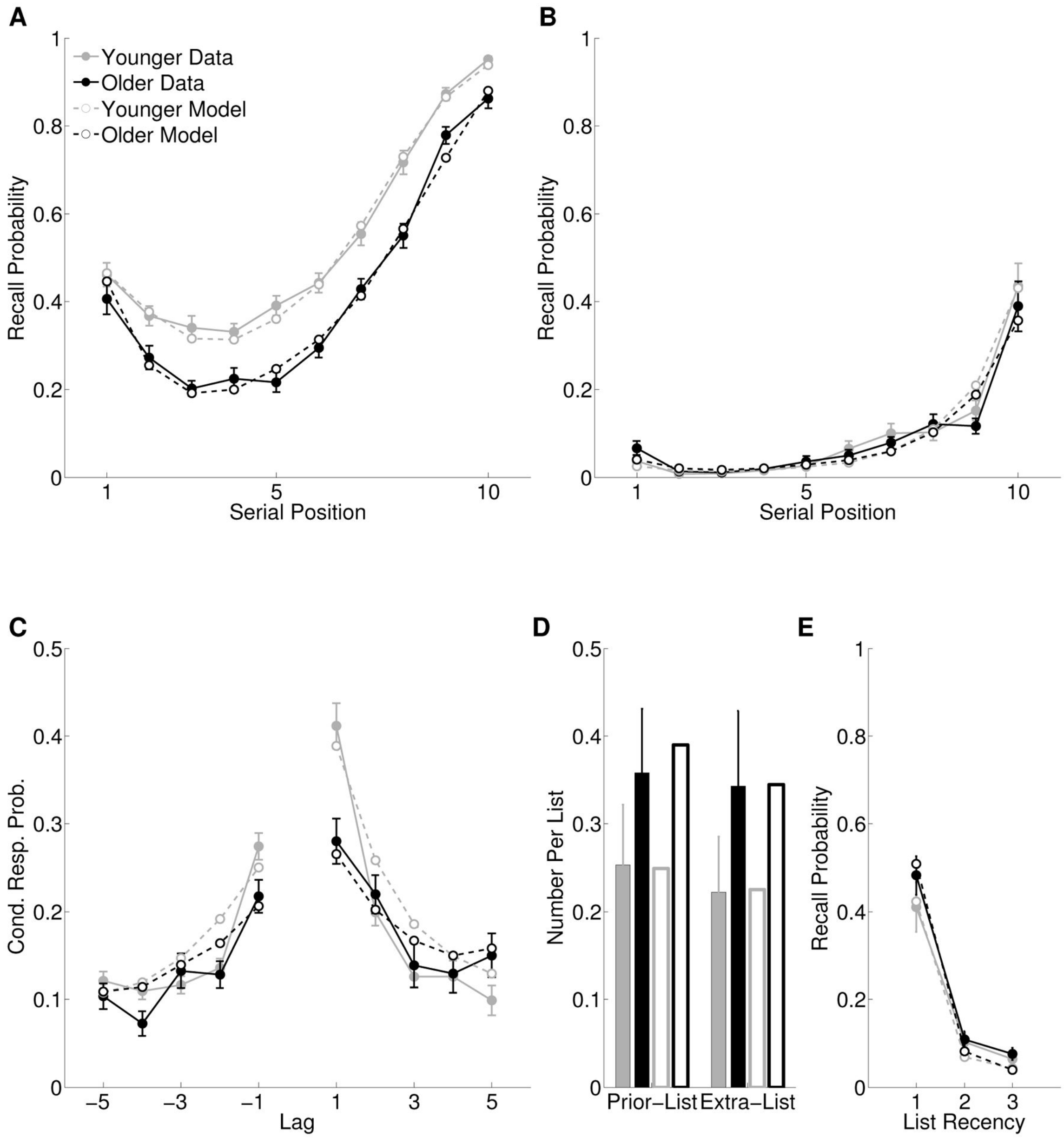

Figure 6.

CMR2 simulations: Serial position curve (A), probability first recall function (B), lag-conditional response probability function (C), Prior–list and Extra–list intrusions (D), and Prior–list intrusion recency effect (E). The model was fit independently to the younger and older adult data and all model parameters were allowed to vary. Black lines or bars are used for older adults. Grey lines or bars are used for younger adults. Solid lines with filled symbols or filled bars are used for participant data. Broken lines with open symbols or unfilled bars are used for model simulations. Error bars represent one standard error of the mean. Data from Kahana et al. (2002).

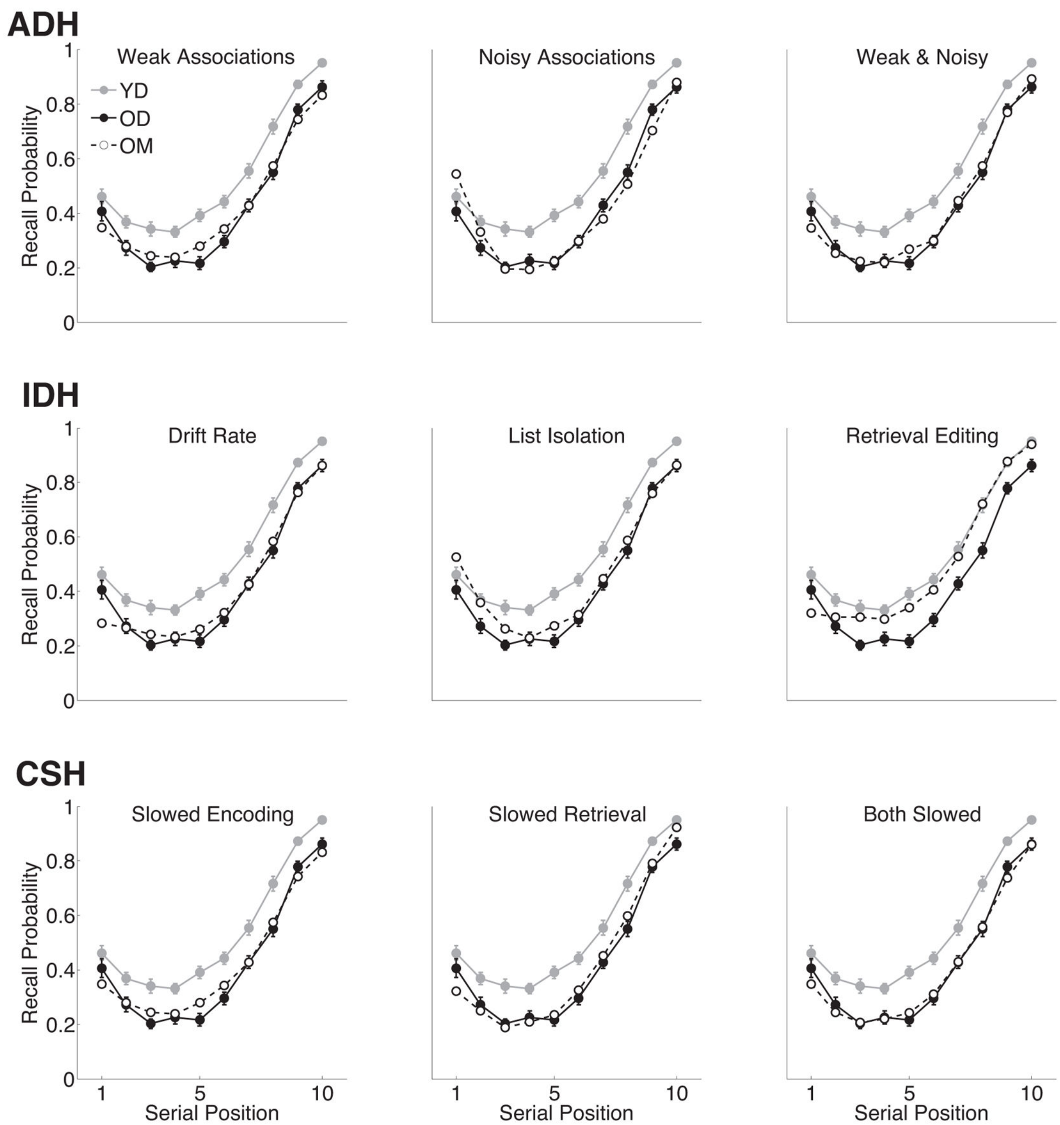

Figure 7.

The lesioned models' simulations of the serial position curve in the Kahana et al. (2002) data. Each panel represents one version of the model and shows the best–fitting simulated data along with the actual data from both older and younger adults. Black lines are used for older adults. Grey lines are used for younger adults. Solid lines with filled symbols are used for participant data. Broken lines with open symbols are used for model simulations. YD = Younger Data; OD = Older Data; OM = Older Model. Error bars represent one standard error of the mean.

