An Embedded Computational Framework of Age-Related Memory Change

Dominic Guitard[1], J. Nick Reid[2], Nathaniel R. Greene[3], Jean Saint-Aubin[4], Randall K. Jamieson[5]

[1] Cardiff University, Cardiff, Wales, United Kingdom

[2] University of Northern British Columbia, Prince George, BC, Canada

[3] Villanova University, Villanova, PA, United-States

[4] Université de Moncton, Moncton, NB, Canada

[5] University of Manitoba, Winnipeg, MB, Canada

Corresponding author information:

Dominic Guitard, School of Psychology, Cardiff University, Tower Building, 70 Park Place, Cardiff, CF10 3AT, Email: guitardd@cardiff.ac.uk

Acknowledgements

This work was supported by an Experimental Psychology Society Small Grant to Dominic Guitard and while working on this manuscript J. Nick Reid, Jean Saint-Aubin, and Randall K. Jamieson were supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC), and Nathaniel R. Greene acknowledges support from National Institutes of Health (NIH) grant F32AG087646.

Abstract

Research on human memory has expanded rapidly over the past decades, leading to the development of numerous models and experimental paradigms. However, this growth has also resulted in increasing specialization, with limited integration across models, tasks, age groups, and domains. Here, we present the Embedded Computational Framework of Memory (eCFM), a simple yet flexible computational model that integrates encoding, storage, retrieval, and decision processes with structured representations that capture visual, orthographic, phonological, and semantic information. We tested the generalizability of the eCFM across 11 experiments involving approximately 40 younger (18 to 25 years) and 40 older adults (65 to 80 years) per experiment. In Experiments 1 to 4, participants completed order reconstruction tasks using words or images varying in semantic, phonological, orthographic, or visual similarity. Experiments 5 to 7 tested serial recall using the same materials. In Experiments 8 to 11, participants performed old/new recognition tests with studied and unstudied items with similar representational forms. Results showed age-related differences in serial recall, smaller differences in order reconstruction, and none in recognition. Across all tasks and domains, participants consistently showed high false memory rates for related foils, with no age-related differences. The eCFM accounted for all main findings and captured age-related differences, and the lack thereof, by manipulating encoding strength and position discriminability. Overall, the model represents a step toward a general account of human memory and highlights the potential and value of integrating traditionally separate areas of memory research.

Keywords: serial recall; order reconstruction; recognition; false memory; aging; computational modelling.

Transparency and Openness Statement

All experiments were theoretically guided by the Embedded Computational Framework of Memory (eCFM) but were not preregistered. We report all sample sizes, and all simulations were conducted as described in the manuscript. The stimuli used in each experiment are detailed in the text, and all stimuli, analyses, vectors, and simulation codes are available on the Open Science Framework (OSF) page associated with this manuscript.

An Embedded Computational Framework of Age-Related Memory Change

Cognitive psychology is the scientific study of how the mind processes sensory information from our environment to encode, maintain, manipulate, transform, and use that information in all mental activities (Neisser, 1967). Cognition is omnipresent in our lives and plays an essential role in all our activities, from speaking to loved ones to cooking a favorite meal. Given the complexity of cognition, carefully controlled experimental methods help to distill this complexity into more basic, constituent processes. The ultimate goal of cognitive psychology is to develop a set of general principles or a unifying framework that can explain how the mind operates in all its forms (Surprenant & Neath, 2009; Kahana & Wagner, 2024). With such a framework, we can better understand the mechanisms of cognition, identify its limitations, and explore how deviations from typical functioning arise, for example, in the context of adult aging where such deviations are often the norm (Salthouse, 2016).

Despite this overarching goal, the field has often shifted its focus toward the specific rather than the general. As Roediger (2008) noted in his review of over a century of research, we have yet to establish general laws or frameworks of memory that hold across variations in materials, encoding conditions, and test procedures. This difficulty may reflect the influence of prevailing research traditions rather than the impossibility of such a goal. Researchers have often become drawn to competing perspectives and narrowly defined paradigms, frequently using small sets of experiments to support opposing theoretical positions. This narrowing of focus is also found in cognitive aging research (Salthouse, 2000). There, efforts to identify general principles of age-related memory (or other cognitive) change have been slowed by the prevailing approach of studying these changes in isolation on singular tasks (cf., Benjamin, 2010). While this work has yielded valuable insights, it has also contributed to fragmentation within the field of cognitive psychology generally, and within cognitive aging more specifically, at times diverting attention from the broader objective of developing a comprehensive account of cognition (Oberauer & Lewandowsky, 2019). As Newell (1973) famously argued, cognitive psychology has lost sight of its most important challenge and should strive instead to develop integrative and computationally explicit theories of cognition. Encouragingly, a growing number of researchers are working in this direction, aiming to develop broader computational accounts of cognition (see Jamieson et al., 2022, for a review). Despite these efforts, the path toward a truly comprehensive framework remains incomplete.

Here, we propose a path forward by demonstrating how a common computational modelling framework – the embedded computational framework of memory (eCFM; Guitard et al., 2025a, 2025b, 2025c) – can identify shared processes underlying performance across a diverse set of tasks and age groups. Although our focus in this review is on illuminating shared principles that can jointly explain age-related stability and differences across tasks of memory that have been largely studied separately, the approach we take holds considerable promise for generalizing the same set of principles to other domains of cognition. This is not to undersell the importance of bridging disparate areas of memory research as a first step toward a more comprehensive framework for studying cognition broadly. Research on human memory constitutes an excellent example of the kind of fragmentary approach to studying cognition that Newell (1973) bemoaned. Much like musicians who specialize in a single instrument, memory researchers often specialize in single tasks (e.g., free or serial recall, recognition), domains (e.g., working memory, long-term memory), or representational modalities (e.g., visual memory, semantic memory, phonological memory). Memory models have also typically been developed independently with minimal integration or dialogue between these distinct research traditions. For example, some models have been designed to account for recognition (see Humphreys et al., 2024, for a review), while others have focused on recall (see Hurlstone, 2024; Hurlstone et al., 2014, Howard & Kahana, 2002; Polyn et al., 2009; Raaijmakers & Shiffrin, 1980, 1981 for reviews). While this specialization has produced many benchmark findings (Oberauer et al., 2018), integration across empirical research and theoretical frameworks remains limited (see for reviews and notable exceptions; Brown et al., 2007; Hintzman, 1986; Hurlstone et al., 2014; Hedayati et al., 2022; Kahana, 2020; Jamieson et al., 2022; Nosofsky, 1988; Logan; 1988; Oberauer & Lin, 2024; Spens & Burgess, 2024).

Specialization undoubtedly provides valuable and deep insights into specific memory processes, which are essential for advancing our broader understanding and refining memory models. Nonetheless, consistent with Newell's (1973) perspective, we argue that the silos created through such specialization must now be bridged to reveal common principles and develop a more comprehensive account of human memory. In this review, we embrace Newell's (1973) recommendation and take a step toward building bridges across memory silos by conducting a large systematic empirical and computational demonstration involving both younger and older adults across several tasks and materials. We intend to show how our recently proposed eCFM (Guitard et al., 2025a, 2025b, 2025c), a simple memory model that integrates processes typically associated with episodic and semantic memory, can offer a general account of memory performance across multiple task types, including order reconstruction, serial recall, and recognition; across multiple materials, including visually, orthographically, phonologically, and semantically related information; and across individuals of different age groups who vary in their mnemonic capabilities.

Précis of the Present Review

Our review aims to show that a common set of processes can explain performance on a wide array of tasks of learning and memory that are typically studied in separate literatures.

We formalize our assumptions about how these unobservable cognitive processes give rise to behavioral outcomes (e.g., true and false temporal order reconstruction, recall, or recognition) through a computational model (eCFM) that integrates instance theories of episodic memory (Hintzman, 1986) with structured semantic, orthographic, phonological, or visual representations derived from established techniques in models of natural language processing (e.g., Landauer & Dumais, 1997). We build on recent applications of eCFM to explain serial recall and order reconstruction in young adults (Guitard et al., 2025a, 2025b, 2025c) by testing the model's generalizability to other common memory tasks (recognition) and to older adults, who are more prone to memory errors (for recent reviews, see Greene & Guitard, Naveh-Benjamin, 2025; Greene & Naveh-Benjamin, 2023; Naveh-Benjamin & Cowan, 2023). These extensions not only test the generalizability of eCFM as a potential general model of memory but also constitute one of few attempts to apply a common computational modeling framework to cognitive aging data across a diverse set of tasks. Computational modeling remains an underutilized tool for understanding age differences in memory, despite its potential to identify complex and interactive mechanisms contributing to these differences (Benjamin, 2010; Healey & Kahana, 2016; Neath, 1999; Surprenant et al., 2006; cf., Greene et al., 2025; Salthouse, 1988). Here, we aim to show that the same underlying processes that can account for memory successes and failures of younger adults can also explain older adults' true and false memories on various tasks of memory, and that age differences in memory can be explained through variations in how these processes are realized.

In what follows, we first introduce the eCFM model, explaining its basic representational and processing assumptions. We review previous applications of the model to serial recall and order reconstruction tasks before describing how we intend to extend the model to recognition. We next revisit the broader theoretical implications of applying the eCFM to cognitive aging data across these various tasks. We then devote a substantial portion of the review to formally testing the eCFM's ability to explain true and false memory in young and older adults. We present 11 novel experiments testing order reconstruction (Experiments 1-4), serial recall (Experiments 5-7), and item recognition (Experiments 8-11), for lists of semantically (Experiments 1, 5, and 8), phonologically (Experiments 2, 6, and 9), orthographically (Experiments 3, 7, and 10) or visually (Experiments 4 and 11) similar items. We apply the eCFM to the observed data of young and older adults in each experiment to test its ability to reproduce these data and to identify processes accounting for both preserved and impaired memory functions associated with normal aging. Finally, we summarize the new theoretical insights afforded by this endeavor, situating our findings in broader conversations about memory theory and cognitive aging.

The Embedded Computational Framework of Memory

The eCFM is a computational memory model that combines the encoding, storage, retrieval, and decision processes of MINERVA 2 (Arndt & Hirshman, 1998; Hintzman, 1984, 1986) with a lexicon (long-term memory) derived from distributional semantic and computational linguistic techniques. This lexicon can represent orthographic, phonological, or semantic relationships between both studied and unstudied items (Guitard et al., 2025a, 2025b, 2025c; Reid et al., 2025). Incorporating structured representations not only aligns the model with contemporary trends in computational approaches of recognition (Chang et al., 2025; Jamieson et al., 2018; Johns & Jones, 2010; Johns et al., 2012; Jones & Mewhort, 2007; Osth & Zhang, 2023; Reid & Jamieson, 2023; Reid et al., 2025) and recall (Kimball et al., 2007; Lohnas, 2024; Mewhort et al., 2018; Polyn et al., 2009; Sirotin et al., 2005), but also enhances its ability to capture memory phenomena at the item level (e.g., recalling individual words from a list or falsely recalling a similar word) – an idea that Jamieson and Mewhort (2010) examined in the context of implicit learning and that they named Dienes'

Dictum after work highlighting the opportunities of item specific prediction has for investigating the completeness and precision of cognitive models (Dienes, 1992).

Initially, the eCFM was designed to explain performance in two specific tasks: order reconstruction, where participants reconstruct the original sequence of a presented list when items are re-presented at test, and serial recall, in which participants recall items in their original order without re-presentation (Guitard et al., 2025a). Early evaluations demonstrated that the model successfully captured the semantic relatedness effect, reflected in superior recall performance for lists of semantically related words (e.g., car, train, plane) compared to lists of semantically unrelated words (e.g., car, dog, banana; see Neath et al., 2022, for a review), across a range of manipulations including, list composition, task difficulty, and presentation rate. Building on these foundations, we expanded the eCFM to address extralist recall errors where participants recall items not previously studied (false memories), across lists comprised of orthographically, phonologically, or semantically related or unrelated items (Guitard et al., 2025b, 2025c). With the embedded representation, the model was able to predict memory performance at the item-level across multiple representational domains, reinforcing the claim that a simple model with structured representations can bridge distinct areas of memory research typically studied separately.

Nevertheless, our initial investigations concentrated on a restricted set of tasks (serial order reconstruction and serial recall) within younger adult populations to simplify model development. However, given its foundation in the versatile MINERVA 2 architecture, eCFM inherently possesses the flexibility to accommodate a broader spectrum of memory tasks (see Jamieson et al., 2022 for a review). This potential of the eCFM to generalize the same formalized principles across multiple tasks of learning and memory also makes it a valuable model for uncovering joint mechanisms accounting for adult age differences in performance across these varied tasks, a point we will soon revisit. Despite this theoretical versatility, the empirical and computational validation of the model's broader applicability remains incomplete.

Therefore, the primary aim of the present work is to explicitly address these limitations by systematically extending and evaluating the generalizability of the eCFM. To establish a comprehensive model of memory, it is essential to account not only for benchmark findings such as true and false memories, but also to demonstrate robust predictive capability across a variety of memory tasks, stimulus types, and age groups. In the following sections, we briefly summarize the model's representational structures and mechanisms for serial recall and serial order reconstruction, before outlining the modifications required to account for recognition. First, however, we justify our decision to extend the model specifically to recognition, a standard and widely used memory task.

On Extending eCFM to Recognition

There are several motivations for extending the model to recognition memory. First, recognition constitutes one of the most widely used tasks for studying memory in both the mainstream and cognitive aging literatures (see, for example, meta-analyses by Fraundorf et al., 2019; Rhodes et al., 2019). Thus, the success of the eCFM as a general model of memory rests on its ability to explain data not only from more obscure tasks like order reconstruction, but also from more common tasks (e.g., serial recall and recognition) according to the same basic theoretical assumptions and model expressions.

Second, several computational models have been advanced specifically to explain recognition (for reviews, see Clark & Gronlund, 1996; Osth & Dennis, 2024), including MINERVA 2 (Hintzman, 1984, 1986). Given that the eCFM builds on MINERVA 2, it might seem obvious that it should be able to explain recognition. However, the eCFM's structured representations could address a major shortcoming of models that represent items with engineered feature vectors, the convention in MINERVA 2 (but see Jamieson et al., 2018; Reid et al., 2025). The use of engineered feature vectors (arbitrary vectors typically chosen to ensure orthogonality among items) poses substantial challenge for a model's ability to predict item-level effects of lure similarity across different and co-occurring representational domains (semantic, orthographic, phonological, or visual). Although one could design the feature vectors to maximize the ability to simulate these effects, this affords potentially too many experimenter degrees of freedom in the engineering of item representations that makes the act of modelling these effects almost tautological in nature (Johns, Jamieson, & Jones, 2020). Using structured representations removes these experimenter degrees of freedom and equips the model with a principled way of measuring item-level effects consistently across representational domains. Because the eCFM builds its lexical representations from instances of item co-occurrences in largescale, real-world data, it can capture variations in item similarities with a high degree of precision (cf., Johns & Jones, 2010).

Third, age differences in item recognition are considerably smaller in magnitude than those obtained in recall procedures (Craik & McDowd, 1987; Danckert & Craik, 2013; Rhodes et al., 2019). Although such differences are typically non-zero, some studies report no or only negligible age differences in item recognition (e.g., Naveh-Benjamin et al., 2003). Thus, for the eCFM to be a viable model for understanding the processes underlying age changes in memory, it must be able to account for memory functions that are typically greatly impaired with aging (e.g., recall) and those that are relatively more preserved (e.g., recognition). With these justifications in mind, we turn now to presenting the nuts and bolts of the eCFM.

Representations in the eCFM

A fundamental issue for any computational memory model is how sensory input from the external environment is translated into internal representations. These representational schemes critically determine the model's ability to simulate observed memory behavior with precision. For a model to successfully account for order memory, it must incorporate two distinct components: item-level information, which encodes individual stimuli (such as words, nonwords, or images), and order-level information, which encodes the sequential position of each item within a list. These two types of representations together underpin performance in serial recall and serial order reconstruction tasks (Guitard et al., 2021, 2022; Guitard & Cowan, 2023; Majerus, 2019). These same underlying representations play a role in recognition; however, order information appears to have limited influence on recognition, with item representations playing a more prominent role (e.g., Gionet et al., 2024). In this section, we first outline the approach we used for item representations, and then briefly describe the representation of order.

Item Representations

The eCFM represents items (typically words) as *n*-dimensional vectors—300 dimensions in all our simulations, based on our previous work (Guitard et al., 2025a, 2025b, 2025c)—with each dimension reflecting aspects of the item's feature (i.e., visual features) or lexical content (i.e., its semantic, phonological, or orthographic). These underlying representations are derived from specialized techniques, briefly described below, which endow the eCFM with realistic structured representations that capture both blunt and subtle effects of inter-item similarity on recall or recognition.