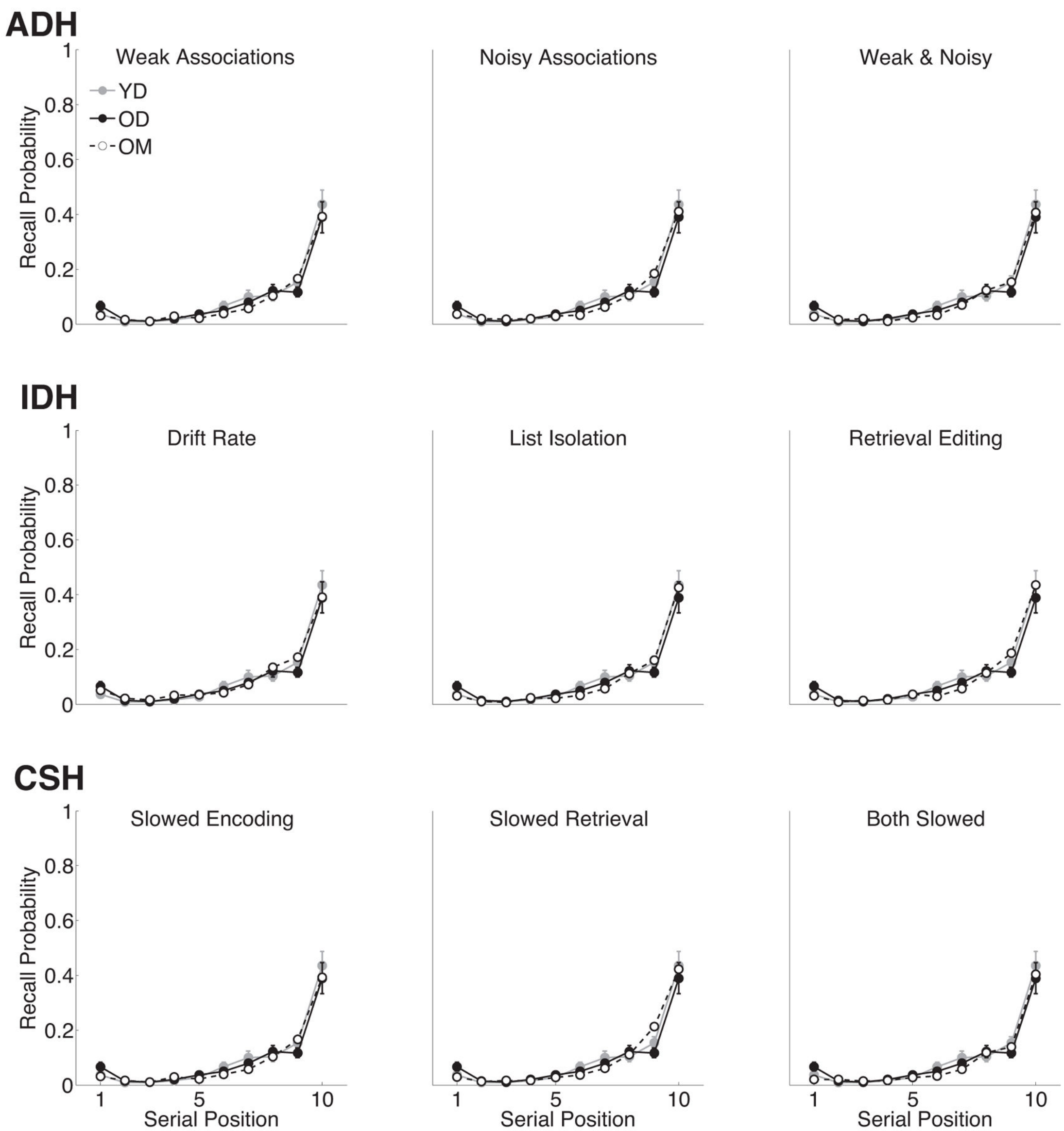

Figure 8.

The lesioned models' simulations of the probability of first recall function in the Kahana et al. (2002) data. Each panel represents one version of the model and shows the best–fitting simulated data along with the actual data from both older and younger adults. Black lines are used for older adults. Grey lines are used for younger adults. Solid lines with filled symbols are used for participant data. Broken lines with open symbols are used for model simulations. YD = Younger Data; OD = Older Data; OM = Older Model. Error bars represent one standard error of the mean.

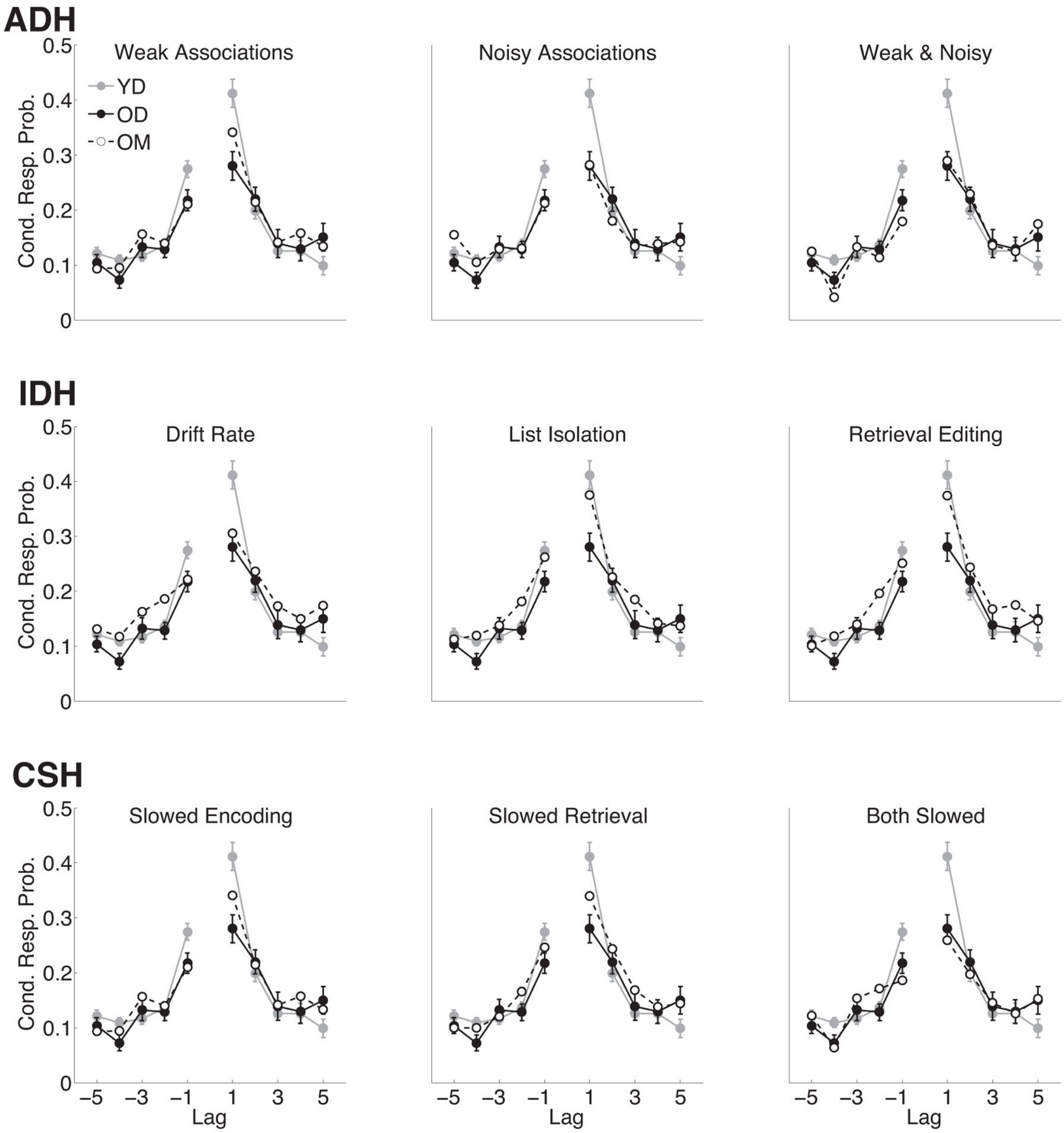

Figure 9.

The lesioned models' simulations of the Lag-CRP Curve in the Kahana et al. (2002) data. Each panel represents one version of the model and shows the best–fitting simulated data along with the actual data from both older and younger adults. Black lines are used for older adults. Grey lines are used for younger adults. Solid lines with filled symbols are used for participant data. Broken lines with open symbols are used for model simulations. YD = Younger Data; OD = Older Data; OM = Older Model. Error bars represent one standard error of the mean.

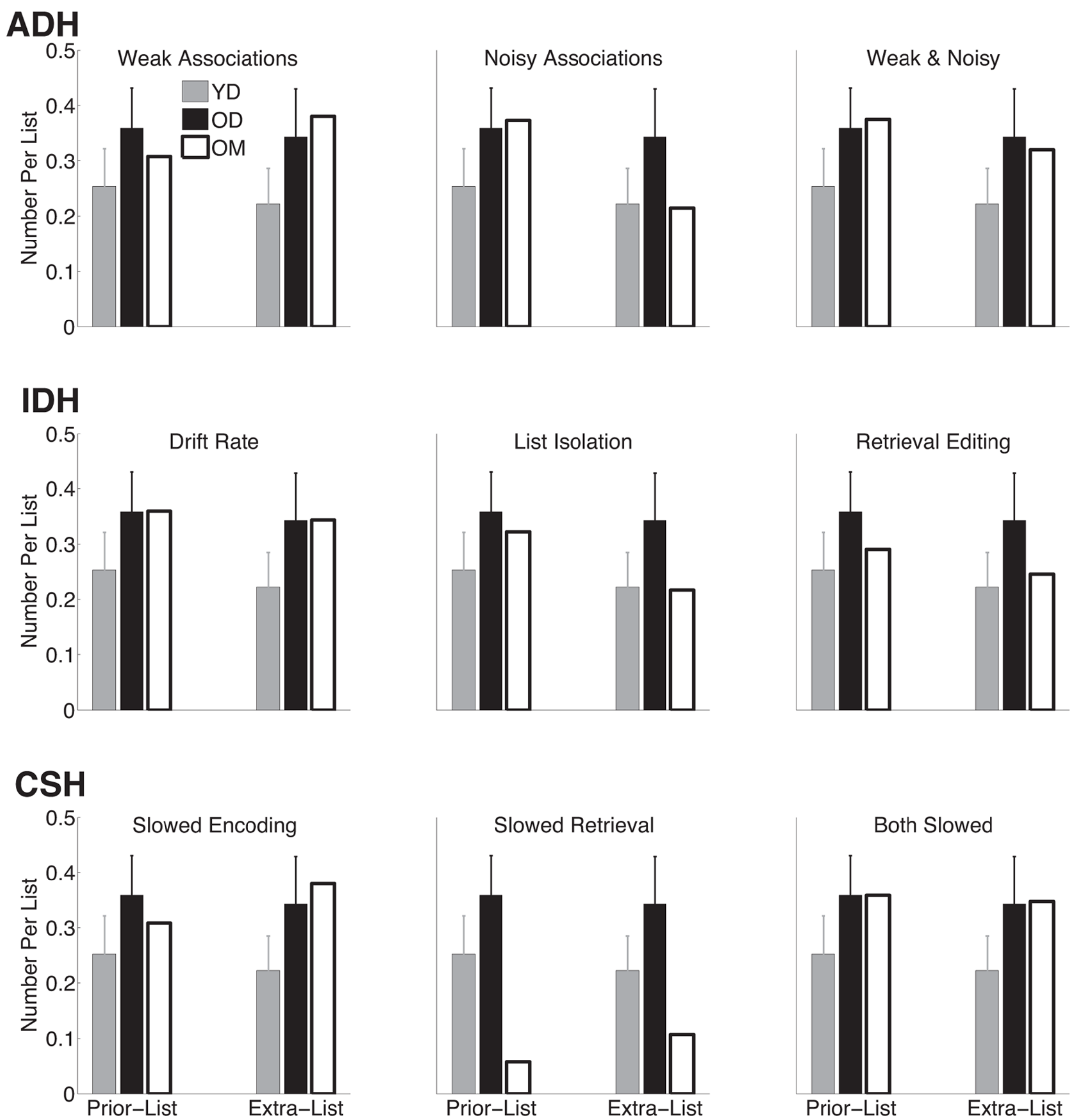

Figure 10.

The lesioned models' simulations of prior–list and extra–list intrusions in the Kahana et al. (2002) data. Each panel represents one version of the model and shows the best–fitting simulated data along with the actual data from both older and younger adults. Black bars are used for older adults. Grey bars are used for younger adults. Filled bars are used for participant data. Unfilled bars are used for model simulations. YD = Younger Data; OD = Older Data; OM = Older Model. Error bars represent one standard error of the mean.

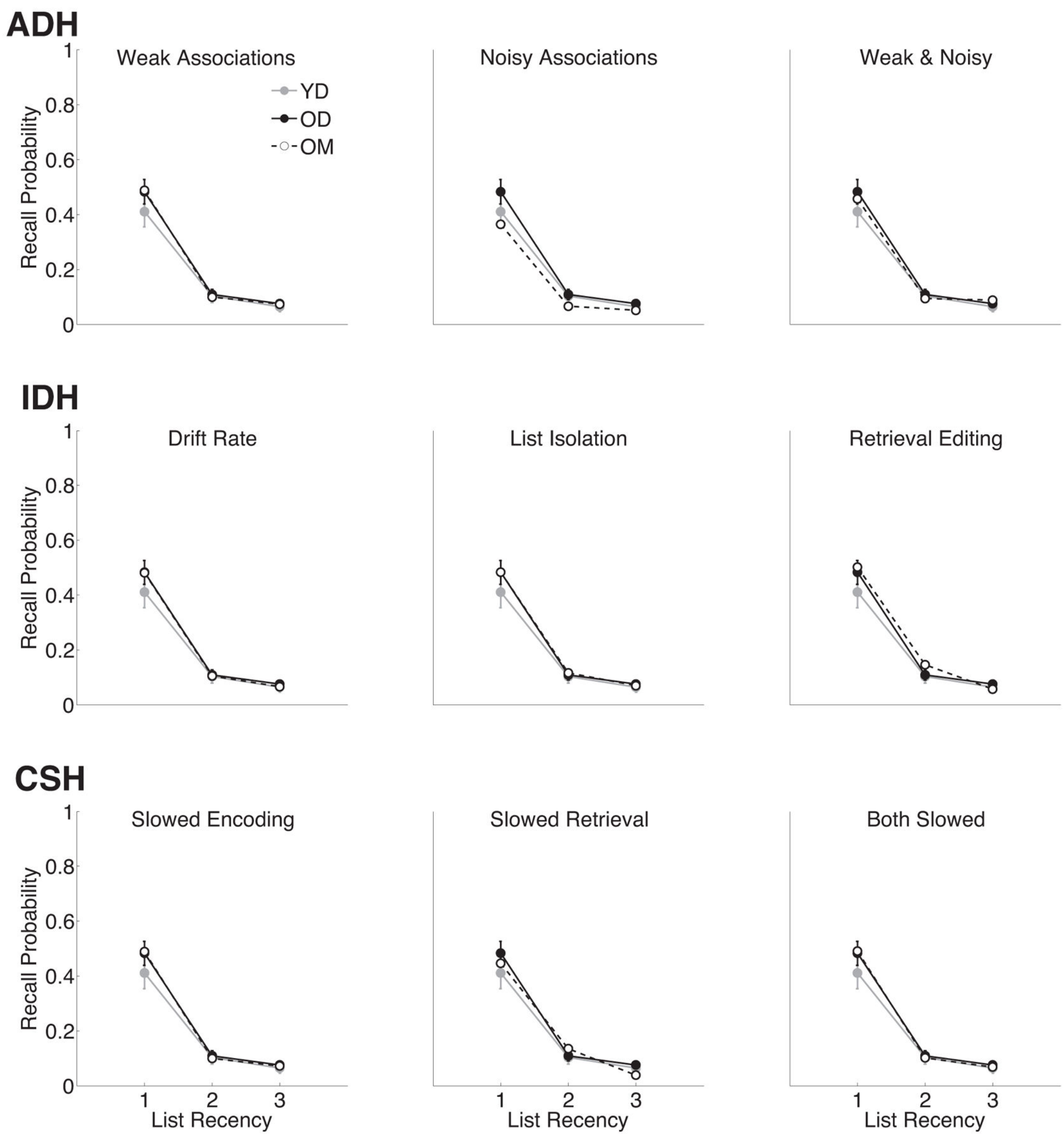

Figure 11.

The lesioned model's simulations of the prior–list recency effect in the Kahana et al. (2002) data. Each panel represents one version of the model and shows the best–fitting simulated data along with the actual data from both older and younger adults. Black lines are used for older adults. Grey lines are used for younger adults. Solid lines with filled symbols are used for participant data. Broken lines with open symbols are used for model simulations. YD = Younger Data; OD = Older Data; OM = Older Model. Error bars represent one standard error of the mean.