Semantic Representation. We derived semantic representations using Latent Semantic Analysis (LSA; Landauer & Dumais, 1997), an established method that captures word meaning based on statistical co-occurrences in large text corpora. We used the

Touchstone Applied Science Associates (TASA) corpus as our source corpus, which comprises a diverse selection of text samples appropriate for modeling general English semantics. First, a word-by-document occurrence matrix was constructed, in which rows represented individual words from the corpus and columns represented text documents. The frequency counts of each word's occurrence within each document were recorded in the matrix. Subsequently, we applied log-entropy weighting to these raw counts to emphasize informative word-document associations and diminish the influence of overly frequent, less informative terms, as is typically done in LSA (Landauer & Dumais, 1997; Martin & Berry, 2007).

To reduce the dimensionality of the resulting weighted matrix, we employed Singular Value Decomposition (SVD), a widely used dimension-reduction technique in semantic modeling (Bullinaria & Levy, 2012; Levy, Goldberg, & Dagan, 2015). SVD decomposes the original matrix into three component matrices (U, S, V), where matrix U represents orthogonal semantic dimensions of the words, and S is a diagonal matrix containing the singular values reflecting the variance accounted for by each dimension in order from largest to smallest. A reduced matrix can then be computed by retaining the *n* largest singular values as well as the corresponding rows in the U matrix and multiplying these matrices, U_nS_n. Landauer and Dumais (1997) found that LSA performs best on synonym selection tasks when the first 300 singular values are retained, which is also what we have applied in our prior studies (Reid et al., 2025; Guitard et al., 2025b) and what we apply in the current simulations. The resulting vectors have been shown to effectively predict semantic similarity judgments, semantic priming effects, and recognition memory performance (Landauer & Dumais, 1997; Landauer et al., 1998; Reid & Jamieson, 2023).

While we adopted LSA because of its demonstrated effectiveness in prior modeling efforts (Reid & Jamieson, 2023), we acknowledge alternative semantic models, such as word embeddings (Mikolov et al., 2013; Bojanowski et al., 2017), HAL-type models (Lund & Burgess, 1996; Durda & Buchanan, 2007; Rohde et al., 2006), random indexing models (Kanerva et al., 2000; Jones & Mewhort, 2007; Sahlgren et al., 2008), or word-by-word models (Bullinaria & Levy, 2007, 2012; Johns et al., 2019), could similarly contribute to understanding recognition memory phenomena (see also McRae et al., 2005; De Deyne et al., 2019; Jamieson et al., 2018).

Phonological Representation. Phonological vectors were constructed using Parrish's (2017) method of representing "interleaved bigrams of phonetic features". This method encodes phonological information by representing interacting sound features of adjacent phonemes in the word, explicitly capturing co-articulatory properties of spoken words. To illustrate the approach, consider the word *chip*, comprising the phonemes /CH/, /IH/, and /P/ (transcribed using Arpabet conventions). Each phoneme contains distinctive phonetic features: for instance, /CH/ is characterized by {affricate, postalveolar, voiceless}, /IH/ by {front, high, unrounded, vowel}, and /P/ by {stop, bilabial, voiceless}. The phonological representation involves pairing each phonetic feature of a given phoneme with each feature of its adjacent phoneme. This results in feature pairs, such as {(affricate, front), (affricate, high), (affricate, unrounded), (affricate, vowel), (postalveolar, front), ...}. By representing pairs of features rather than individual features, this captures the fact that the sound features of preceding and subsequent phonemes influence the pronunciation of the current phoneme (i.e., co-articulation; Connine & Darnieder, 2009; Daniloff & Hammarberg, 1973; Fowler, 1980; Fowler & Saltzman, 1993; Laplante et al., 2023). To mark word boundaries, two pseudophonemes, BEG and END, are incorporated, each containing a single distinctive feature {beg} and {end}, respectively.

We constructed a word-by-feature-pair matrix based on this phonological encoding scheme, recording the counts of each feature pair for all words in the corpus. Following

Parrish (2017), the matrix was then mean-centered (columns adjusted to have a mean of zero), and dimensionality was reduced via SVD to yield phonological vectors with 300 dimensions, ensuring an equal number of dimensions with the semantic vectors. These phonological vectors have demonstrated alignment with human phonetic similarity ratings (Parrish, 2017) and have proven effective in modeling memory phenomena (Guitard et al., 2025b; Reid et al., 2025).

Orthographic Representation. We derived orthographic vectors, which represent the written structures or spelling patterns of words, using an *open-bigram* coding approach inspired by the SERIOL and SERIOL2 models (Whitney, 2001; Whitney & Marton, 2013). In this framework, orthographic information is encoded as pairs of letters (bigrams) that occur within a small positional window. Specifically, pairs of letters with zero, one, or two intervening letters are represented, with activation weights inversely proportional to their separation: adjacent letters (no intervening letters) receive a weight of 1.0, pairs separated by one letter receive 0.7, and pairs separated by two letters receive 0.5.

For example, the word *FISH* is represented by the bigrams *FI*, *IS*, and *SH* (adjacent, weight = 1.0); *FS* and *IH* (one-letter separation, weight = 0.7); and *FH* (two-letter separation, weight = 0.5). Together, these weighted bigrams define the orthographic representation of the word, capturing the spatial relationships amongst its letters independently of their sounds. We also included special edge markers to encode word boundaries (denoted by "*"). In SERIOL2 (Whitney & Marton, 2013), letters are paired with the edge markers to create edge bigrams. These bigrams receive graded activation based on proximity of the letter to the respective word boundary. For instance, the word *FISH* would activate the edge bigrams **F* and *H** with an activation weight of 1.0, **I* and *S** with a weight of 0.7, and **S* and *I** with a weight of 0.5. Open-bigram schemes effectively capture transposed-letter priming effects and the edge bigrams help to emphasize the importance of the first and last letters of the word, which play a relatively larger role in word recognition (Whitney & Marton, 2013)

The orthographic bigram activations were compiled into a word-by-bigram matrix, with cells reflecting the summed weights of bigram activations for each word. We applied SVD to reduce dimensionality to a 300-dimensional orthographic vectors, aligning the dimensionality of our orthographic representations with those of our semantic and phonological representations. This orthographic scheme has effectively captured orthographic similarity effects on recognition memory in prior studies (Reid et al., 2023; see also Zhang & Osth, 2024, for further evidence using a different method of open-bigram coding to model recognition memory).

While alternative orthographic representation schemes exist (Hannagan et al., 2011; Cox et al., 2011; Chang et al., 2025), our choice of open-bigram coding was motivated by its demonstrated efficacy in previous modeling work (Reid et al., 2023, 2025; Zhang & Osth, 2024).

Visual Representation. In several of the to-be-modelled experiments of the present study, we aimed to extend the model beyond verbal materials by using abstract images. This required us to develop a method for representing visual information within our existing framework. To create a vector representation for an image, we first extracted the color values for each pixel in the image. Each pixel in an image has three values for the colors red, green, and blue that range from values of 0 (color is not present) to 255 (color is fully present). The pixels are typically represented in a three-dimensional array, where one dimension corresponds to the vertical position of the pixel, one dimension corresponds to the horizontal position, and one dimension corresponds to the color (red, green, or blue). To create a vector for the image, we flattened the array into a long vector where each value corresponded to a certain color at a certain horizontal and vertical location. For instance, for a 50 × 50 resolution image, the dimensionality of the vector would be equal to 7,500 (50 × 50 × 3). The flattened image vectors were combined into a matrix where each row of the matrix corresponded to a particular image. The column values in the matrix were mean-centered before dimensionality reduction was performed.

To reduce the dimensionality of the matrix, we used random projection, a simple but popular method of dimension reduction that has been applied to represent both images (Bingham & Mannila, 2001) and the meanings of words in distributional semantic models (Sahlgren et al., 2008; Jones & Mewhort, 2007). In this method, each visual dimension (corresponding to columns in the matrix) is assigned a randomly generated index vector with a much smaller dimensionality than in the original vectors. Here, we created the index vectors by sampling values from a Gaussian distribution with a mean of zero and a standard deviation of $1/\sqrt{d}$ where d is the number of dimensions in the index vector (Jones & Mewhort, 2007; see Achlioptas, 2003; Li et al., 2006, for other random indexing methods). The vectors are combined into a matrix, $R_{k\,X\,d}$, where k is equal to the number of dimensions in the original vectors. The reduced matrix, X^{RP}, can then be computed by multiplying the original matrix, X, by the matrix of index vectors, R:

$$X^{RP}_{n\times d} = X_{n\times k}R_{k\times d} \qquad [1]$$

resulting in a matrix of n (the number of images) by d dimensions. For our simulations, we set d to 300. Although random projection is a fairly simple method of dimension reduction, it maintains the Euclidean distances between the vectors in the original matrix quite well in the reduced matrix, and its accuracy is comparable to other more computationally costly methods such as Principal Component Analysis (Bingham & Mannila, 2001).

Order Representations

To encode order information, we used a feature-dissimilarity vector approach commonly employed in item-independent context models (e.g., Logan & Cox, 2023; Osth & Hurlstone, 2023). Each serial position is assigned a high-dimensional vector of 300 dimensions, with successive positions generated by probabilistically modifying features of the previous vector. Specifically, we first generate a random n-dimensional vector for the first serial position, where each dimension is randomly sampled from a $N(0, 1/\sqrt{n})$ distribution. In our simulations, we set n to 300, equating the dimensionality of the order representations to the lexical and visual representations. We then successively generate serial position vectors for the ith item (p_i) by iteratively copying the preceding item's position vector (p_{i-1}) and sampling, with probability d, a new random deviate from the same $N(0, 1/\sqrt{n})$ distribution for each dimension in p_i. The parameter d controls the degree of similarity/dissimilarity between position representations. When d is low (near 0), adjacent positions are highly similar (more features from the previous position vector are retained), increasing intralist errors (recalling a word in the wrong serial position); when d is high (near 1), position vectors become more distinct (less features from the previous position vector are retained and more a resampled), reducing such errors (Guitard et al., 2025a; 2025b; 2025c; Saint-Aubin et al., 2021). In effect, parameter d determines the precision of each item's position representation, as lower values of d blur distinctions among these representations.

Having described the basic representational structure of the model, we turn now to explaining how the eCFM accounts for serial recall and order reconstruction as a prelude to extending the model to account for recognition.

Modeling Serial Recall and Order Reconstruction

Encoding and Storage

In the eCFM, both serial recall and order reconstruction rely on the model's ability to encode and temporarily store the study list as a sequence of memory traces. Each trace consists of an imperfect copy of the item's lexical and order representation. Specifically, the first 300 dimensions of each trace represent the serial position (i.e., the order representation), while the remaining 300 dimensions represent the lexical features of the item (i.e., the item representation). To implement this computationally, we construct a memory matrix, *M*, where each row is a 600-dimensional vector. For a six-item list, *M* is a 6 × 600 matrix, with each row storing one item's order and lexical representations.

Consistent with previous models of serial order memory (e.g., Brown et al., 2000, 2007; Henson, 1998; Page & Norris, 1998; Nairne, 1990), we assume that encoding is not uniform across positions. Items at the beginning of the list are encoded with greater fidelity, reflecting the idea that they benefit from increased attentional resources and greater opportunity for maintenance strategies such as rehearsal (e.g., Bhatarah et al., 2009; Rundus, 1971; Popov & Reder, 2020). For computational simplicity, we assume that the final item is encoded as well as the item that precedes it. This assumption aligns with evidence suggesting that final list items are relatively protected from retroactive interference (e.g., Nairne, 1990; Saint-Aubin et al., 2021) and may benefit from their privileged position at the end of the sequence (e.g., Henson, 1998; Brown et al., 2007).

To incorporate this assumption, we copy each dimension in a trace at serial position *p* with probability L_p.

$$L_p = \begin{cases} L - (\mathrm{p} - 1)\mathrm{g} & , \ p < N \\ L - (\mathrm{p} - 2)\mathrm{g} & , p = N \end{cases} \qquad [2]$$

In Equation 2, L denotes the base-rate learning strength of the first item in the list (i.e., the proportion of dimensions encoded)[1], p refers to the serial position of an item, g reflects the rate at which encoding effectiveness declines across serial positions, and N denotes the total number of items in the study list. In general, each successive item is encoded with slightly less precision than the one before it, at a rate determined by g. An exception is made for the final item in the list, which is assumed to be encoded with the same effectiveness as the next to last item. Within this scheme, L_1 denotes the encoding strength of the first item, L_2 the second item, and so on.

Retrieval and Decision

In the eCFM, retrieval in both serial recall and order reconstruction is guided by cue-based similarity, consistent with the principles of MINERVA 2 (Hintzman, 1984, 1986) and earlier models of serial recall (e.g., Nairne, 1990). Retrieval begins by presenting a cue vector q that corresponds to the intact representation of a given serial position (300 dimensions). This process is currently deterministic, proceeding from the first to the last serial position, in line with the experimental procedure in which participants were cued to recall specific list positions (e.g., 'Recall the first word,' 'Recall the second word,' and so on). This cue interacts in parallel with the memory matrix M, activating each trace in proportion to its similarity with q. The result is an *echo*, a composite vector e that reflects an activation-weighted sum of all traces in memory. These weights are determined by the similarity between the cue and the serial position components stored in each trace. The echo then serves

[1] In essence, one can think of L as a representational fidelity parameter. Higher values of L result in stronger representations of items in the list, as more features are faithfully encoded into the model's memory. Conversely, lower values of L result in impoverished item representations, with fewer features of the item represented in memory. Because L operates on the combined lexical and order representation of the item, it affects the fidelity of both types of representations, in contrast to parameter d, which only affects the precision of order representations.

as a retrieval signal from which the model attempts to recover the corresponding lexical information, represented in the final 300 dimensions of each trace.

To decide which item to recall, the model compares the lexical portion of the echo (the last 300 dimensions) with all entries in its lexicon (for serial recall) or all test items available at retrieval (for order reconstruction). A word is selected if its similarity to the echo exceeds a predefined threshold *T* (see also Johns et al., 2020). If no similarity exceeds this threshold, the model produces an omission. In order reconstruction, the item amongst the presented alternatives (e.g., only the words from the studied list rather than the entirety of memory) with the highest cosine similarity is selected from the available alternatives.

The computational steps of this retrieval process are summarized in Equation 3. For each serial position *p*, the model provides the corresponding cue *q* to memory and retrieves an echo *e* based on its similarity to each of the *m* traces stored in memory. The equation is defined as follows:

$$\mathrm{e} = \sum_{i=1}^{i=m} \left(\frac{\sum_{j=1}^{j=n/2} q_j \times M_{ij}}{\sqrt{\sum_{j=1}^{j=n/2} q_j^2}\sqrt{\sum_{j=1}^{j=n/2} M_{ij}^2}} \right)^{\tau} \times M_{i\bullet} \qquad [3]$$

In this equation, q_j represents the *j*th feature of the cue, M_{ij} is the *j*th feature of the *i*th memory trace, *n/2* refers to the dimensionality of the serial position representation (1:300), and *m* denotes the number of traces in memory (equal to the list length). The parameter τ controls the selectivity of the retrieval process (i.e., which candidates contribute the most to the echo) by scaling the similarities between the cue and each memory trace. Higher values of τ make the retrieval process more selective, favoring only those memory traces with high similarity to the cue. Conversely, lower values of τ make retrieval less selective as even memory traces with only moderate or low degrees of similarity to the cue become active. In all simulations, τ was fixed at 3, consistent with our prior work (Guitard et al., 2025a, 2025b, 2025c), but we consider the implications of changes in τ in a later section (see Recognition).

The resulting echo typically corresponds most strongly to the item presented at the cued position, but it also contains residual activation from neighboring positions due to representational overlap in positional features, governed by d, that can yield intralist errors. Additionally, because lexical representations of related items tend to be similar, echoes can sometimes activate items that were not presented, resulting in extralist errors or "false memories."

After generating the echo, the model computes cosine similarity between the lexical portion of the echo (the final 300 dimensions) and each word in the lexicon (serial recall) or the candidate words at test (order reconstruction). The word with the highest similarity is selected if its similarity exceeds the decision threshold T. If no word exceeds this threshold, the model produces an omission (no word recalled).

To minimize repetition errors, a soft output suppression mechanism is included (see e.g., Hurlstone et al., 2014). Following the recall of a word, its activation value is reduced by a fixed proportion s in subsequent retrieval attempts within the same trial. For example, if *dog* is recalled at position 1 with a similarity of 0.30, and it has a similarity of 0.20 at position 2, applying $s = 0.10$ reduces its activation at position 2 to 0.10. This reduces, but does not eliminate, the likelihood of repetition. For order reconstruction, where each item can be selected only once, this suppression parameter is set to 1 to strictly prevent the reselection of the same item within a trial.

Following Equation 3, the model, like human participants, either produces an item or fails to respond (an omission) at each serial position based on whether the value resulting from Equation 3 is above the decision threshold T. Based on these outputs, we apply the same scoring criteria used in serial order memory to directly compare the model's performance at the item level. If the model recalls the correct item in its original position, the response is scored as correct using a strict criterion. If the model recalls a studied item but places it in the wrong serial position, this is classified as an intralist error, also known as an order error. If the model produces an item that was never studied, the response is considered an extralist error, or false memory.

Modeling Recognition

Thus far, we have described how the eCFM represents information, encodes study lists, retrieves information, and makes decisions for both order reconstruction and serial recall. But how can the model be extended to handle recognition? Somewhat ironically, echoing Newell's (1973) critique, we (the authors of this manuscript) have historically contributed to computational modeling in recognition (e.g., Caplan & Guitard, 2024a, 2024b, 2025; Reid et al., 2023; Reid & Jamieson, 2022, 2023; Spear et al., 2024) and recall (Saint-Aubin et al., 2021, 2024; Cyr et al., 2022; Guitard et al., 2025a, 2025b) largely separately, with limited cross-talk between domains. In the present work, we bridge this gap by extending the eCFM to account for recognition memory, drawing on principles from both domains. Below, we outline the key modifications needed for this extension.

Encoding and Storage

In traditional implementations of MINERVA 2, each item (e.g., word, nonword, image) is encoded and temporarily stored in memory at a fixed probability L, without position-specific encoding. We adopt a similar approach to model recognition in eCFM by setting $g = 0$, which eliminates the decline in encoding effectiveness across serial positions (see Equation 2), thereby ensuring uniform encoding as in previous recognition modeling (see, e.g., Reid et al., 2025; Reid & Jamieson, 2022, 2023; Spear et al., 2024). This decision is motivated by the fact that recognition accuracy remains relatively stable across serial positions for long lists of items, like those to-be-modeled here (see, for example, Experiment 1 of Gionet et al., 2024; cf., Fawcett et al., 2023). Nonetheless, some empirical studies have reported modest serial position effects in recognition tasks for short lists (Neath, 1993) or modest between-condition differences in recognition accuracy at earlier versus late serial positions (e.g., Gionet al., 2024). We therefore retain serial position vectors as an available feature for future modeling, but we render them functionally inert in the current simulations by assigning identical values to the first 300 dimensions of each vector. As a result, these dimensions represent a constant context that does not influence retrieval outcomes.

Thus, for a recognition task involving 40 studied items, the memory matrix *M* consists of 40 rows and 600 columns. The first 300 dimensions (context) are identical across items, and the remaining 300 dimensions encode lexical or item-specific information, as in prior serial recall and order reconstruction.

Retrieval and Decision

The retrieval process in recognition mirrors recall in structure but differs in its target. Rather than reconstructing an echo, the model calculates the *familiarity* or *intensity* evoked by a cue. The cue is the intact vector representation of an item presented at test, whether previously studied or novel, where for consistency the first 300 elements represent the list context (i.e., dimensions 1:300) and the second 300 represent the word (i.e., dimensions 301:600). The formula to calculate familiarity, *f*, is computed as,

$$f = \sum_{i=1}^{i=m} \left(\frac{\sum_{j=\frac{n}{1}+1}^{j=n} q_j \times M_{ij}}{\sqrt{\sum_{j=\frac{n}{2}+1}^{j=n} q_j^2} \sqrt{\sum_{j=\frac{n}{2}+1}^{j=n} M_{ij}^2}} \right)^{\tau} \qquad [4]$$

where, f is the sum of activation in memory elicited by the cue q, q_j is the jth feature of the cue, M_{ij} is the jth feature of trace i in memory, indices $(n/2+1){:}n$ refers to the dimensionality of the lexical representation (elements 301:600), m is the total number of memory traces (equal to the number of studied items), and τ is a scaling parameter that governs the conversion of trace similarity (in parentheses) into trace activation. The higher the value of τ, the more sharply the model favors highly similar traces over less similar ones. Critically, we draw the reader's attention to an instructive comparison of Equations 3 and 4. As one will see, the two equations are nearly identical, except that f is the sum of trace activation computed over the item field (i.e., dimensions 301:600) and **e** is the activation-weighted sum of traces conditional on trace activation computed over the serial position field (i.e., dimensions 1:300).

To convert familiarities into binary "old/new" responses, we compare those against a threshold-based decision rule. Specifically, familiarity values for all test items are rank-ordered, and a percentage-based response threshold is applied. For example, applying a 50 percent threshold would result in classifying half of the test items as "old" and half as "new." In our current study, we selected a threshold of 42.86 percent to precisely align with the proportion of old items included in our recognition test phase (60 old items, 70 new items, and 10 related items) [2]. Further details on this implementation are provided in the recognition experiments section.

[2] It is important to acknowledge that this approach simplifies the decision-making process by assuming a uniform decision bias across all simulated participants. Additionally, the threshold is set after computing familiarity for all items, a procedure that differs from human behavior, where participants evaluate test items sequentially and may adjust their decision criteria dynamically during the test phase. A more realistic method would involve using an absolute familiarity threshold (as discussed by Hintzman, 1988; Chang & Johns, 2023), augmented by randomly varying the threshold for each simulation to reflect inter-individual differences in decision biases. However, interpreting absolute familiarity thresholds is inherently less intuitive. Given that the primary objective of this study is not to precisely replicate individual differences in recognition bias but rather to clearly illustrate the decision process, we opted for a simpler, percentage-based threshold. This choice facilitates ease of interpretation and maintains a clear focus on the core objectives of our investigation, essentially

In summary, the eCFM accounts for recognition by (1) applying uniform encoding across items, rather than position-dependent encoding as in serial order memory, and (2) computing familiarity instead of generating an echo, using this value to determine whether an item is classified as old or new.

Summary

We have presented the eCFM (Guitard et al., 2025a, 2025b, 2025c) and explained how this framework, originally developed for order reconstruction and serial recall, can, in principle, be extended to account for recognition. The parameters of the model and their implications across the three tasks are summarized in Table 1. Although other memory tasks exist, we chose to begin with these three (serial recall, order reconstruction, and recognition) because they are widely studied; represent distinct yet traditionally separate areas within the field of human memory; produce different gradients of age-related impairment (discussed next); and place differing demands on the need to remember items, their orders, or both. Our aim is not to account for every empirical finding across all tasks, but to introduce a common foundation that can be adapted to test predictions across various tasks and promote a coherent more integrated view of memory and how memory applies across different remembering contexts. We turn now to discussing the implications of generalizing the eCFM to explain patterns of age-related differences and similarities across these diverse tasks of memory.

measuring the model's recognition performance under the assumption of a specified and in our case mildly conservative response bias.

Table 1

Model Parameters, Definitions, and Their Application Across Recall, Reconstruction, and Recognition Tasks

Parameter	Definition	Effect	Recall	Reconstruction	Recognition
L	Learning probability: likelihood that a feature is encoded to memory	Lower values reduce the number of features stored in memory	✓	✓	✓
g	Gradient of learning: position-dependent encoding adjustment	Controls the slope of the serial position curve; higher values reduce encoding of later items	✓	✓	✗
d	Context drift: degree of change and thus similarity across context/position vectors	Controls context distinctiveness across items; higher values reduce interference	✓	✓	✗
τ	Sensitivity of retrieval to similarity; translation of similarity into activation	Higher values amplify similarity differences (strong cues dominate); lower values make retrieval more tolerant to weaker matches	✓	✓	✓
T	Decision threshold: minimum similarity required to recall or recognize	Higher values make recall or recognition more conservative (more omissions or fewer old responses)	✓	✓ (fixed at 0)	✓
s	Repetition suppression: reduces the likelihood that the model recalls the same response more than once	Higher values reduce repeated responses; used to avoid repetition errors in Recall and Reconstruction	✓	✓ (fixed at 1)	✗
Response Pool	Set of candidates considered during retrieval	Determines what can be recalled or recognized: full vocabulary in Recall; only test items in Reconstruction and Recognition	Vocabulary	Test Items	Test Items

Explaining Adult Age Differences in Memory with the eCFM

Given eCFM's current success in explaining serial recall and order reconstruction in young adults (Guitard et al., 2025a, 2025b, 2025c), along with its potential to apply the same principles to explain recognition, the eCFM holds great promise for uniting these disparate tasks under a common framework. This makes the eCFM a strong candidate model for explaining the source of adult age differences in memory, given that such differences have been detected to varying degrees on all these tasks. Older adults particularly struggle with remembering the precise order of events (Healey & Kahana, 2016; Kahana et al., 2002), resulting in substantial age differences in serial recall (Golomb et al., 2008; Maylor et al., 1999; Naveh-Benjamin et al., 2007) and in tests of temporal order similar to those used in order reconstruction tasks (e.g., Cabeza et al., 2000). Aging is also associated with more modest but typically non-zero deficits in recognition memory (for meta-analyses, see Fraundorf et al., 2019; Rhodes et al., 2019). Yet, attempts to identify common processes underlying age-related performance differences across these various tasks have been few and far between, owing in part to the comparatively limited application of computational modeling in cognitive aging research.

Prior Computational Modeling Work on Age Differences in Memory

Although we have amassed a rich understanding of which aspects of memory are most susceptible to age-related changes (Greene & Naveh-Benjamin, 2023; Light, 1991; Naveh-Benjamin & Cowan, 2023; Zacks et al., 2000), prevailing theoretical explanations for these changes almost exclusively rely on verbal terminology that is difficult to map to specific processes of impairment. Theories attributing age-related memory change to impairments in attention (Craik & Byrd, 1982), inhibition (Hasher & Zacks, 1988), associative binding (Naveh-Benjamin, 2000), recollection (Jennings & Jacoby, 1997), or other processes rarely formalize their assumptions about *how* these underlying processes give rise to observed differences in memory performance. This relative lack of formalization affords great flexibility for these theories (Oberauer & Lewandowsky, 2019) while leaving unaddressed important questions about the complex ways in which unobservable cognitive processes effectuate age-related memory change. As but one example, these theories are typically vague about whether the proposed process(es) of impairment occurs during encoding or at the time of retrieval. In contrast to verbal theories, computational modeling explicates the ways in which an underlying process gives rise to behavior through precise mathematical equations that can be evaluated and falsified against observed data.

The importance of this issue is not lost on cognitive aging researchers (Salthouse, 1988), yet efforts to explain age differences in memory with formal computational models have typically been constrained to single tasks (cf., Greene et al., 2025). There have been isolated attempts to computationally model age differences in serial recall (Neath & Surprenant, 2007; Maylor et al., 1999; Surprenant et al., 2006), two-choice discrimination (e.g., old/new recognition; Bayen et al., 2000; Ratcliff et al., 2004), source/context recognition (Benjamin, 2010), and associative recognition (Buchler & Reder, 2007; Darby & Sederberg, 2022; Li et al., 2005; Stephens & Overman, 2018) with a panoply of models that make different processing assumptions. For example, age differences in serial recall have separately been accounted for with Nairne's (1990) Feature Model (Neath & Surprenant, 2007), with the OSCillator-based Associative Recall (OSCAR) model (Brown et al., 2000; Maylor et al., 1999), and with the Scale-Invariant Memory, Perception, and Learning (SIMPLE) model (Brown et al., 2007; Surprenant et al., 2006). These models advance competing explanations for why age differences in serial recall emerge. Both the Feature Model and SIMPLE propose that older adults encode stimulus features less efficiently, resulting in impoverished memory representations that impair serial recall accuracy.

However, unlike the Feature Model, SIMPLE additionally assumes that older adults' degraded item representations reduce the distinctiveness of competing stimuli in multidimensional space, leading to more confusability among recall candidates. Meanwhile, OSCAR attributes age differences in serial recall to less precise temporal-context signals that blur older adults' perception of which items were mapped to specific context states.

A similar diversity of model-based explanations has been advanced to account for age differences in recognition memory, though most of these explanations converge on the general suggestion that noisier or impoverished memory representations lead to age-related impairments in discriminating old from new experiences. For example, Benjamin's (2010, Benjamin et al., 2012) Density of Representations Yields Age-Related Deficits (DRYAD) model attributes age differences in recognition accuracy to a global reduction in memory fidelity in older adults. This global deficit manifests especially on tests of context memory because DRYAD assumes that context features are more sparsely represented than item features. Similarly, Stephens and Overman (2018) simulated older adults' reduced associative recognition accuracy with the Retrieving Effectively from Memory (REM) model (Criss & Shiffrin, 2005; Shiffrin & Steyvers, 1997) by degrading the quality of feature representations stored in REM's memory. Models based on connectionist principles, like the source of activation confusion (SAC) model (Buchler & Reder, 2007; Reder et al., 2000) and Li et al.'s (2005) neural network model of associative binding, make slightly different assumptions about the source of age differences in recognition memory, yet these assumptions still ultimately align with the idea that representations become noisier or less distinctive with age. For example, the SAC model attributes age differences in recognition to over-saturated (or more diffuse) semantic association networks in older adults, resulting in each new experience being less distinctive by virtue of it activating many similar prior experiences (Buchler & Reder, 2007). Finally, the Item-Context-Ensemble (ICE) model proposes that older adults' recognition errors arise from less unitized memory representations, whereby items are less well integrated into the contexts in which they appear (Murnane et al., 1999; Bayen et al., 2000).

Computational models have rarely been applied to explain age differences in memory across multiple tasks in tandem. A recent notable exception comes from Healey and Kahana (2016). Using the Context Maintenance and Retrieval (CMR2) model (Lohnas et al., 2015; Polyn et al., 2009), they identified combined impairments in attentional control, context retrieval, output monitoring, and increased representational noise could account for both the pronounced age differences in free recall and the more subtle ones in item recognition. This impressive modeling work notwithstanding, efforts to identify shared processes underlying age differences across various tasks of memory remain limited. We intend to address this gap by applying the eCFM to age differences in serial recall, order reconstruction, and item recognition across a diversity of representational domains.

Modeling Age Differences in the eCFM

As noted earlier, a key innovation of the eCFM is its use of structured representations. This allows the eCFM to apply the same principles to explain the occurrence of true and false memories across multiple representational domains, including semantic, orthographic, phonological, and visual. This makes the eCFM an especially promising tool for studying age differences in memory, given that older adults exhibit an increased susceptibility to false memories, erroneously recalling or recognizing events that share salient features with previously learned experiences (Devitt & Schacter, 2016; Greene & Naveh-Benjamin, 2020; Koutstaal & Schacter, 1997; Schacter et al., 1997; Tun et al., 1998; Yassa et al., 2011). Critically, older adults' heightened susceptibility to false memories have been reported on a variety of tasks, including recall (Balota et al., 1999; Norman & Schacter, 1997), item recognition (Dennis et al., 2007; Stark et al., 2013; Trelle et al., 2017), and associative recognition (Greene & Naveh-Benjamin, 2020, 2024); and have also been reported across multiple representational modalities, including semantic (Abadie et al., 2021; Rankin & Kausler, 1979; Smith, 1975), visual (Koutstaal, 2003; Koutstaal & Schacter, 1997), phonological (Budson et al., 2003; Rankin & Kausler, 1979; Watson et al., 2001) and orthographic (Piguet et al., 2008). Thus, applying a common modeling framework (the eCFM) to a diverse set of tasks and representational modalities has enormous potential for advancing an integrated theory of memory that can explain both age-related constancies and differences in true and false memories.

The eCFM also offers considerable advantages over similar models, particularly those rooted in the MINERVA 2 (Hintzman, 1984, 1986) framework (e.g., DRYAD; Benjamin, 2010), in its ability to uncover processes of age-related memory change. The eCFM contains several viable candidate mechanisms represented by different parameters in the model. These include the size or composition of the lexicon (the model's "long-term memory") that controls item similarities, the learning rate L that determines how many features of an item are faithfully represented in the model, the positional distinctiveness of each item dictated by its representational similarity (d) to previous items, the cue-specificity of retrieval (τ), and the decision threshold T for outputting a response or recognizing an item as "old." Although variations in each of these parameters might be necessary to account for the patterns of age differences across our diverse set of tasks, our goal is to identify a common mechanism that can explain these differences with minimal assumptions (i.e., the fewest number of parameter differences). To this end, we appeal primarily to parameters dictating the representational fidelity of the model's newly learned experiences (parameters L and d).

Our emphasis on representational fidelity is motivated by extensive empirical work converging on a potential universal principle of age-related memory change: episodic memories become “fuzzier” or less precise with aging (Greene & Naveh-Benjamin, 2023). Support for this principle comes from findings in item recognition (e.g., Koutstaal & Schacter, 1997), cued recall (e.g., Castel, 2005), associative recognition (e.g., Greene & Naveh-Benjamin, 2020), free recall (e.g., Norman & Schacter, 1997), reading comprehension (e.g., Radvansky et al., 2001), autobiographical memory (e.g., Levine et al., 2002), and visual feature reproduction (e.g., Peich et al., 2013). Consistently, studies in these domains have reported age differences in memory for the specific details of past experiences but preserved memory for the meaning (or gist) of those experiences (cf., Brainerd & Reyna, 2015). Yet, much of the evidence supporting this principle comes from conventional analyses of memory performance (e.g., recognition or recall accuracy) with relatively few attempts to measure the latent processes underlying these performance metrics. By tuning the representational fidelity of the eCFM model’s learned experiences, our approach holds promise for directly testing that proposed principle of age-related memory change.