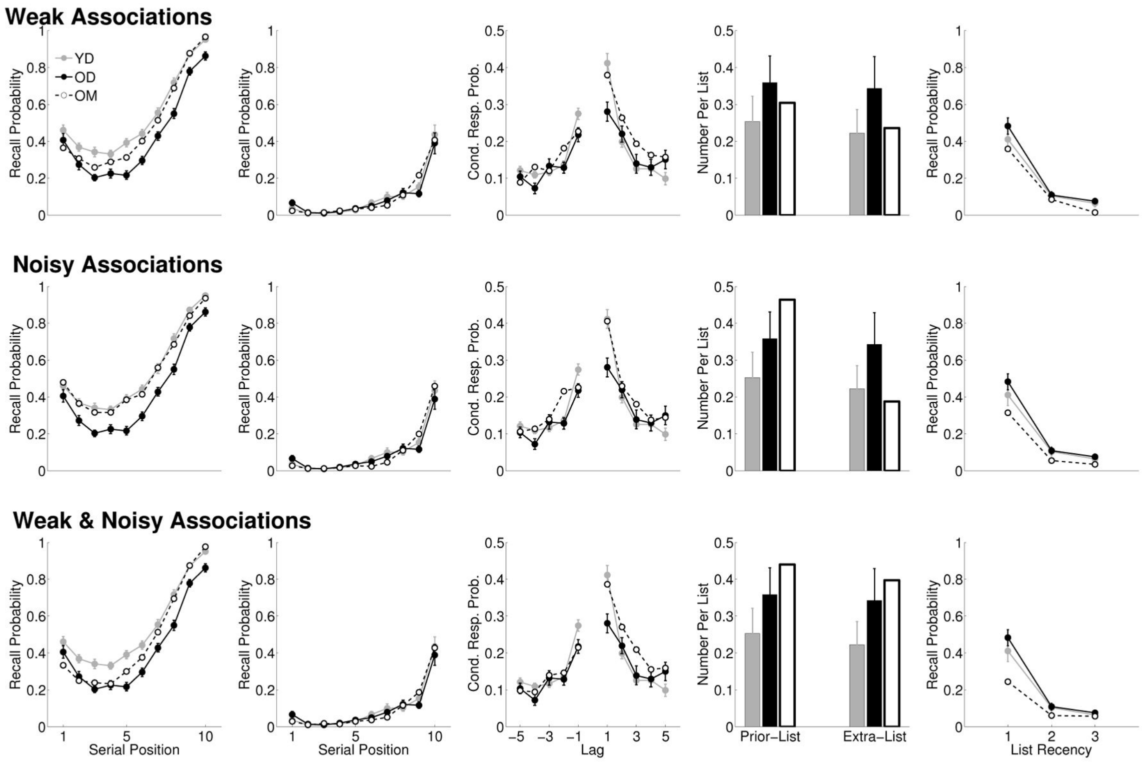

Figure 12.

The Associative Deficit Hypothesis simultaneously simulating all effects in the Kahana et al. (2002) data. Each row represents one implementation of the theory and shows the best–fitting simulated data from that model along with the actual data from both older and younger adults. See the text for full details on each implementation. Black lines and bars are used for older adults. Grey lines and bars are used for younger adults. Solid lines with filled symbols and filled bars are used for participant data. Broken lines with open symbols and unfilled bars are used for model simulations. YD = Younger Data; OD = Older Data; OM = Older Model. Error bars represent one standard error of the mean.

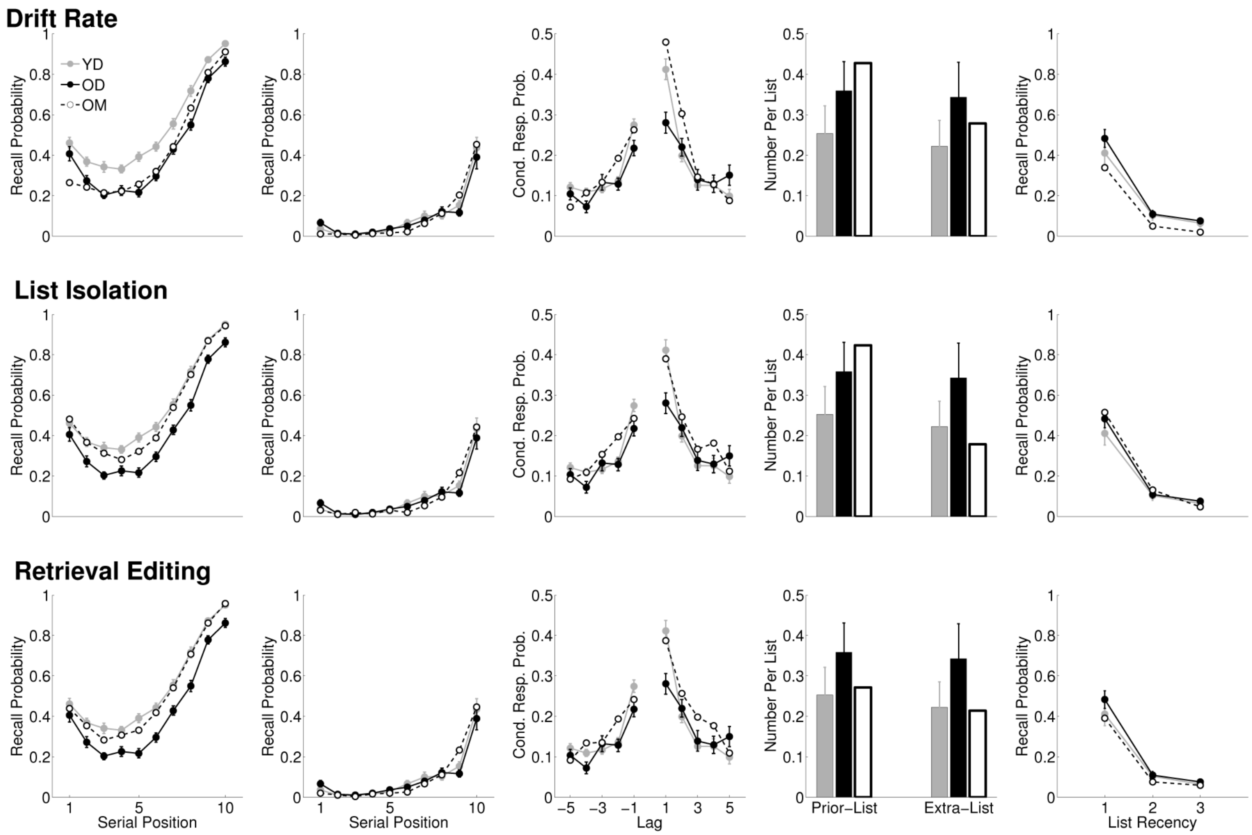

Figure 13.
The Inhibitory Deficit Hypothesis simultaneously simulating all effects in the Kahana et al. (2002) data. Each row represents one implementation of the theory and shows the best–fitting simulated data from that model along with the actual data from both older and younger adults. See the text for full details on each implementation. Black lines and bars are used for older adults. Grey lines and bars are used for younger adults. Solid lines with filled symbols and filled bars are used for participant data. Broken lines with open symbols and unfilled bars are used for model simulations. YD = Younger Data; OD = Older Data; OM = Older Model. Error bars represent one standard error of the mean.