This also aligns our approach with earlier models of age differences in recognition (Benjamin, 2010; Buchler & Reder, 2007; Li et al., 2005; Stephens & Overman, 2018) and recall (Neath & Surprenant, 2007; Surprenant et al., 2006), while extending on these prior models in important ways. These extensions include generalizing the representational fidelity account of age differences in memory across multiple tasks, including previously un-modeled tasks like serial order reconstruction, and across multiple representational modalities (semantic, phonological, orthographic, and visual) over which age differences in true and false memory have been observed empirically. This generalizability is critical for advancing a more comprehensive theory of age-related memory change.

Another key extension that the eCFM affords over these earlier models is its level of precision in localizing which components of memory are less precise with age and how these components vary based on task demands. The two parameters that we manipulated to capture age-related differences in memory performance across our experiments (L and d)[3] affect the quality of the model's representations in different ways. Lowering L results in a more impoverished memory representation (fewer encoded lexical and order features) for each item, while lowering d amplifies similarities among the position features associated to each item, blurring distinctions about which items appeared in which positions. An age difference could arise singularly from either L or d or through a joint impairment in both. Thus, the eCFM can identify distinct mechanisms of age-related representational change that affect different components of memory. There could be an overall reduction in the fidelity of memories with age ($L_{\text{old}} < L_{\text{young}}$) with no difference in the precision of positional features that are encoded ($d_{\text{old}} = d_{\text{young}}$). Alternatively, there could exist a selective deficit in the precision of order/context representations ($d_{\text{old}} < d_{\text{young}}$) with no concomitant reduction in the number of lexical and positional features of an item that are stored in memory ($L_{\text{old}} = L_{\text{young}}$). Finally, there could be a reduction in both overall item fidelity ($L_{\text{old}} < L_{\text{young}}$) and order precision ($d_{\text{old}} < d_{\text{young}}$), such that older adults store fewer lexical and order features of each item in memory and, for the limited positional features they do encode, they represent these features more diffusely across different items.

The degree to which L, d, or both are affected by age could depend on the demands that the task of memory places on remembering item or order information. For example, whereas item recognition prioritizes memory for which items were encountered with little need to remember their order, order reconstruction prioritizes memory for the orders in which items appeared without needing to reconstruct memory for the items themselves. Meanwhile, serial recall demands one remember both the items and their orders with no explicit cues

[3] Because we zeroed out the order vectors in modeling recognition memory, we did not attribute age differences in recognition results in Experiments 8-11 to differences in d. However, by allowing for such representations to be stored, the eCFM has potential to be extended to account for an age difference in order memory that manifests even on tests where order information typically plays a less important role, like recognition.

available. By systematically investigating age differences in each of these tasks within a common modeling framework (the eCFM), our approach aims to refine current theories of age-related changes in the representational precision of memories by revealing which specific components of memory are degraded with age across widely different tasks. We turn now to introducing our novel experimental methods and simulating age differences in true and false memory with the eCFM.

Overview of the Present Experiments

Our aim in the remainder of this review is to evaluate the generalizability of the eCFM as a step toward developing a more general model of human memory, one that is capable of explaining memory functioning across numerous tasks, representational modalities, and individuals of different ages with varying degrees of memory capability. As an empirical lens, we focus on true and false memories, instances in which individuals recall, reconstruct, or recognize information that was or was not actually encountered. False memory is a powerful tool for understanding the structure and dynamics of human memory because it reveals the influence of prior lexical and conceptual knowledge on memory construction. These errors reflect not random noise, but systematic activation of representations shaped by experience, semantic organization, and feature overlap (see Chang & Brainerd, 2021, for a review; Chang et al., 2025, for modeling work).

A widely used method for studying false memory is the Deese–Roediger–McDermott (DRM or "Dream"; Deese, 1959; Roediger & McDermott, 1995) paradigm, in which participants study lists of semantically, phonologically, orthographically, or visually similar items that are all related to an unpresented thematic lure. Early DRM studies reported high rates of false recall or recognition of the critical lure among young adults (e.g., Roediger & McDermott, 1995), and ensuing studies found that these false memory effects were even more pronounced in older adults (Balota et al., 1999; Koutstaal & Schacter, 1997; Norman & Schacter, 1997; Tun et al., 1998; for a review, see Devitt & Schacter, 2016). The DRM and related paradigms have been instrumental in demonstrating how memory is shaped by the structure of prior perceptual (orthographic, phonological, and/or visual information) and semantic information (Guitard et al., 2025b). The present set of experiments draws upon the DRM methodology to provide the eCFM with a rich set of false memory experiences to simulate.

Each experiment tasks participants, young and older adults alike, with encoding lists of items that are related along one common representational dimension (semantic, phonological, orthographic, or visual), while varying how memory for those items is assessed. Experiments 1 through 4 test participants' order reconstruction, that is, whether they can correctly place previously studied items in the same order in which they were studied (*true reconstruction*) while avoiding erroneously assigning an unstudied lure to a list position (*false reconstruction*). Experiments 5 through 7 test participants' serial recall, the ability to recall in order the exact items that were presented on a list (*true recall*) and to inhibit recall of an unstudied lure (*false recall*). The eCFM has previously explained the performance of younger adults on both these tasks with great success (Guitard et al., 2025a, 2025b, 2025c), but there are two key innovations of our first seven experiments. First, these experiments generalize the eCFM to explaining serial recall and order reconstruction performance of older adults, who typically commit more errors on these tasks than younger adults do (Cabeza et al., 2000; Golomb et al., 2008). This represents a rare application of computational modelling to explain age differences in memory across multiple tasks, while providing to our knowledge the first formal modelling of these differences in temporal order reconstruction. Second, we assess the model's ability to reconstruct the order of *visually* similar experiences in Experiment 4. This provides a proof of whether the eCFM can generalize beyond verbal lists, with implications for the model's ability to explain the dynamics of visual memory.

We then extend the eCFM to the domain of recognition memory. Experiments 8 through 11 test whether participants can recognize which items were studied (*true recognition*) while rejecting unstudied related lures (*false recognition*) and completely unrelated items. We aim to show that the eCFM can generalize to this widely used task of memory by applying the same principles it uses to explain performance on serial recall and order reconstruction.

Through our modelling of these diverse sets of tasks, we aim to show that a common set of processes can explain the occurrence of true and false memories across varied task structures and representational modalities. Furthermore, by integrating age as a core variable in our empirical and computational framework, we aim to provide a more general account of memory that is applicable across age groups and memory domains, thereby contributing a simple yet scalable tool to bridge theoretical and age-specific research gaps. We turn now to our first simulation results, applying the eCFM to tests of order reconstruction.

Order Reconstruction

In Experiments 1 to 4, we employed an order reconstruction task involving young adults (ages 18–25) and older adults (ages 65–80). Participants studied sequences comprising six items, which varied across experiments: words (Experiments 1 and 2), nonwords (Experiment 3), or images (Experiment 4). Following each study phase, participants engaged in a brief, 6-second parity judgment task as a distraction. Subsequently, they reconstructed the original order of the studied items. To systematically investigate false memory, half of the trials included either a related or unrelated lure item that had not been presented during the study phase (i.e., the candidate response set included all studied items and one extra lure).

The objectives were twofold: firstly, to systematically explore false reconstruction across different stimulus types and age groups, and secondly, to evaluate the capability of the eCFM to account for these empirical observations.

In our previous research (Guitard et al., 2025a, 2025b, 2025c) we have demonstrated that the eCFM effectively modeled false recall in tasks employing verbal materials with younger adults. The current set of experiments, however, introduces novel elements by incorporating older adults and employing nonverbal stimuli (images), neither of which has previously been modeled within eCFM. This extension allows us to explicitly assess whether the structured representational mechanisms and similarity-based retrieval processes inherent in eCFM generalize beyond the conditions in which they were originally validated.

We expected that the fundamental mechanisms of eCFM—structured representations and similarity-based retrieval—will effectively generalize to older adults and nonverbal stimuli. However, we anticipate that accurately capturing age-related differences may require adjusting existing model parameters. The order reconstruction task itself is particularly valuable, as it directly measures the retrieval of precise temporal information and the ability to differentiate true memories from memory distortions. eCFM's structured representation and similarity-based retrieval processes are ideally suited to explaining performance in this task, naturally accommodating both accurate and distorted memory retrieval.

Furthermore, examining older adults' order reconstruction abilities addresses a gap in the literature. Although there are established age differences in serial recall (e.g., Golomb et al., 2008), whether such differences would emerge in a test of order reconstruction – in which older adults are provided the items explicitly and must place them in the correct order in which they were studied – remains less well understood. The present set of experiments addresses this empirical gap while simultaneously modeling processes underlying potential age differences in order reconstruction.

Experiment 1

In Experiment 1, we examined serial reconstruction of semantically related word lists. During the reconstruction test, half of the trials included an unstudied semantically related lure, while the other half included an unstudied semantically unrelated lure. The goal was to demonstrate that participants are more likely to falsely reconstruct unstudied but semantically related lures (i.e., critical lures) compared to semantically unrelated lures, and to investigate age-related differences between young and older adults.

We predicted that older adults would commit more false reconstructions given their heightened susceptibility to false recall and recognition on semantic DRM procedures (e.g., Norman & Schacter, 1997) and that they would make more errors in correct reconstructions given their difficulties remembering when things occurred (e.g., Kahana et al., 2002). However, most of the evidence attesting to age differences in temporal memory has come from serial recall procedures (Golomb et al., 2008; Maylor et al., 1999), which require participants to remember both item and order information with no external cues provided. An exception comes from a study by Cabeza et al. (2000), in which participants were provided two studied items side-by-side at test and had to decide which item had been studied more recently. Older adults made more errors on this task, indicative of reduced temporal order memory. Unlike these procedures, in an order reconstruction test, all studied items are provided as cues, and the participant must reconstruct the order in which these items were presented, while avoiding placing any unstudied item in a position. The increased environmental support might reduce age differences in temporal memory (e.g., Craik, 1983). Experiment 1 allowed us to assess this possibility, providing to our knowledge the first joint empirical and computational modelling of older adults' order reconstruction for semantically related items.

Method

Participants. Eighty participants (40 young adults and 40 older adults) were recruited via Prolific (https://www.prolific.com/). Each participant received £9.00 per hour (pro-rated) for their participation. All participants in all experiments provided their informed consent to participate. All experiments were approved by the School of Psychology Ethics Committee of Cardiff University. Below, we describe sample size determination, inclusion and exclusion criteria, and the demographic composition of the sample.

Sample Size Determination. The experiment was modeled after Experiment 1 of Guitard et al. (2025b), who used a serial recall task and found a large false memory effect (Cohen's d = 0.97). An a priori two-tailed paired-sample t-test was conducted using G*Power 3.1.9.7 (Faul et al., 2007) with default parameters and the effect size reported by Guitard et al. This analysis revealed that a total of 16 participants would be needed to achieve a power of .95. However, to obtain more stable estimates, we increased the sample size to 40 participants per age group. A sensitivity analysis, conducted with α = .05 and power = .95, indicated that a sample of 40 participants per group would allow us to detect a medium effect size of Cohen's d = 0.58. This sample size was adopted for all experiments unless otherwise specified.

Inclusion Criteria. Eligibility criteria were as follows: (a) native English speaker, (b) nationality from the United Kingdom, United States, or Canada, (c) normal or corrected-to-normal vision, (d) no color-blindness, (e) no cognitive impairment, mild cognitive impairment, or dementia, (f) no language-related disorders or literacy difficulties, (g) aged between 18 and 25 years for the young adult group or between 65 and 80 years for the older adult group, (h) a Prolific approval rating of at least 90%, and (i) consumption of no more than 5–9 units of alcohol per week. All inclusion criteria were self-reported by participants, except for the approval rating, which was provided by Prolific. The same inclusion criteria applied in all experiments.

Demographic Composition. In Experiment 1, the mean age of participants in the young adult group was 21.77 years (SD = 2.03). Eighteen participants identified as female, 20 as male, and 2 preferred not to specify their gender. In the older adult group, the mean age was 69.08 years (SD = 3.71), with 29 participants identifying as female and 11 as male.

Ethical approval. All experiments were reviewed and approved by the School of Psychology Ethics Committee at Cardiff University. Participants provided electronic informed consent before beginning the experiment and received a debriefing form upon completion. The research was conducted in accordance with the approved protocol *Toward a Theoretical Understanding of the Relationship Between Human Memory and Attention* (EC.22.09.20.6627G).

Materials and Design. The stimuli in Experiment 1 were drawn from Experiment 1 of Guitard et al. (2025b). A total of 20 lists, each consisting of six words thematically related to an unpresented critical lure, were chosen based on the University of South Florida's free association, rhyme, and word fragments norms (Nelson et al., 2004). These words were carefully selected to maximize the likelihood of detecting a false memory for the critical unstudied but related lures. Two versions of each list were created (see Appendix A) by assigning the critical lure either to its respective list (Related condition) or to a different list (Unrelated condition), with the aim of minimizing the likelihood of detecting a false memory for these lures in the unrelated condition (see Guitard et al., 2025b). Each participant received 10 lists in which the critical lure was related to the items in the list and 10 lists in which the critical lure was unrelated to the studied items. To ensure that a word was never presented more than once to any given participant, we generated two sets of the 20 lists. The order of the versions was counterbalanced across participants, with half of the participants tested on version 1 and the other half on version 2. The order of the conditions (related lure lists or unrelated lure lists) and the order of words within a list were presented in randomized order for each participant.

Procedure. The experiment was programmed with PsyToolKit (Stoet, 2010, 2017), and each online experimental session lasted approximately 12 minutes. The experiment was self-paced; participants initiated each trial by pressing the space bar. If a participant did not commence the trial within the 60-second window, the trial was automatically initiated to ensure completion within the anticipated timeframe.

Immediately following trial initiation, six words to be remembered were sequentially presented at the center of the computer screen, with a rate of one word per second (1000 ms on, 0 ms off). The words were displayed in white lowercase 30-point Times font against a black background. After the presentation of the last word, participants completed a parity judgment task lasting 6 seconds. In this task, a random integer from 0 to 9 appeared at the center of the screen, accompanied by the instruction “Press the Z key for an odd number” displayed at the bottom left and “Press the M key for an even number” at the bottom right. During these 6 seconds, participants were directed to complete as many parity judgments as possible.

Following the parity judgment task, all studied words and the unstudied critical lure were presented in a randomized order on the screen, arranged in three vertical lists (three items, one item, three items), accompanied by a reconstruction cue (“Reconstruct the order by clicking on the words”) at the top of the computer screen. Participants were instructed to click on the words in the order they were presented. Once a participant clicked on a word, it transitioned from yellow to blue, indicating its selection. This process continued until the sixth-to-be-remembered word was selected. Upon selecting the sixth item, the trial concluded. Participants were not permitted to backtrack and modify a response once it was registered.

Data Analysis. All data related to the experiments can be found on the Open Science Framework page linked to this project (OSF). Additionally, R markdown files for each experiment, including analyses and modelling codes, are available on the same page.

Scoring. In Experiments 1 to 4, we compute the proportion of correct reconstruction and the proportion of false memories. The proportion of correct reconstruction was calculated using a strict serial reconstruction criterion, where an item must be reconstructed in its presented serial position to be considered correct. A false memory was recorded when the critical lure was reconstructed, regardless of the position.

Statistical Analysis. In all experiments, we conducted both frequentist and Bayesian statistical analyses using R (R Core Team, 2024). ANOVAs were performed with condition (unstudied related critical lure vs. unstudied unrelated critical lure) and age group (younger adults vs. older adults) as factors, analyzing both proportion correct and false memory rates. For transparency, we report serial position curves in our figures; however, these were not included in the main analyses, as we had no specific predictions regarding them in order reconstruction. However, for interested readers they are presented on the OSF page associated with the manuscript. Frequentist analyses are presented for descriptive purposes, whereas Bayes factors (BF) are used to guide our inferences.

Frequentist ANOVAs were conducted using the *ez* package (version 4.4-0; Lawrence, 2016). BF analyses were performed using the *BayesFactor* package (version 0.9.12-4.2), with default prior settings (Morey & Rouder, 2024; Rouder et al., 2009, 2012). Each Bayesian

ANOVA was based on 100,000 iterations, followed by an additional 10,000 iterations to ensure the proportional error of the estimates was reduced to below 10%. Main effects and interactions were tested by removing each term from the full model, with participants included as a random factor. For BF results, we report the strength of evidence in favor of the alternative hypothesis (BF_{10}) or, conversely, in favor of the null hypothesis (BF_{01}), where $BF_{01} = 1 / BF_{10}$. Interpretation of BF values follows guidelines suggested by van Doorn et al. (2021): values between 1 and 3 indicate weak or anecdotal evidence, values between 3 and 10 reflect moderate evidence, values between 10 and 30 denote strong evidence, values between 30 and 100 correspond to very strong evidence, and values above 100 represent extreme evidence. However, we encourage readers to interpret Bayes Factors as continuous values representing a spectrum of evidence, rather than relying strictly on categorical thresholds.

Results

The experimental and simulation results for Experiment 1 are shown in Figure 1. The figure illustrates the proportion of correct responses and the proportion of false reconstructions as a function of condition (related, unrelated), age group (younger adults, older adults), and serial position.

Strict Correct. For the proportion of correct responses, participants performed better in the related condition (M = .57, SD = .21) than in the unrelated condition (M = .54, SD = .20), $F(1, 78) = 6.08$, $\eta_p^2 = .072$, $BF_{10} = 3.16$. Younger adults (M = .59, SD = .19) outperformed older adults (M = .51, SD = .20) at the descriptive level; however, the evidence only weakly supported the absence of an age difference, $F(1, 78) = 3.05$, $\eta_p^2 = .04$, $BF_{01} = 2.52$. There was also strong evidence against a two-way interaction between age group and condition, $F < 1$, $\eta_p^2 = .01$, $BF_{01} = 31$.

False Memory. Participants produced a higher proportion of false memories in the related condition (M = .08, SD = .05) compared to the unrelated condition (M = .04, SD = .04), $F(1, 78) = 87.12$, $\eta_p^2 = .53$, $BF_{10} > 10{,}000$. Younger adults (M = .05, SD = .03) and older adults (M = .07, SD = .03) produced a comparable number of false reconstructions, $F(1, 78) = 4.74$, $\eta_p^2 = .06$, $BF_{01} = 6.10$. There was also strong evidence against an interaction between age group and condition, $F < 1$, $\eta_p^2 = .01$, $BF_{01} = 19.45$.

Empirical Summary

To summarize, younger and older adults alike were more likely to falsely reconstruct the unstudied critical lure when it was semantically related to the items in the list than when it was unrelated. There was no credible age-related difference in the rate of false reconstructions, and although younger adults descriptively outperformed older adults in correct reconstructions, this difference was not supported in the statistical analysis. Thus, unlike previous studies showing age differences in temporal memory with serial recall (Golomb et al., 2008) or forced choice (Cabeza et al., 2000) procedures, these differences were not as apparent in an order reconstruction procedure in which participants had access to all the studied items as retrieval cues, potentially providing older adults with enough environmental support to perform at a level similar to younger adults (cf., Craik, 1983). In the next section, we evaluate whether the eCFM can capture these key findings.

eCFM Simulation of Experiment 1

In this section, we evaluate the capacity of the eCFM (Guitard et al., 2025a, 2025b, 2025c) to reproduce the key empirical findings presented in Figure 1. Consistent with our prior work, the present simulations were conducted using representations tailored to the demands of the task, in this case, semantic representations. For ease of visual comparison, the empirical data are presented as bar plots with the final model simulations overlaid as lines.

This format allows for a straightforward assessment of the model's ability to reproduce observed behavioral patterns. To evaluate the quality of the model fit, we report both the coefficient of determination (R^2) and the mean absolute deviation (MAD). These metrics provide complementary perspectives: R^2 captures the proportion of variance in the data explained by the model, while MAD reflects the average absolute difference between observed and predicted values. In line with conventions in cognitive modeling (e.g., Cohen, 1988; Roberts & Pashler, 2000), we consider R^2 values above .80 and MAD values below 0.10 as indicative of a good quantitative fit, though we also consider theoretical interpretability and qualitative correspondence between predicted and observed patterns in our overall evaluation.

Simulation Parameters. We first conducted a grid search to identify the parameter space best suited to capture the key experimental findings. In this initial step, the selection parameter s (controlling how much an item's activation is suppressed following its retrieval) was fixed at 1, reflecting the constraint of the reconstruction task in which each word could only be selected once. The parameter g (dictating the decay in the baseline learning rate across serial positions) was set to 0.02. The parameter τ (scaling the influence of similarity on retrieval) was fixed at 3. The encoding base rate L (governing the fidelity of encoding for the first item in the list) was varied from 0.15 to 0.40 in increments of 0.01, and the order discriminability parameter d was varied from 0.10 to 0.30 in similar steps.

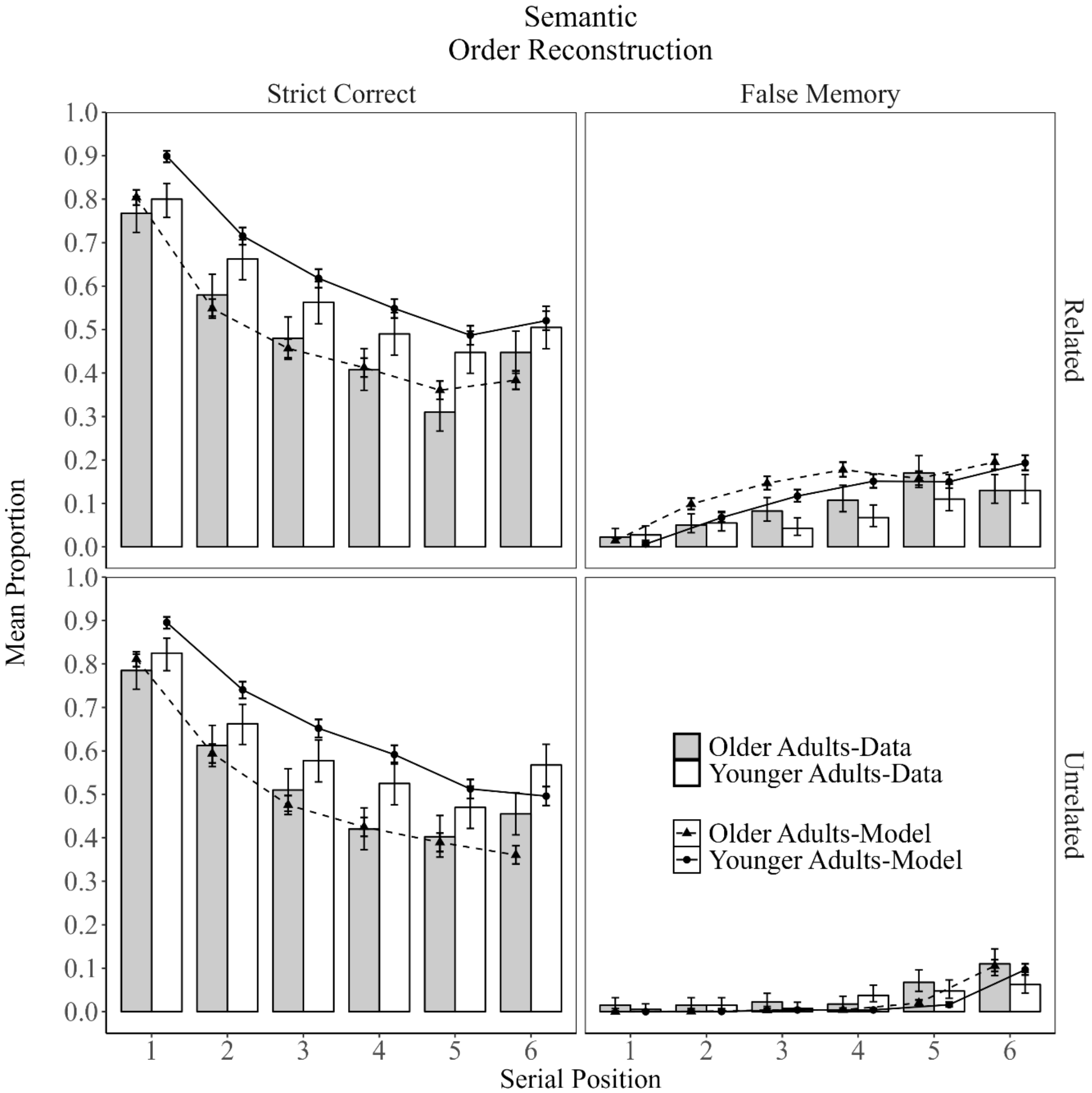
Following the grid search, we ran full simulations for each experiment using 100 simulations. Each simulation included 20 related lists and 20 unrelated lists, equivalent to 200 participants each completing 10 related and 10 unrelated lists. For Experiment 1, the model was trained on the same word lists given to participants, allowing it to build structured representations of these items using LSA (Landauer & Dumais, 1997) applied to a larger corpus in which the studied words were embedded (see introduction for description of the semantic representations).[4] To capture the data of young and older adults in Experiment 1, the following parameter values were applied: $L = .17$ for both younger and older adults; $d = .26$ for younger adults and $d = .17$ for older adults. These settings reveal that, although the empirical results failed to detect age differences, a difference in positional precision (d) was necessary to ensure the model could recreate the empirical patterns of true and false reconstruction within each age group, where serial positions were more similar to one another in the simulation of older adults' performance than in the simulation of younger adults' performance.

Simulation Results. The simulation results for true and false reconstructions among young and older adults and across list types are presented in Figure 1, alongside the empirical data (bars). Overall, the model captures all of the key findings, including the typical serial position curve with superior memory for early list items and a modest boost for the final items, although this effect was more pronounced in the related than in the unrelated lists.

The model produces a slight advantage in correct reconstruction for younger adults compared to older adults, as well as comparable rates of false memory across age groups. It accurately reflects the higher prevalence of false memories in the related condition relative to the unrelated condition. The model exhibited excellent fits to the data of younger adults (R^2 = .98, MAD = .05) and older adults (R^2 = .97, MAD = .03).

[4] The word *misfiling* was replaced with *misfile* due to its unavailability in the lexicon used for simulations. This change applied to all experiments using the semantic lists available in Appendix A.

Figure 1. *Empirical (bars) and computational (lines) results for the mean proportion of correct reconstructions and false memories for the critical lure in Experiment 1 (semantically related vs. semantically unrelated), shown for younger and older adults.*



Note. *Error bars represent 95% credible intervals. The left panels display the mean proportion of correct responses as a function of serial position, while the right panels show the mean proportion of false memories by serial position. The top row presents data for the related lure condition, and the bottom row for the unrelated lure condition.*

Discussion

In Experiment 1, younger and older adults were tasked with reconstructing the order of six studied, semantically related words when provided, at the time of retrieval, those same words and a critical lure that was either related or unrelated to them. There were slight age differences in order reconstruction accuracy, with young adults at least descriptively exhibiting a slight advantage over older adults. The rate of false reconstructions (erroneously assigning the unstudied item to a list position) was comparable between age groups and was higher for related than unrelated lures. Crucially, by incorporating semantic representations into the eCFM, we were able to successfully reproduce false reconstructions while accurately capturing key patterns of memory performance across age groups. The small age-related difference in correct order reconstruction was accounted for by reducing the d parameter for older adults, resulting in less precise order representations due to increased similarity among successive positions. However, other parameter configurations could likely produce similar results, for instance, maintaining a constant d value while allowing L to vary. Here, we chose to vary d, guided by prior theoretical and empirical work suggesting that older adults exhibit less precise order memory (e.g., Brown et al., 2000; Maylor et al., 1999).

Experiment 2

In Experiment 2, we tested order reconstruction for lists of phonologically similar items. The goal was to extend the findings of Experiment 1 from the semantic to the phonological domain and to evaluate whether the model could capture the main effects by incorporating phonological word representations based on the method inspired by the work of Parrish (2017) described above in the phonological representation section of our introduction.

Method

Participants. The sample size justification, recruitment procedure, and inclusion criteria were the same as those used in Experiment 1, with one additional exclusion criterion: participants who took part in Experiment 1 were not eligible to participate in Experiment 2. As a result, 80 different participants were recruited through Prolific. In the younger adult group, the mean age was 22.32 years (SD = 1.90); 24 participants self-identified as female, 15 as male, and 1 preferred not to specify their gender. In the older adult group, the mean age was 69.22 years (SD = 3.90); 30 participants self-identified as female and 10 as male.

Materials. The stimuli in Experiment 2 were taken from Experiment 2 of Guitard et al. (2025b). A total of 20 lists, each consisting of six words phonologically related to an unpresented critical lure, were utilized. These words were carefully chosen to maximize the likelihood of detecting a false memory for the critical unstudied but related lures. Like Experiment 1, the critical lures were reorganized to generate an additional 20 lists in which the critical lure was unrelated to the studied items. The stimulus lists are included in Appendix B (see Guitard et al., for more details), with participants receiving all the lists in either version 1 or version 2 in a random order. Assignment to version was randomized across participants, with an equal number of participants assigned to each.

Procedure and Data Analysis. The experimental procedure and data analysis methods in Experiment 2 were identical those of Experiment 1, except for the stimuli that we changed to phonologically related lists rather than semantically related lists.

Results

The results of Experiment 2 are presented in Figure 2, which displays simulation and empirical results for the proportion of correct responses and false reconstructions as a function of condition (related vs. unrelated), age group (younger vs. older adults), and serial position.

Strict Correct. Participants' performance with the strict correct scoring was comparable across the related (M = .40, SD = .15) and unrelated (M = .41, SD = .15) conditions, $F < 1$, $\eta_p^2 = .004$, $BF_{01} = 28.34$. Younger adults (M = .44, SD = .15) outperformed older adults (M = .36, SD = .11), $F(1, 78) = 8.34$, $\eta_p^2 = .09$, although the evidence was only modest, $BF_{10} = 1.98$. There was no interaction between age group and condition, $F < 1$, $\eta_p^2 = .00$, $BF_{01} = 30.00$.

False Memory. Replicating the pattern observed in Experiment 1, participants produced a higher proportion of false memories in the related condition (M = .12, SD = .03) than in the unrelated condition (M = .04, SD = .03), $F(1, 78) = 556.59$, $\eta_p^2 = .88$, $BF_{10} >$ 10,000. Younger adults (M = .08, SD = .03) and older adults (M = .09, SD = .02) produced similar rates of false reconstructions, $F(1, 78) = 4.57$, $\eta_p^2 = .06$, $BF_{01} = 12.63$. There was strong evidence against an interaction between age group and condition, $F < 1$, $\eta_p^2 = .00$, $BF_{01} = 32.26$.

Empirical Summary

Overall, the empirical findings on order reconstruction with phonologically similar lists were consistent with those of Experiment 1 with semantically similar lists. Both younger and older adults made more false reconstructions for unstudied lures that were phonologically related versus unrelated to the studied items. There was no substantial age-related difference in the rate of false memories, although younger adults performed slightly better in terms of correct reconstructions. In the following section, we assess whether eCFM can account for these key results.

eCFM Simulation of Experiment 2

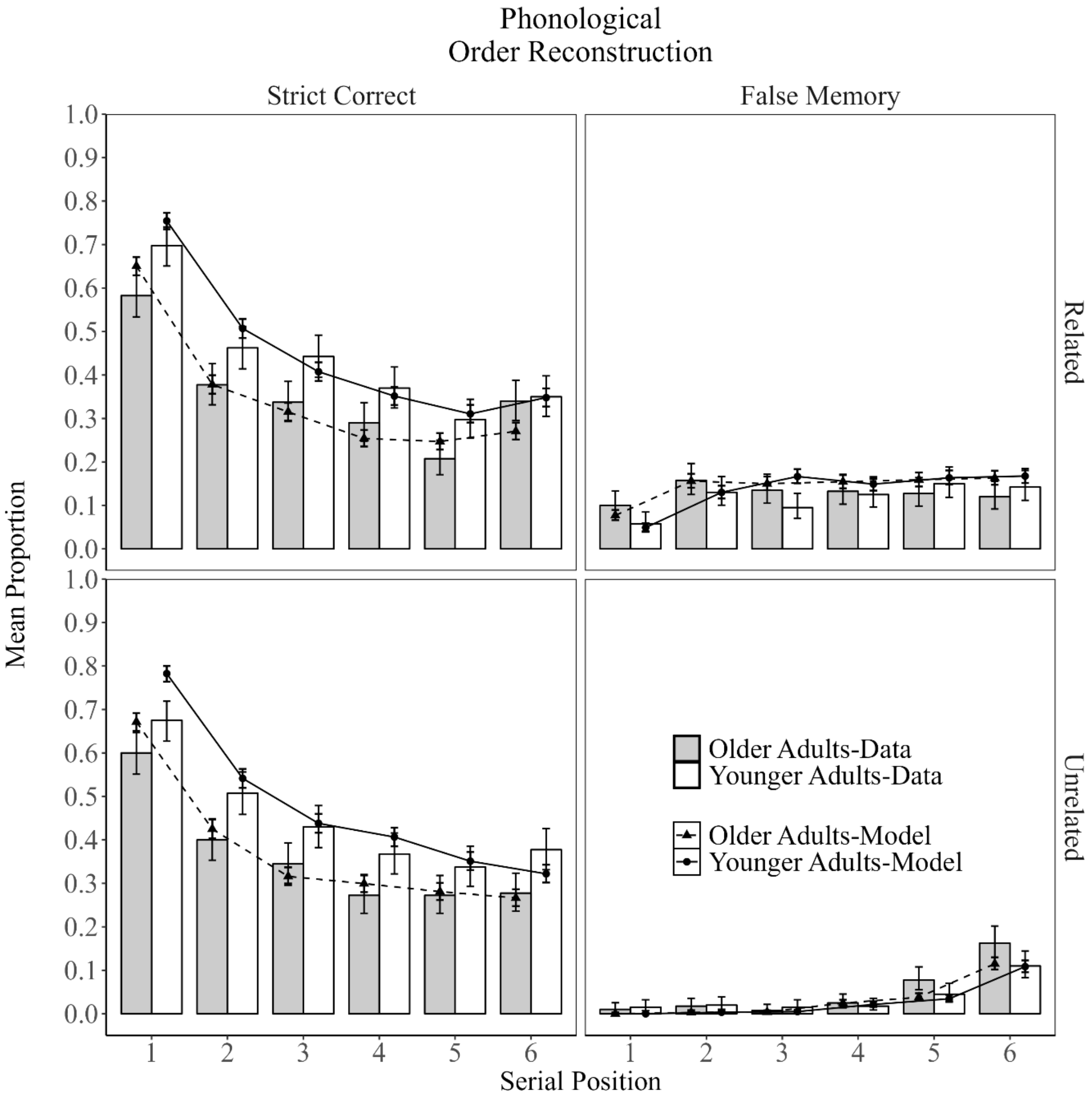
Here, we evaluate the ability of the eCFM (Guitard et al., 2025a, 2025b, 2025b) to reproduce the key empirical patterns shown in Figure 2. As in the simulation for Experiment

1, we used representations tailored to the specific demands of the task, in this case, phonological representations derived from Parrish (2017) for the phonologically related materials.