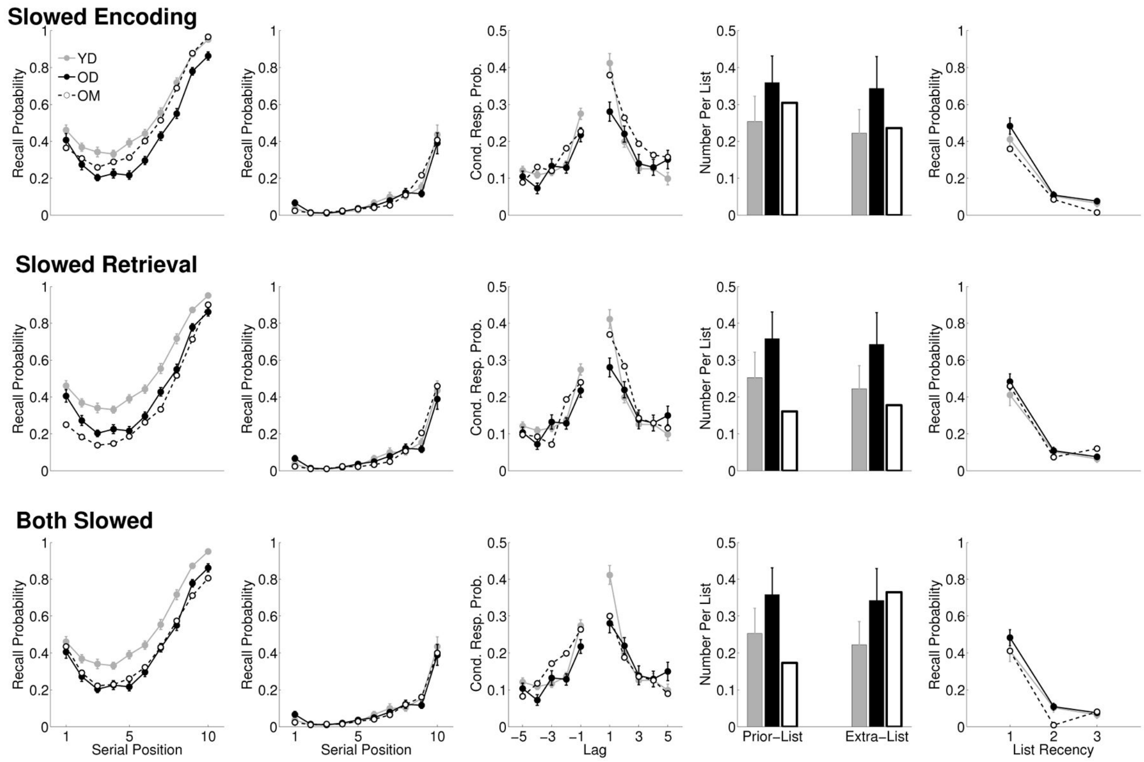

Figure 14.

The Cognitive Slowing Hypothesis simultaneously simulating all effects in the Kahana et al. (2002) data. Each row represents one implementation of the theory and shows the best–fitting simulated data from that model along with the actual data from both older and younger adults. See the text for full details on each implementation. Black lines and bars are used for older adults. Grey lines and bars are used for younger adults. Solid lines with filled symbols and filled bars are used for participant data. Broken lines with open symbols and unfilled bars are used for model simulations. YD = Younger Data; OD = Older Data; OM = Older Model. Error bars represent one standard error of the mean.

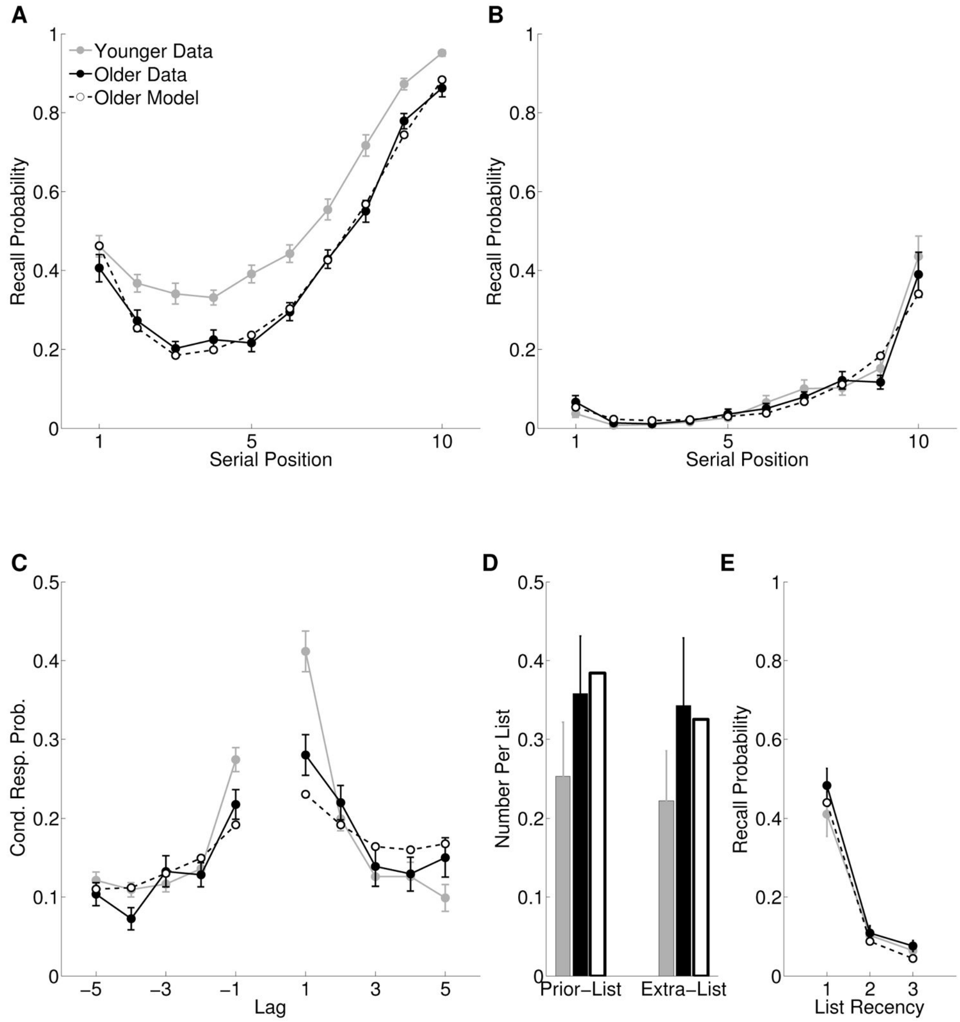

Figure 15.

Fit of the of the combined aging theories to data from Kahana et al. (2002). All parameters implicated by the ADH, IDH, and CSH were allowed to vary. The best–fitting simulated data is shown along with the actual data from both older and younger adults. See the text for full details on how each model was implemented. Black lines or bars are used for older adults. Grey lines or bars are used for younger adults. Solid lines with filled symbols or filled bars are used for participant data. Broken lines with open symbols or unfilled bars are used for model simulations. Error bars represent one standard error of the mean.

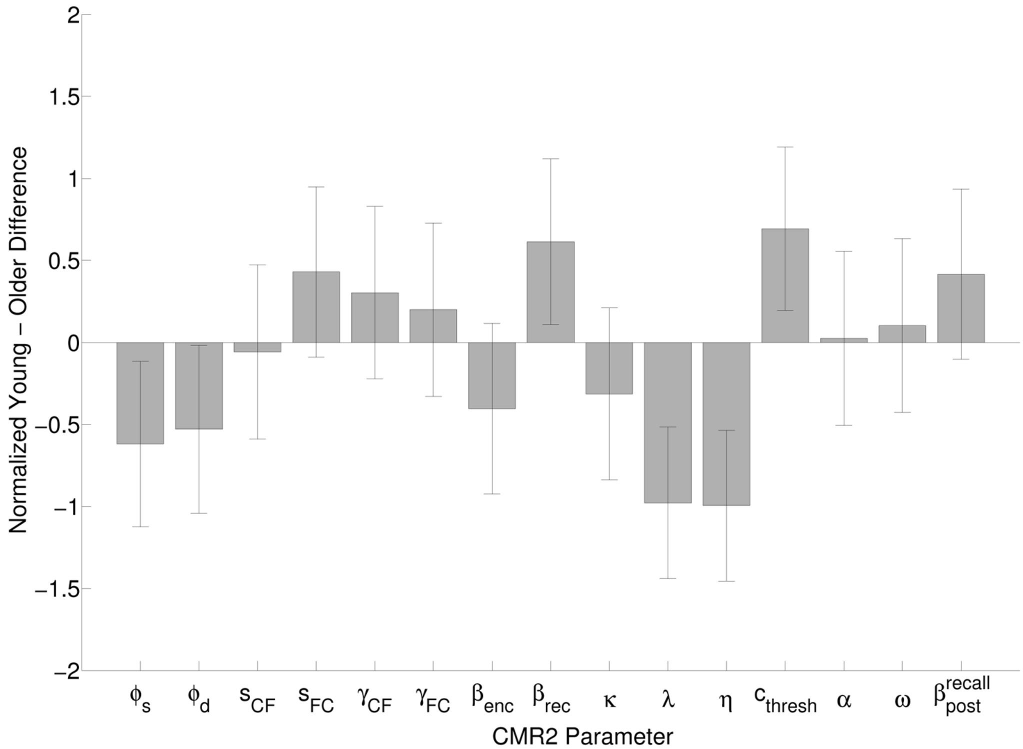

Figure 16.