Simulation Parameters. The simulation procedure was identical to that used in Experiment 1, except the model was trained on the phonological word lists. Specifically, we conducted 100 simulations, each comprising 20 related and 20 unrelated lists, equivalent to 200 simulated participants, each completing 10 related and 10 unrelated lists. As in Experiment 1, we used the same settings of $g = 0.02$, $s = 1$, and $\tau = 3$ to model the data of young and older adults but we allowed L and d to vary between age groups. Nonetheless, the same value of $L = .15$ was specified for both younger and older adults, while $d = .17$ for younger adults and $d = .10$ for older adults. Both L and d values were lower than those used in Experiment 1 to reflect the reduced proportion of correct responses, likely due to the greater confusability associated with phonological information (see Roodenrys et al., 2022). As in simulations for semantic order reconstruction (Experiment 1), the lower d value for older compared to younger adults reflects reduced discriminability between serial positions, while the consistent value of L indicates no age-related difference in the global fidelity of item representations.

Simulation Results. The simulation results are presented in Figure 4, which shows the proportion of correct responses and false memories as a function of serial position, lure relatedness, and age group (younger vs. older adults), alongside the empirical data (bars). Overall, the model provided an excellent fit to the data (for younger adults, $R^2 = .98$ and MAD = .03; for older adults, $R^2 = .97$ and MAD = .03). It successfully captured the serial position curve, including higher accuracy for early list items and a modest recency boost, an effect that was more pronounced in the related condition and reduced or absent in the unrelated condition. The model also reproduced the slight advantage in correct recall observed in younger adults compared to older adults, as well as the comparable rates of false memories across age groups. Crucially, the model, like the human participants, produced more false memories in the related condition than in the unrelated condition; and, like in the data, produced false memories more evenly across serial positions in the related than in the unrelated condition.

Figure 2. *Empirical (bars) and computational (lines) results for the mean proportion of correct reconstructions and false memories for the critical lure in Experiment 2 (phonologically related vs. phonologically unrelated), shown for younger and older adults.*



Note. *Error bars represent 95% credible intervals. The left panels display the mean proportion of correct responses as a function of serial position, while the right panels show* *the mean proportion of false memories by serial position. The top row presents data for the related lure condition, and the bottom row for the unrelated lure condition.*

Discussion

Experiment 2 tested order reconstruction of phonologically related word lists. Young adults exhibited a slight correct reconstruction advantage over older adults, but there were no age-related differences in the observed rates of false reconstructions. These false reconstructions were higher for unstudied words that were phonologically similar to the studied items than for those that were dissimilar. By incorporating structured phonological representations into the model, the eCFM effectively captured all key aspects of memory performance, including the serial position curve, false memory errors, and small age-related differences in correct reconstructions. These age differences were attributed to a reduced precision of order representations among older adults (lower d) with no concomitant decline in the overall fidelity of item representations (equal L). Thus, the same representational mechanism underlies the rather subtle age differences in temporal order reconstruction for both semantically (Experiment 1) and phonologically (Experiment 2) similar lists of items.

Experiment 3

Experiments 1 and 2 provide compelling evidence that the presence of an unstudied item, semantically related in Experiment 1 and phonologically related in Experiment 2, can systematically induce false reconstructions, with comparable rates observed in both younger and older adults. Importantly, these experiments also demonstrate that embedding structured representations into the model allows for accurate prediction of both correct recall and false memory rates and discovery of key representational differences between young and older adults. In Experiment 3, we extend this investigation by examining whether younger and older adults will falsely reconstruct an unstudied, orthographically related nonword (critical lure), and whether the model can successfully account for these findings and any age differences in true or false order reconstruction.

Method

Participants. In Experiment 3, we applied the same criteria as in the previous experiments, while excluding any participant who had taken part in either of the prior experiments. A new group of 80 participants (40 younger adults and 40 older adults) was recruited through Prolific. In the younger adult group, the average age was 23.02 years (SD = 1.84); 29 participants identified as female, 9 as male, and 2 preferred not to specify their gender. In the older adult group, the average age was 68.83 years (SD = 3.40); 27 participants identified as female and 13 as male.

Materials. In Experiment 3, we generated 20 lists, each consisting of six nonwords ranging in length from 4 to 7 letters. These nonwords were orthographically related to one another and to an unpresented critical lure. To ensure that all items were nonwords, each was checked against the English Lexicon Project database (Balota et al., 2007) to exclude existing English words. As in the previous experiments, the items were selected to maximize the likelihood of eliciting false memories for the unstudied, but orthographically related, critical lures. To create a control condition, the critical lures were rearranged to form 20 additional lists in which the lure was orthographically unrelated to the studied items. The complete set of stimulus lists is provided in Appendix C. An equal number of participants were tested on the 20 lists in version 1 and on the 20 lists in version 2, with the 20 lists presented in a random order for each participant.

Procedure and Data Analysis. The experimental procedure, scoring criteria, and data analysis methods used in Experiment 3 were identical to those employed in the previous experiments, with the exception of the study materials. In this experiment, the studied items were orthographically related nonwords.

Results

The empirical and computational results of Experiment 3 are presented in Figure 3, which shows the proportion of correct responses and false memories for younger and older adults as a function of serial position (1 to 6).

Strict Correct. For strict scoring, performance was nearly identical across the related (M = .39, SD = .17) and unrelated (M = .40, SD = .16) conditions, $F < 1$, $\eta_p^2 = .004$, $BF_{01} = 33.52$. Descriptively, older adults (M = .42, SD = .17) outperformed younger adults (M = .37, SD = .14), $F(1, 78) = 1.53$, $\eta_p^2 = .02$, although the Bayesian analysis provided evidence in favor of the absence of difference, $BF_{01} = 6.25$. There was also no evidence of an interaction between age group and condition, $F < 1$, $\eta_p^2 = .00$, $BF_{01} = 27.82$.

False Memory. In line with Experiments 1 (semantic) and 2 (phonological), participants exhibited a higher rate of false memories in the related condition (M = .10, SD = .04) compared to the unrelated condition (M = .03, SD = .03), $F(1, 78) = 192.67$, $\eta_p^2 = .71$, $BF_{10} > 10{,}000$. False memory rates were similar across age groups, with younger adults (M = .07, SD = .03) and older adults (M = .06, SD = .02) showing comparable performance, $F < 1$, $\eta_p^2 = .00$, $BF_{01} = 23.07$. There was no evidence for an interaction between age group and condition, $F(1, 78) = 5.09$, $\eta_p^2 = .00$, $BF_{01} = 3.05$.

Empirical Summary

Overall, the empirical findings were clear and largely consistent with those from Experiments 1 and 2. Younger and older adults alike were more likely to falsely reconstruct an unstudied lure when it was orthographically related versus unrelated to the studied items. There was no credible age-related difference in false or correct reconstructions, although older adults surprisingly showed slightly better performance at the descriptive level for correct reconstruction, a finding at odds with the slight age deficit in correct reconstruction for semantically and phonologically similar lists. In the following section, we evaluate whether the eCFM can account for these key findings.

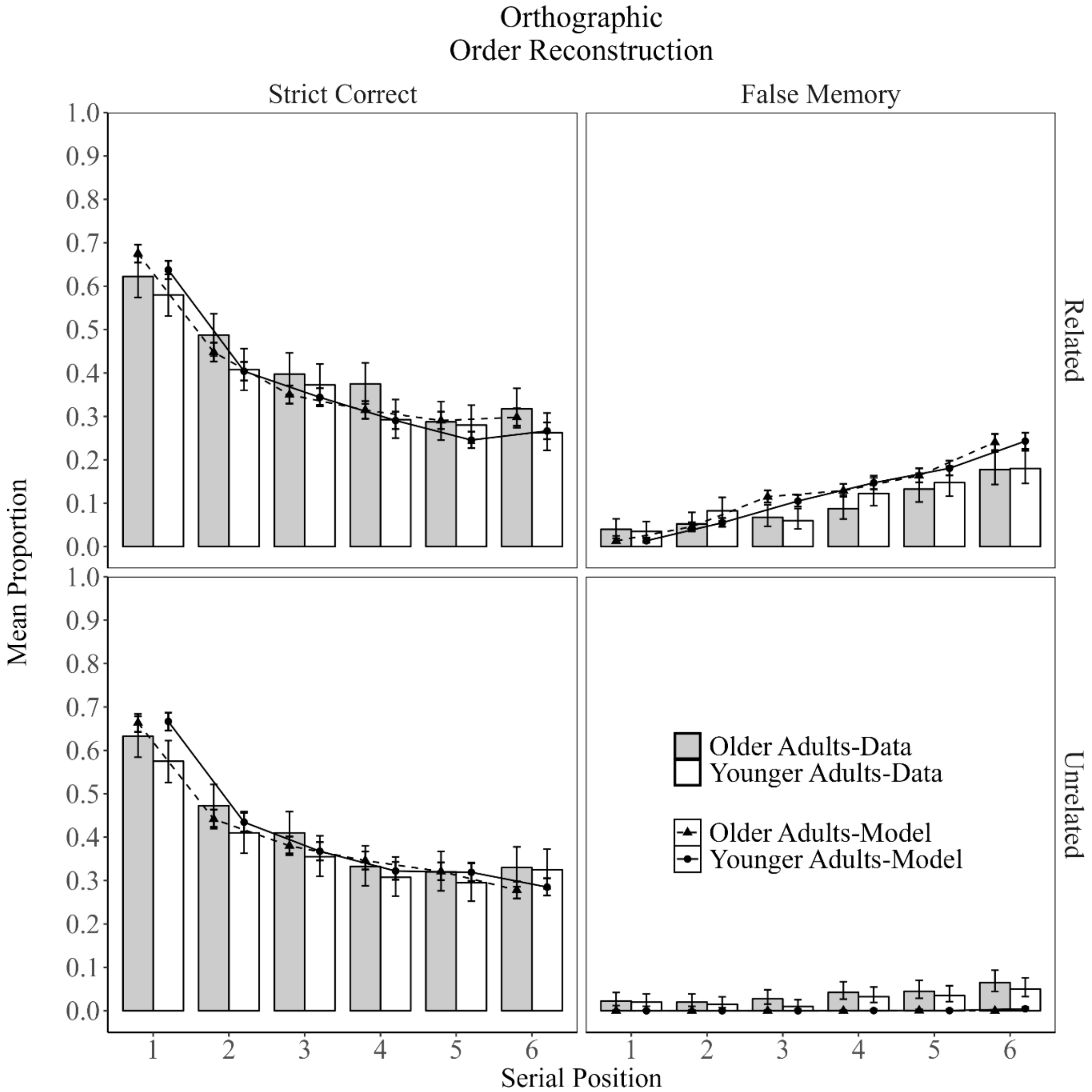
eCFM Simulation of Experiment 3

Here, we evaluate the ability of the eCFM (Guitard et al., 2025a, 2025b, 2025c) to reproduce the key empirical patterns observed in Experiment 3. As in the simulations for Experiments 1 and 2, we used task-specific representations, here, orthographic representations based on the open-bigram scheme from SERIOL and SERIOL2 (Whitney, 2001; Whitney & Marton, 2013). For further methodological details, see orthographic representation in our introduction.

Simulation Parameters. We conducted 100 simulations, each comprising 20 related and 20 unrelated lists, equivalent to 200 simulated participants, each completing 10 related and 10 unrelated lists. For this simulation, orthographic representations were embedded in the eCFM. To reflect the slight descriptive memory advantage observed in older adults, we used $L = .15$ for younger adults and $L = .17$ for older adults. The discriminability parameter d was held constant across age groups at $d = .15$. All other parameter settings were as defined in Experiment 1 and equal across age groups ($g = 0.02$, $s = 1$, $\tau = 3$).

Simulation Results. The simulation results are presented in Figure 3, which displays the proportion of correct responses and false memories as a function of serial position, lure relatedness, and age group (younger vs. older adults), alongside the corresponding empirical data (bars). Overall, the model provided an excellent fit to the data, with only minor discrepancies (for younger adults, $R^2 = .98$ and MAD = .03; for older adults, $R^2 = .97$ and MAD = .03). It successfully captured the serial position curve, showing higher accuracy for early list items and a modest recency effect. The model also reproduced the slight advantage in correct recall for older adults over younger adults, as well as the similar false memory rates observed across age groups. Crucially, like the participants, the model produced more false memories in the related condition than in the unrelated lure condition. However, unlike participants, the model did not produce any false memories for the unrelated critical lure condition. This discrepancy may reflect an idiosyncratic feature of the participants, as the unrelated lures were rarely produced. Another possibility is that participants simply exhibit poor memory for nonwords, leading to extreme forgetting and guessing behaviour, whereas the model treats unrelated nonwords as orthogonal representations, thereby underestimating such response variability.

Figure 3. *Empirical (bars) and computational (lines) results for the mean proportion of correct reconstructions and false memories for the critical lure in Experiment 3 (orthographically related vs. orthographically unrelated), shown for younger and older adults.*



Note. *Error bars represent 95% credible intervals. The left panels display the mean proportion of correct responses as a function of serial position, while the right panels show the mean proportion of false memories by serial position. The top row presents data for the related lure condition, and the bottom row for the unrelated lure condition.*

Discussion

When younger and older adult participants were asked to reconstruct the order of orthographically related nonwords, they consistently produced more false memories for the related, unpresented lure than for the unrelated one. Performance was comparable across age groups, with older adults actually exhibiting, at least descriptively, a slight correct reconstruction *advantage* over younger adults. Both groups showed similar patterns of false reconstruction. The key empirical features were successfully captured by the eCFM: by embedding orthographic representations into the model, we were able to generate specific false memories that closely mirrored participants' behavior. However, the model's failure to falsely remember unrelated nonwords represents a difference between the empirical and simulated outcomes. These findings are consistent with Guitard et al. (2025b) in the context of serial recall but constitute the first application of eCFM to order reconstruction of lists of orthographically related non-words.

The slight correct reconstruction advantage of older adults was accounted for in eCFM by setting the item fidelity parameter *L* (controlling how many lexical and order features of each item are successfully stored in memory) to a slightly higher value for older adults, with no age difference in order discriminability (*d*). This was an unexpected outcome at odds with the age-related reduction in positional precision (lower *d*) of Experiments 1 and 2. One possible explanation is that the abstractness of the nonwords in Experiment 3 suppressed the retrieval of similar features from established long-term memory that could otherwise blur older adults' memories for which items were encountered in which list positions. Support for this view comes from empirical evidence showing no age differences in true and false recognition for similar, abstract stimuli that lack any inherent meaning (Sekuler et al., 2006), but adding semantic labels (i.e., meaning) to abstract stimuli elevates older adults' false recognitions (Koutstaal et al., 2003).

Alternatively, participants of both age groups might have adopted a common encoding strategy that made it easier for young and older adults alike to group (or chunk) the nonwords in a list. For example, given a list of nonwords like "bapte", "dapte," "vapte," etcetera (see Appendix C), participants might have sounded these items out (internally or aloud) – in effect using phonological coding – and detected a potential rhyming pattern that facilitated chunking the items. By contrast, the familiarity of the word stimuli in Experiments 1 and 2 might have yielded a host of different strategies that young and older adults employed, with older adults likely using less effective encoding strategies (e.g., Bailey et al., 2009; Hertzog et al., 1998; Perlmutter, 1978).