Age differences in CMR2 parameters derived by fitting individual participants from Kahana et al. (2002). To put each parameter on the same scale, the individual participant estimates for each parameter were z-scored (ignoring age), and then averaged within age group. The bars represent differences in the mean z-scored parameter estimates for younger versus older adults; error bars are 95% confidence intervals on that difference. Parameters for which the confidence interval does not include zero differed significantly between age groups.

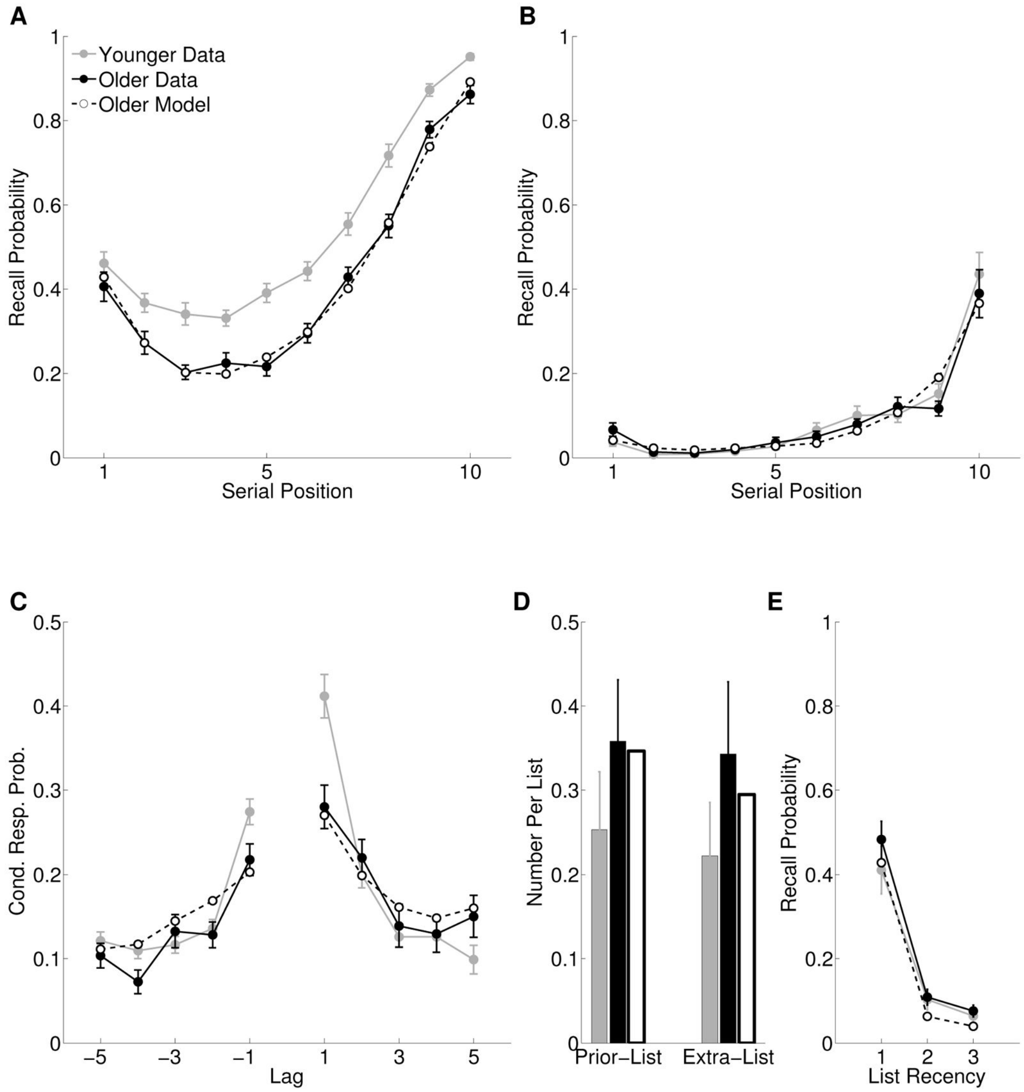

Figure 17.

Fit of the four–component model to data from Kahana et al. (2002). The best–fitting simulated data is shown along with the actual data from both older and younger adults. Black lines or bars are used for older adults. Grey lines or bars are used for younger adults. Solid lines with filled symbols or filled bars are used for participant data. Broken lines with open symbols or unfilled bars are used for model simulations. Error bars represent one standard error of the mean

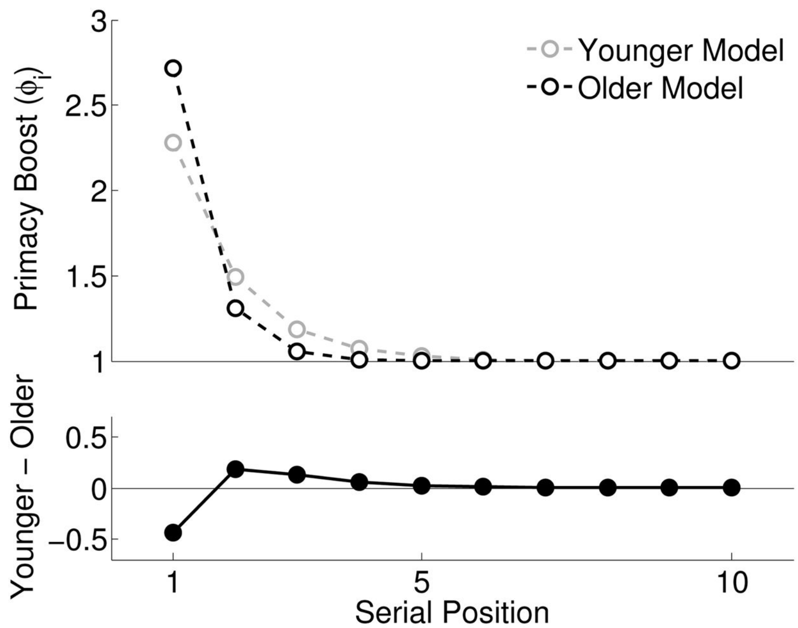

Figure 18.

Age differences in primacy gradients implied by the four–component model. The top panel shows the primacy gradients for older and younger adults. For each serial position, *i*, the gradients give the value of ϕ_i by which the strengths of newly formed context–to–feature associations are multiplied. The bottom panel shows the younger–older difference scores. The gradients were derived by entering the ϕ_s and ϕ_d parameter values form the younger adult model and the four–component older adult model in to: $\phi_i = \phi_s e^{-\phi_d(i-1)} + 1$ (Equation 6, Appendix B).

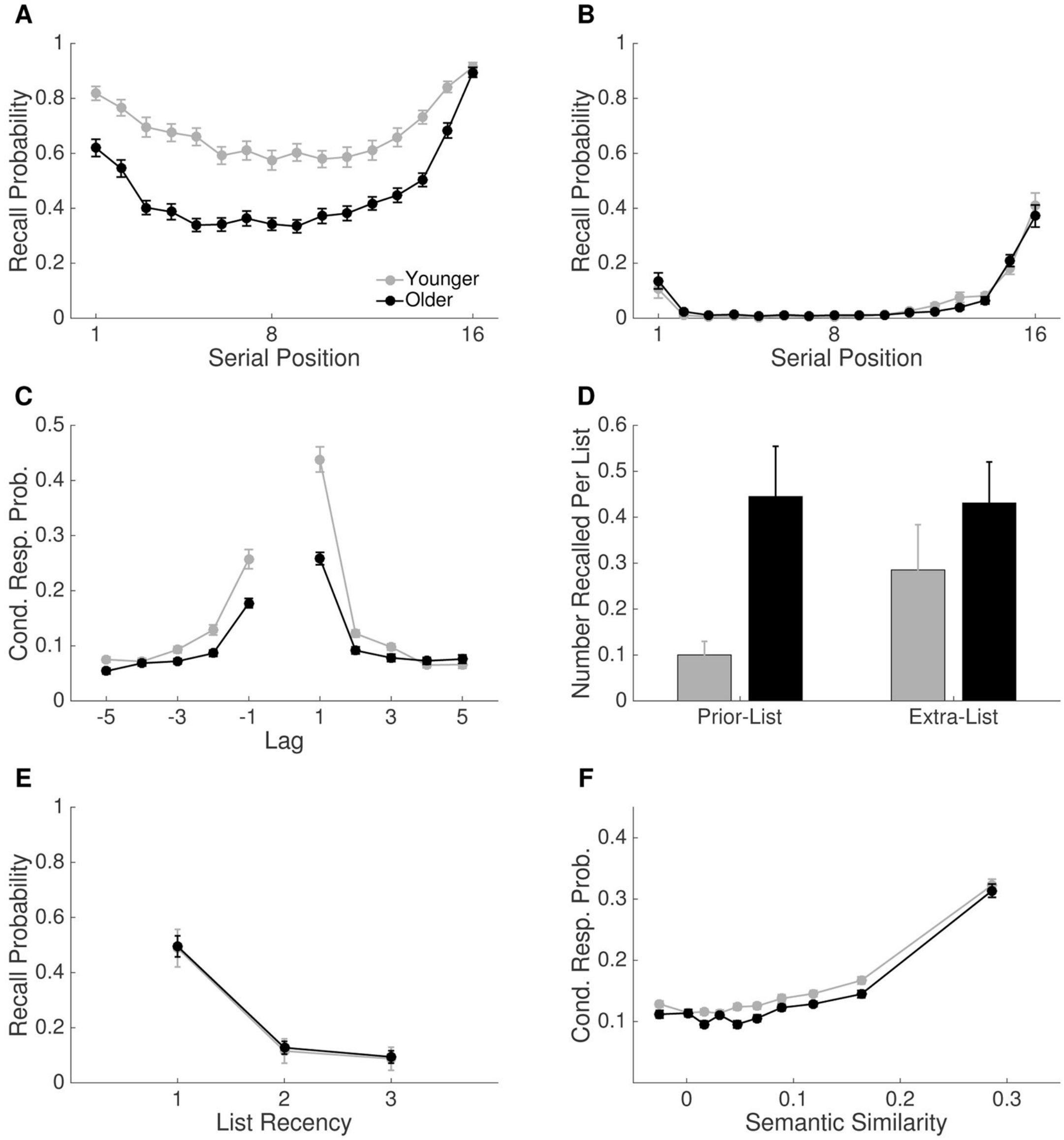

Figure 19.

The Free Recall Aging Pattern in the Penn Electrophysiology of Encoding and Retrieval Study: Serial position curve (A), probability first recall function (B), lag-conditional response probability function (C), Prior–list and Extra–list intrusions (D), Prior–list intrusion recency effect (E), and semantic-conditional response probability function (F). Data are from 38 Older adults and a random subsample of 38 Younger adults that do not differ (α = .15) from the full sample on these data points; see text for full details. Error bars represent one standard error of the mean.

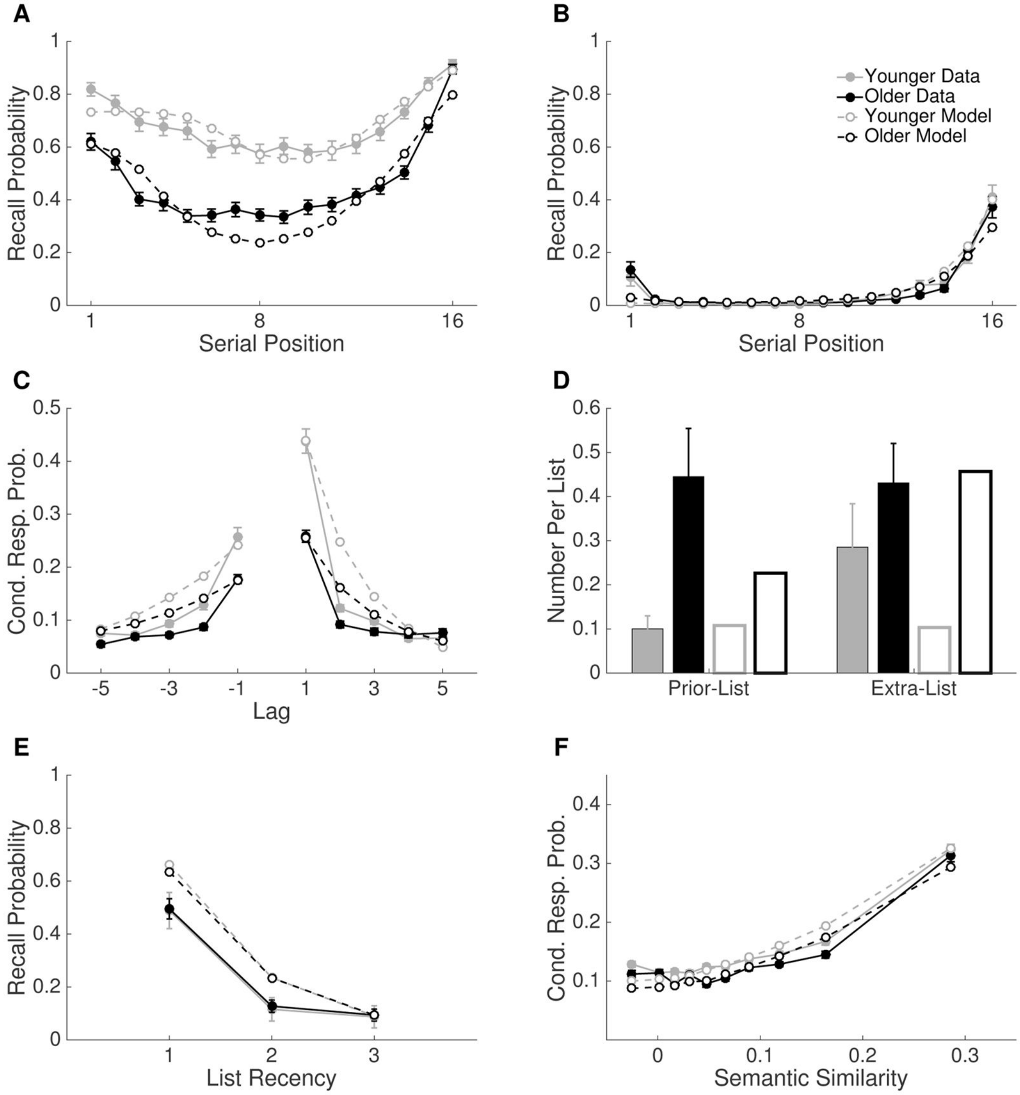

Figure 20.

Fit of the four–component model to data from the Penn Electrophysiology of Encoding and Retrieval Study. Data are from 38 Older adults and a random subsample of 38 Younger adults that do not differ (α = .15) from the full sample on these data points; see text for full details. In fitting the younger adult data, all parameters were allowed to vary. In fitting the older adult data, parameters implicated by the four–component model were allowed to vary and all others were fixed at the best–fitting younger adult values. The best–fitting simulated data is shown along with the actual data from both older and younger adults. Black lines or bars are used for older adults. Grey lines or bars are used for younger adults. Solid lines with filled symbols or filled bars are used for participant data. Broken lines with open symbols or unfilled bars are used for model simulations. Error bars represent one standard error of the mean.