One way to tease apart these alternative hypotheses is to examine whether a similar age-related stability in order reconstruction would be obtained with highly similar abstract visual stimuli. If it was the abstractness that eliminated age-related declines in order precision in Experiment 3, then we should expect similar findings in Experiment 4 with abstract visual stimuli.[5] If instead the shared orthography of the nonwords in a list encouraged a similar phonological-based grouping strategy among young and older adults, then we should anticipate that this benefit would be eliminated when participants study abstract images that do not promote such strategies.

Experiment 4

In Experiment 4, we extend both the empirical and computational investigations of temporal order reconstruction to visually related images. This constitutes a powerful test of the generalizability of the eCFM to one of the most widely studied domains of memory research: memory for visual information. Given the abstract nature of the visual stimuli, Experiment 4 also allows us to follow-up on the alternative explanations for the age-related

[5] Indeed, Koutstaal et al. (2003) and Sekuler et al. (2006) both relied on abstract visual stimuli in their experiments and found that there were no age differences in recognition when these stimuli lacked any inherent meaning, supporting the abstractness account. Yet unlike their recognition procedures, we employ a test of order reconstruction, which places more demands on the need to remember the exact order of items rather than the exact identity of the items. Thus, it remains to be seen whether the abstractness account extends fully to order reconstruction.

stability in order precision and slight enhancement in overall item fidelity obtained with abstract orthographically similar non-word stimuli in Experiment 3.

Method

Participants. The eligibility criteria for Experiment 4 were identical to those used in the previous experiments, while extending the exclusion criteria to individuals who had participated in any of the prior experiments. A new sample of 80 participants (40 younger adults and 40 older adults) was recruited via Prolific. In the younger adult group, the average age was 22.45 years (SD = 2.05); 24 participants identified as female and 16 as male. In the older adult group, the average age was 68.53 years (SD = 3.74); 23 participants identified as female and 17 as male.

Materials. In Experiment 4, we created 20 lists, each containing seven abstract images generated using Craiyon, an AI image generator (version 3, https://www.craiyon.com/). Six of the images served as the to-be-remembered items, while the seventh image was designated as the critical related lure. As in the previous experiments, the critical lures were rearranged to form 20 additional lists in which the lure was visually unrelated to the studied items. The names of the image sets are provided in Appendix D, and the full image set is available on the OSF page associated with this manuscript. Half of the participants received the 20 lists in version 1, and half the 20 lists in version 2, in a random order for each participant.

Procedure and Data Analysis. The experimental procedure and data analysis methods for Experiment 4 were identical to those used in the previous experiments, with the exception of the memoranda. In this experiment, participants studied visually related images rather than words (Experiments 1 and 2) or nonwords (Experiment 3).

Results

The proportions of correct responses and false memories for younger and older adults are presented in Figure 4 with the simulation results as a function of serial position

Strict Correct. For strict scoring, performance was comparable between the related (M = .30, SD = .14) and unrelated (M = .32, SD = .14) conditions, $F(1, 78) = 1.86$, $\eta_p^2 = .002$, $BF_{01} = 5.31$. Younger adults (M = .38, SD = .13) outperformed older adults (M = .25, SD = .07), $F(1, 78) = 31.14$, $\eta_p^2 = .29$, $BF_{10} > 10{,}000$. There was no evidence of an interaction between age group and condition, $F(1, 78) = 2.29$, $\eta_p^2 = .03$, $BF_{01} = 2.42$.

False Memory. As in the previous experiments, participants produced more false memories in the related condition (M = .13, SD = .03) than in the unrelated condition (M = .01, SD = .03), $F(1, 78) = 803.85$, $\eta_p^2 = .91$, $BF_{10} > 10{,}000$. False memory rates were slightly lower in younger adults (M = .06, SD = .02) compared to older adults (M = .08, SD = .02), $F(1, 78) = 15.91$, $\eta_p^2 = .17$; however, the Bayesian analysis provided evidence in favor of the absence of such difference, $BF_{01} = 5.57$. There was no evidence for an interaction between age group and condition, $F(1, 78) = 2.02$, $\eta_p^2 = .03$, $BF_{01} = 17.64$.

Empirical Summary

In an order reconstruction test for visually similar abstract items, young adults exhibited higher rates of correct reconstruction than older adults did, but there was no age difference in false reconstruction rates. In both age groups, false reconstructions were higher for visually related lures than for unrelated lures. In the following section, we assess whether the eCFM can account for these key findings.

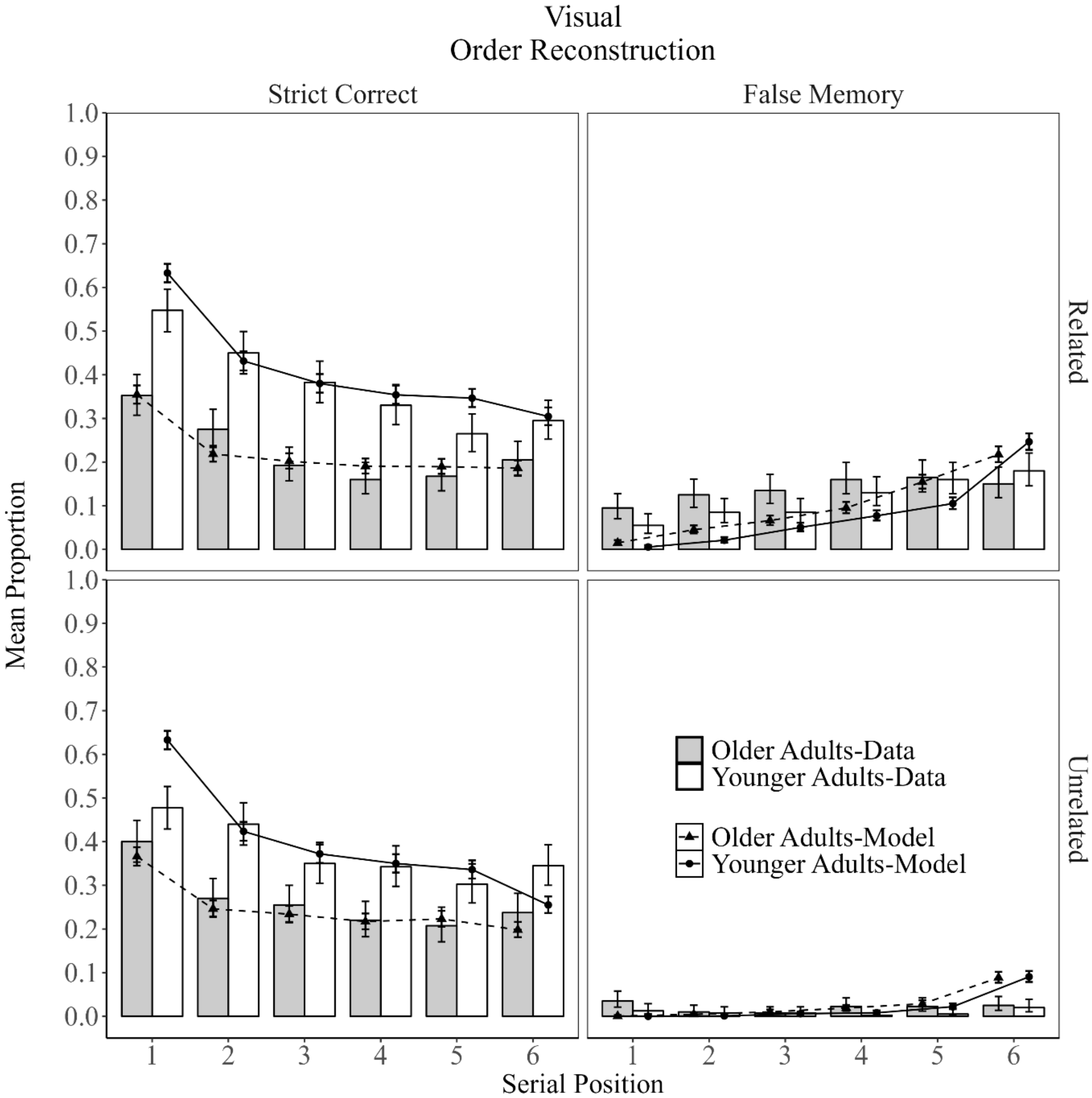
eCFM Simulation of Experiment 4

As in the previous experiments, we evaluated the extent to which the eCFM (Guitard et al., 2025a, 2025b, 2025c) could simulate the core empirical patterns observed in Experiment 4. In this case, the model incorporated visual representations to reflect the nature of the image-based task (see the Introduction for details on how these visual representations were constructed).

Simulation Parameters. We ran 100 simulations, each consisting of 20 related and 20 unrelated lists, representing 200 simulated participants who completed 10 related and 10 unrelated trials each. To model the young adult data, we set $L = .15$ and $d = .15$, but to model the older adult data, we reduced these values to $L = .10$ and $d = .05$ to capture the relative recall disadvantage. Additionally, to reflect the increased difficulty of encoding abstract images and relatively flat serial position curves, we lowered g (the decay of representational fidelity over serial positions) to .005 in both groups, minimizing the influence of serial position on encoding. Otherwise, we retained the same parameter settings as in the previous experiments, with $s = 1$ and $\tau = 3$.

Simulation Results. Figure 4 presents the model's predictions for correct responses and false memories across serial positions, lure type, and age groups, alongside the empirical data (bars). The model captured several key aspects of participant behavior, including the relatively flat serial position curve and the superior recall performance of younger adults. While it slightly overestimated the recall advantage of younger adults, the overall trends aligned well with the data. The model also reproduced the critical false memory pattern: higher rates of false recall in the related condition than in the unrelated one. However, it slightly underestimated the false memories in the related condition and slightly overestimated them in the unrelated condition. Despite these minor deviations, the model demonstrated good alignment with the empirical data: for younger adults, $R^2 = .93$ and MAD = .04; for older adults, $R^2 = .87$ and MAD = .03.

Figure 4. *Empirical (bars) and computational (lines) results for the mean proportion of correct reconstructions and false memories for the critical lure in Experiment 4 (visually related vs. visually unrelated), shown for younger and older adults.*



Note. *Error bars represent 95% credible intervals. The left panels display the mean proportion of correct responses as a function of serial position, while the right panels show the mean proportion of false memories by serial position. The top row presents data for the related lure condition, and the bottom row for the unrelated lure condition.*

Discussion

When younger and older adult participants were asked to reconstruct the order of visually related abstract images, they consistently produced more false reconstructions for unstudied lures that were related versus unrelated to these studied images. Although the rate of false reconstructions did not differ between age groups, younger adults were more accurate in their correct temporal order reconstruction. Model-based simulations demonstrated that embedding task-relevant visual representations into the eCFM allowed it to generate specific true and false reconstructions that closely reflected human performance. To account for the age-related decline in true reconstruction, the eCFM simulated lower estimates of L and d in older adults. Thus, the eCFM attributed these declines to both a reduction in the fidelity of item representations (i.e., fewer visual and positional features of each item were faithfully encoded into memory) and in the precision of positional features that were encoded (i.e., resulting in relatively the same limited positional features being used diffusely to represent the positions of multiple items). That such differences were observed with abstract visual stimuli argues against the possibility that the age-related stability in these representations for abstract nonword stimuli in Experiment 3 was due to the abstractness of the stimuli.

The fact that different age-related patterns emerged across these experiments implicates some other process beyond abstractness is necessary to explain the puzzling exception to the rule in Experiment 3. For instance, focusing on the common features of the orthographically similar nonwords in a list (e.g., that all the items in a specific list end in "azy") could have contributed to a phonological-based rehearsal strategy that was equally advantageous for young and older adults. Although older adults might similarly have prioritized focusing on shared global features of the visually similar abstract images in Experiment 4 (e.g., the presence of a pink swirl), these abstract images likely exist in a lower dimensional space to a human observer (e.g., Kahana et al., 2007), reducing the number of unique, discriminable features of the stimuli that an older adult could perceive. By contrast, even if the orthographically related nonword stimuli exist in a similarly low dimensional feature space, as modeled in the eCFM, some of these dimensions could offer more diagnostic information to a human observer (e.g., the presence of a given letter like *s* or *h*). We turn now to a summary of our empirical and computational modeling results of order reconstruction from Experiments 1-4.

Summary of Order Reconstruction

Across semantically (Experiment 1), phonologically (Experiment 2), orthographically (Experiment 3), and visually (Experiment 4) related materials, younger and older adults made comparable numbers of false memory errors in temporal order reconstruction. Both groups were more prone to falsely assigning unstudied but similar items to a list position than they were to reconstructing unstudied, unrelated items. Although no age differences in false reconstruction were detected, there were typically modest age differences in correct reconstruction. In most cases, the expected pattern was obtained, with young adults outperforming older adults. The lone exception was in Experiment 3, with orthographically similar items, in which older adults slightly outperformed younger adults. Critically, the eCFM was able to account for all the major findings in each experiment through its use of structured representations. It successfully captured the higher rates of false reconstruction for similar than unrelated lures, along with the serial position curves and age differences in correct temporal order reconstruction.

For lists of semantically (Experiment 1) or phonologically (Experiment 2) similar words, eCFM attributed the subtle age-related declines in correct order reconstruction to reduced precision of the positional representations with no concomitant reduction in the overall fidelity of item representations. Thus, while the eCFM assumed an equal proportion of lexical (semantic or phonological) and order features were stored for each item in each age group ($L_{\text{old}} = L_{\text{young}}$), it required that the features that represented each item's position in the list were more similar across items among older adults ($d_{\text{old}} < d_{\text{young}}$). No such impairments were necessary to capture the age-related stability of order reconstruction for lists of orthographically similar nonwords in Experiment 3. In fact, in this instance, the eCFM assumed that older adults represented each item with slightly *more* features than younger adults did ($L_{old} > L_{young}$). Finally, for lists of visually similar abstract images (Experiment 4), the eCFM attributed age-related declines in order reconstruction to reduced global fidelity ($L_{old} < L_{young}$) and positional precision ($d_{old} < d_{young}$).

Overall, the eCFM simulations reinforce claims that older adults' episodic memories are generally fuzzier in nature (Brainerd & Reyna, 2015; Greene & Naveh-Benjamin, 2023), while providing a more nuanced perspective. It is not simply the case that *all* aspects of an older adult's memory are universally fuzzier than those of a younger adult. When the purpose of remembering is to decide the order in which items occurred without needing to reconstruct memory for the items themselves, older adults often retain the same number of features of each item in memory but use more of the same features to represent the positions of the items. That is, their memory for the order of the items, but not strictly speaking for the items themselves, is fuzzier. As noted above, there were exceptions to this pattern for lists of abstract stimuli, like orthographically similar nonwords (no age-related representational impairment) or visually similar abstract images (age-related reduction in overall item fidelity and in the precision of order representations).

Having established that the eCFM can successfully account for both true and false temporal order reconstruction across various representational modalities and age groups, we turn now to investigating its ability to explain true and false serial recall and age differences therein.

Serial Recall

Building on its success in capturing young and older adults' true and false temporal order reconstructions across lists of semantically (Experiment 1), phonologically (Experiment 2), orthographically (Experiment 3), or visually (Experiment 4) similar items, a natural extension of this work is to apply the eCFM to serial recall, a task for which the eCFM has previously successfully accounted for the performance of younger adults (Guitard et al., 2025a, 2025b, 2025c). Serial recall places similar demands as order reconstruction on the need to remember the precise order of events, but unlike order reconstruction, the studied items are not re-presented at test. Instead, participants must recall the items in the exact order of presentation, typically by typing them or saying them aloud. Thus, serial recall requires that participants remember both item and order information with no cues provided. This increased demand with reduced environmental support could exacerbate age differences that were not as pronounced on tests of order reconstruction (e.g., Craik, 1983). Indeed, older adults often exhibit worse performance than young adults in serial recall accuracy (Golomb et al., 2008; Maylor et al., 1999).

In Experiments 5 through 7, we systematically investigate true and false serial recall among younger and older adults across DRM-style lists (Deese, 1959; Roediger & McDermott, 1995) of semantically (Experiment 5), phonologically (Experiment 6), or orthographically (Experiment 7) related items. Given the recall nature of the task, we did not include lists of visually related items. Our goals were to systematically assess age-related differences in true and false recall across different types of study material and to evaluate whether and how the eCFM can accurately reproduce the observed behavioral patterns in serial recall as in order reconstruction.

A related goal was to assess whether the same representational processes that accounted for age-related differences on order reconstruction would also explain age differences in serial recall, particularly for matched stimulus types. For instance, the eCFM identified that age-related declines in order reconstruction for lists of semantically or phonologically similar items stemmed from poorer precision of order representations (i.e., many of the same features were used to represent different serial positions), with no overall reduction in the fidelity of encoding lexical and order features of items (i.e., the same overall number of features were used to represent each item across age groups). Here we ask whether age differences in serial recall for semantically and phonologically similar lists also arise solely from reduced discriminability of order representations or whether the need to reconstruct the items themselves yields an additional deficit in the global fidelity of item representations, as previously assumed by SIMPLE (Surprenant et al., 2006) and the Feature Model (Neath & Surprenant, 2007). Meanwhile, we also ask if older adults still have a slight advantage in the fidelity of their underlying item representations for orthographically similar nonwords, as was observed in order reconstruction, when the task now requires the reconstruction of the items in addition to their orders? As these questions reveal, there is much to be learned by attempting to explain age differences across different memory tasks with a single model.