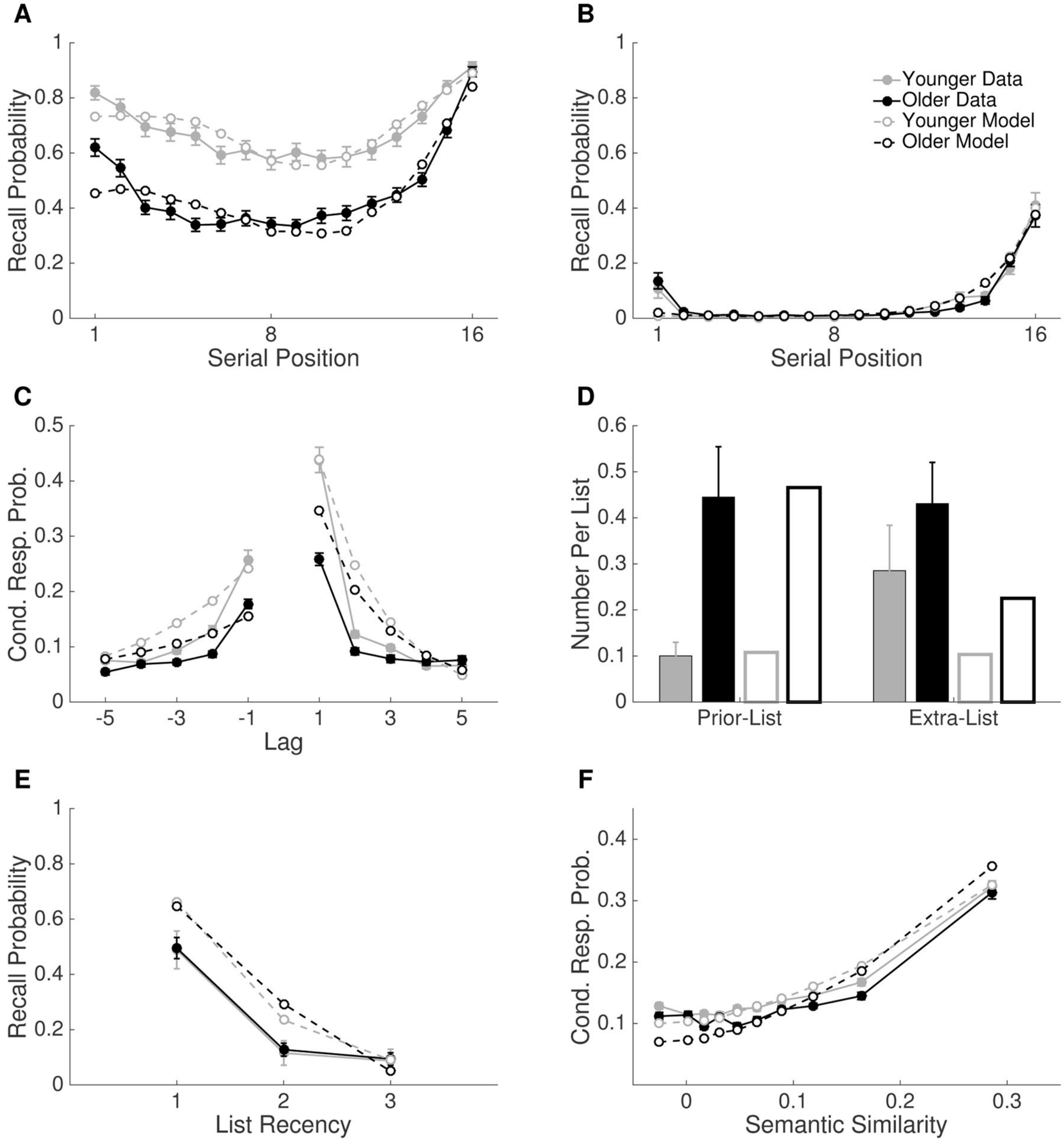

Figure 21.

Fit of the of the combined aging theories to data from the Penn Electrophysiology of Encoding and Retrieval Study. Data are from 38 Older adults and a random subsample of 38 Younger adults that do not differ (α = .15) from the full sample on these data points; see text for full details. In fitting the younger adult data, all parameters were allowed to vary. In fitting the older adult data, parameters implicated by the ADH, IDH, and CSH were allowed to vary and all others were fixed at the best–fitting younger adult values. The best–fitting simulated data is shown along with the actual data from both older and younger adults. Black lines or bars are used for older adults. Grey lines or bars are used for younger adults. Solid lines with filled symbols or filled bars are used for participant data. Broken lines with open symbols or unfilled bars are used for model simulations. Error bars represent one standard error of the mean

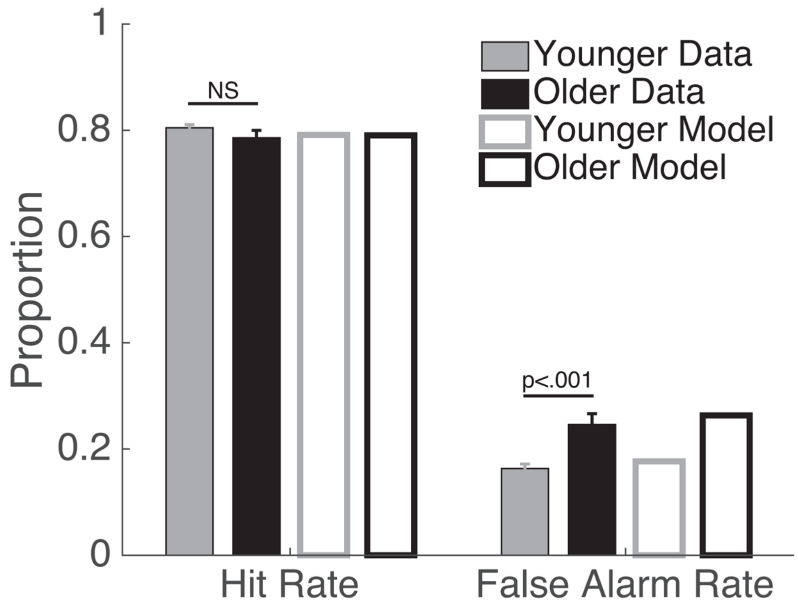

Figure 22.
Fit of the four–component model to recognition data from Kahana et al. (2002). Hit rate is the percentage of old probes correctly identified as being previously studied. false alarm rate is the percentage of new items correctly rejected as lures. The best–fitting simulated data is shown along with the actual data from both older and younger adults. Black bars are used for older adults. Grey bars are used for younger adults. Filled bars are used for participant data. Unfilled bars are used for model simulations. Error bars represent one standard error of the mean.

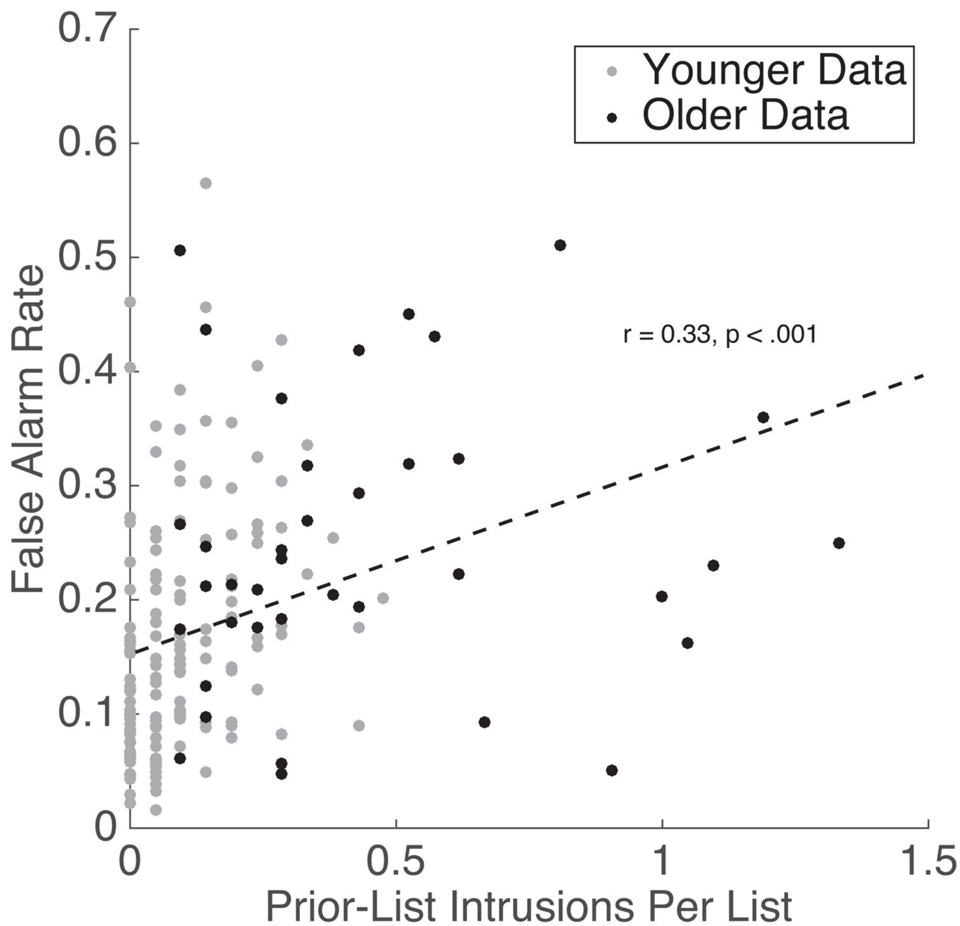

Figure 23.

Confirming a prediction of the four–component model, intrusion rates in free recall correlate positively with false alarm rates in recognition. The solid line is a least squares regression line computed on the combined older and younger adult data from the PEERS study. Black dots are used for older adults. Grey dots are used for younger adults.

Table 1

BIC values for simulation of all effects simultaneously

Theory	Figure	BIC
Fit to Kahana et al. (2002) Data		
Full Model	6	−251
ADH Weak Associations	12A	−185
ADH Noisy Associations	12B	−172
ADH Weak & Noisy	12C	−176
IDH Drift Rate	13C	−185
IDH List Isolation	13A	−182
IDH Retrieval Editing	13B	−180
CSH Slowed Encoding	14A	−185
CSH Slowed Retrieval	14B	−183
CSH Both Slowed	14C	−201
All Theories Combined	15	−223
Four–Component Model	17	−230
Fit to PEERS Data		
All Theories Combined	21	−289
Four–Component Model	20	−297

The figure column gives the figure in which the simulations are shown.