Experiment 5

In Experiment 5, we examined serial recall of semantically related word lists. Participants studied lists of six words, each semantically related to an unstudied, unpresented critical lure. After a 6-second parity judgment task, participants were asked to recall the words in the order in which they were presented. The goal was to examine age-related differences in true and false recall between younger and older adults and to assess whether the model could successfully capture these findings.

Method

Participants. The sample size, recruitment procedure, and inclusion criteria for Experiment 5 were identical to those used in the previous four experiments, with the added requirement that individuals who had participated in any earlier experiments were excluded. A new sample of 81 participants (41 younger adults and 40 older adults) was recruited via Prolific. One additional younger adult was included due to slight over-recruitment. In the younger adult group, the mean age was 22.73 years (SD = 2.03); 25 participants identified as female, 15 as male, and 1 preferred not to specify their gender. In the older adult group, the mean age was 68.30 years (SD = 2.86), with 30 participants identifying as female and 10 as male.

Materials and Design. The stimuli were the 20 related lists used in Experiment 1, with each list consisting of six words thematically related to an unpresented critical lure (see Appendix A). All participants in Experiment 5 were tested on these 20 related lists.

Procedure. The procedure was identical to that used in Experiment 1, except for the memory test. Immediately following the parity judgment task, participants were prompted with the instruction “Type the first word” displayed above an empty blue rectangle where they typed their response. After each response, participants pressed the Enter key to register their answer. The prompt then updated to “Type the second word,” and this continued until all six responses were entered. As in previous experiments, participants were not allowed to go back and change previous answers.

Data Analysis. All serial recall data, along with R Markdown files for analyses and modeling, are available on the OSF page linked to this project.

Scoring. In Experiments 5 to 7, we applied a strict spelling criterion, whereby a response was considered correct only if it exactly matched the target word’s spelling (Guitard et al., 2025b). Serial recall provides rich data, which we leveraged to extract several key measures. A response was classified as *strict correct* when the word was recalled in the exact serial position in which it was presented. An *intralist error* occurred when a studied word was recalled but placed in the wrong position, for example, recalling the first word in the second serial position. *Omissions* were defined as instances in which no response was given, including blanks or placeholders such as “don’t know” or “NA”; these were coded as 0 in the dataset. *False memories* or *critical lure errors* were instances in which the unpresented but semantically related lure associated with the studied list was recalled. Finally, *extralist intrusions* referred to recalled words that were neither studied nor the critical lure but were part of a participant’s (and the model’s) lexicon. To ensure consistency with prior experiments, where repeated responses were not possible, we counted only the first occurrence of each critical lure or extralist intrusion. For example, if the extralist error “cat” was recalled multiple times, only its first appearance was included in the analysis. This approach allows for direct comparability across experiments while preserving the interpretive clarity of our recall measures.

Statistical Analysis. Statistical analyses followed a similar approach to that used in previous experiments. However, because there was no unrelated condition in this experiment, analyses focused on age group (younger vs. older adults) using both frequentist and Bayesian analyses.

Results

The empirical and computational findings are presented in Figure 5, which shows the proportions of correct recall, intralist errors, omission errors, false memories (i.e., critical lures), and extralist errors for younger and older adults. These results are plotted across serial positions.

Strict Correct. Younger adults showed higher accuracy in recalling items in their original positions (M = .57, SD = .14) compared to older adults (M = .47, SD = .18). The results from the analysis of variance confirmed these trends. There was a main effect of age group, $F(1, 79) = 7.89$, $\eta_p^2 = .091$, $BF_{10} = 3.094$, and a main effect of serial position, $F(5, 395) = 281.56$, $\eta_p^2 = .781$, $BF_{10} > 10{,}000$. There was no interaction between age group and serial position, $F(5, 395) = 2.01$, $\eta_p^2 = .025$, $BF_{01} = 27.75$.

Intralist Errors. Younger adults made slightly fewer intralist errors (M = .18, SD = .08) than older adults (M = .21, SD = .10), but this difference was not significant, $F(1, 79) = 1.99$, $\eta_p^2 = .025$, $BF_{01} = 6.883$. There was a strong main effect of serial position, $F(5, 395) = 56.16$, $\eta_p^2 = .416$, $BF_{10} = 254.331$. There was also clear evidence for an interaction between age group and serial position, $F(5, 395) = 4.34$, $\eta_p^2 = .052$, $BF_{10} > 10{,}000$.

Omission Errors. The rate of omission errors was descriptively lower for younger adults (M = .09, SD = .11) compared to older adults (M = .16, SD = .16), but the ANOVA produced weak evidence *against* this effect, $F(1, 79) = 4.44$, $\eta_p^2 = .053$, $BF_{01} = 1.057$. There was a strong main effect of serial position, $F(5, 395) = 60.50$, $\eta_p^2 = .434$, $BF_{10} > 10{,}000$. There was also clear evidence for an interaction between age group and serial position, $F(5, 395) = 2.60$, $\eta_p^2 = .032$, $BF_{10} > 10{,}000$.

False Memories (Critical Lures). Younger adults recalled the critical lure slightly more often (M = .015, SD = .014) than older adults (M = .010, SD = .011) at the descriptive level. However, the analysis of variance showed no clear effect of age group, $F(1, 79) = 3.30$, $\eta_p^2 = .040$, $BF_{01} = 16.356$, and no effect of serial position, $F(5, 395) = 3.47$, $\eta_p^2 = .042$, $BF_{01} = 12.061$. There was no interaction between age group and serial position, $F(5, 395) = 1.11$, $\eta_p^2 = .014$, $BF_{01} = 3792.93$.

Extralist Errors. Younger adults (M = .093, SD = .07) and older adults (M = .092, SD = .08) produced a comparable number of extralist errors. The results from the analysis of variance showed no main effect of age group, $F<1$, $\eta_p^2 = .000$, $BF_{01} = 13.468$, a strong main effect of serial position, $F(5, 395) = 22.51$, $\eta_p^2 = .222$, $BF_{10} > 10{,}000$, and no interaction between age group and serial position, $F(5, 395) = 0.33$, $\eta_p^2 = .004$, $BF_{01} > 10{,}000$.

Empirical Summary

Overall, the results indicate that younger adults outperformed older adults under strict scoring criteria. However, for intralist errors, omission errors, false memories (critical lures), and extralist errors, there were no credible overall age-related differences. Closer inspection of the interactions revealed that older adults were more prone to making intralist errors, as serial position increased, suggesting that associative or order information became less distinct later in the list. Older adults also tended to omit more items toward the end of the sequence, indicating a sharper decline in recall for later list positions. In the following section, we evaluate whether the eCFM can account for these key empirical patterns.

eCFM Simulation of Experiment 5

Here, we assess the ability of the eCFM (Guitard et al., 2025a, 2025b, 2025c) to reproduce the core empirical patterns shown in Figure 5. Building on our prior work, the simulations presented here employed representations specifically suited to the task, in this case, semantic representations. We sought to capture age-related differences in memory performance in serial recall through the same representational parameters that governed age differences in temporal order reconstruction in Experiments 1–4: L, the base encoding rate dictating the fidelity of item representations, and d, which governs how distinctive are the features that represent the orders of each item. Unless otherwise noted, all other model parameters were held constant across age groups. As before, the empirical data are drawn as bar plots, with the corresponding model simulations overlaid for direct comparison.

The eCFM adopts a nearly identical process for modelling serial recall as it uses to model order reconstruction (see Equation 3 and Table 1). The key difference rests in the size of the memory search set over which similarities to the echo are computed. In order reconstruction, eCFM restricts its focus to the subset of studied words (plus one foil), selecting, without repetition, the word from this subset that has the highest similarity to the echo generated for each serial position. In serial recall, the model compares the echo to all words in the lexicon. The word with the highest cosine similarity to the echo is selected for recall, provided that its similarity exceeds a predefined threshold. If no word exceeds this threshold, the model produces an omission. Given the expanded search set in serial recall, it is necessary to define the model's lexicon.

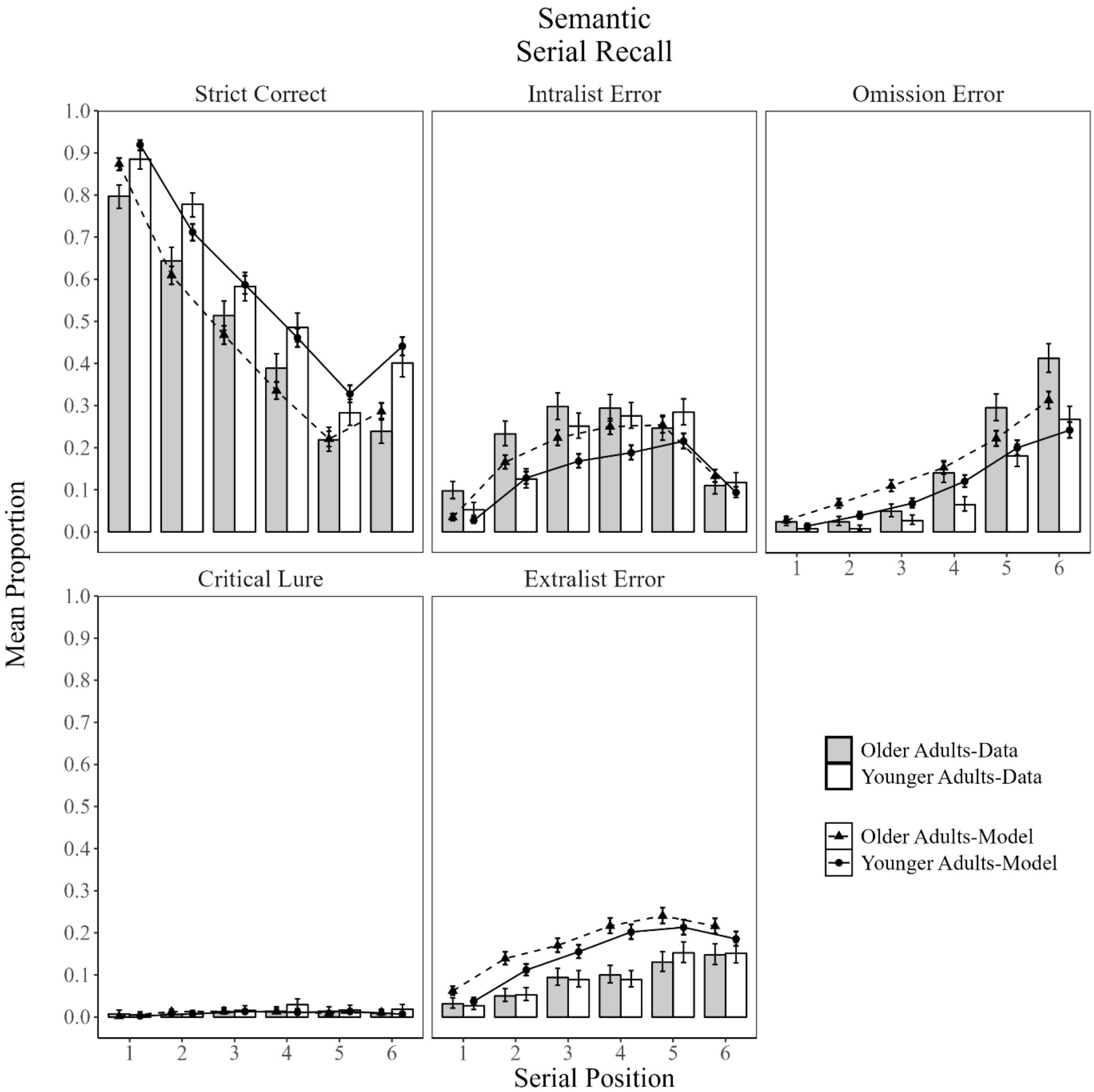
Simulation Lexicon. Following the approach of Guitard et al. (2025a, 2025b, 2025c), we first constructed a semantic lexicon consisting of 95,540 words derived from the Touchstone Applied Science Associates (TASA) corpus. Semantic representations for these words were generated using Latent Semantic Analysis (LSA; Landauer & Dumais, 1997), a well-established method that captures word meaning based on patterns of word co-occurrence in large text corpora. The TASA corpus offers a broad and representative sample of general English language usage, making it suitable for modeling the linguistic environment typically encountered by participants in verbal memory tasks.

This initial lexicon was then refined using frequency and part-of-speech information from the SUBTLEXus database (Brysbaert et al., 2012). Specifically, we retained words with Zipf frequencies between 1 and 7 and limited inclusion to parts of speech commonly recalled in memory experiments (e.g., adjectives, adverbs, names, nouns, numbers, verbs, and interjections). Additionally, 442 high-frequency words exhibiting excessive co-occurrences were excluded to avoid distortion in vector similarity (see OSF for the full list). The final lexicon contained 50,331 words, each represented as a unique 300-dimensional semantic vector. This lexicon was intended to reflect the linguistic environment of typical participants in serial recall experiments.

Simulation Parameters. For each experiment, we conducted 100 simulation runs, each involving 20 related lists, approximating a sample size of 100 participants per age group. In Experiment 5, where semantic representations were embedded within the eCFM, we adjusted the following parameter values to capture the performance of younger and older adults: for younger adults, $L = 0.26$ and $d = 0.28$; for older adults, $L = 0.24$ and $d = 0.25$. These values reflect slightly stronger encoding and greater positional discriminability in younger adults. The following parameters were held constant across groups: $g = 0.03$, $s = 0.01$, $\tau = 3$, and the recall threshold $T = 0.35$.

Simulation Results. As shown in Figure 5, the model captures many of the key empirical patterns, with good fits for both younger adults ($R^2 = .96$, MAD = .04) and older adults ($R^2 =.92$, MAD = .05), despite some minor discrepancies. Specifically, the model reproduces the serial position curve, characterized by enhanced recall for early list items and a modest recency boost. It also captures the intralist error pattern, with higher rates of positional confusion for middle serial positions, as well as the gradual increase in omissions over the course of the list. The low prevalence of false memories is also well reflected, along with a tendency for extralist intrusions to occur more frequently at later positions, though these are somewhat overpredicted by the model. Finally, the model successfully simulates age-related differences, including higher proportions of correct recall for younger adults, and subtle group differences in omission and intralist error rates.

Figure 5. *Empirical results (bars) and computational results (lines) showing the mean proportions of correct recall, intralist errors, omission errors, false memories (critical lures), and extralist errors for younger and older adults as a function of serial position for Experiment 5.*



Note. *Error bars represent 95% credible intervals.*

Discussion

When younger and older adults were asked to recall in order six previously studied, semantically related words following a brief parity judgment task, they produced comparable levels of false memories, both in terms of critical lures and other non-studied intrusions. However, a reliable age difference emerged in correct recall, with younger adults outperforming older adults. Crucially, by incorporating semantic representations into the eCFM, we were able to replicate all major patterns of memory performance across age groups. Consistent with order reconstruction of semantically similar lists, the eCFM implicated a slight age-related reduction in the precision of positional encoding (parameter *d*). However, unlike in order reconstruction, in serial recall of lists of semantically similar items, the eCFM also identified a slight age-related reduction in the overall strength of item encoding (parameter *L*) dictating how many semantic and order features were used to represent each item. Thus, when the task of memory is to remember both the items and their order without being provided the items as cues, older adults retain fewer features of each item in memory while also representing the positions of each item less distinctively.

Experiment 6

The results of Experiment 5 were clear: younger adults outperformed older adults in correct serial recall, but both groups produced a relatively similar number of false recalls when presented with lists that were semantically related to an unstudied critical lure. The eCFM attributed these age differences to reductions in item fidelity and in the precision with which the positions associated with each word were represented. In Experiment 6, we extended this investigation, paralleling the approach taken in Experiment 2, by using phonologically related word lists.

Method

Participants. The sample size, recruitment procedure, and inclusion criteria for Experiment 6 were identical to those used in the previous experiments. To avoid repeat participation, individuals who took part in earlier experiments were excluded and thus not eligible for this study. A new sample of 80 participants was recruited through Prolific, comprising 40 younger adults and 40 older adults. In the younger adult group, the mean age was 22.66 years (SD = 1.84); 29 participants identified as female and 11 as male. In the older adult group, the mean age was 70.12 years (SD = 3.88), with 26 participants identifying as female and 14 as male.

Materials, Design, Procedure, and Data Analysis. The materials were identical to the phonologically related word lists used in Experiment 2. Each of 20 lists contained six words that were phonologically related to an unpresented critical lure (see Appendix B). All participants in Experiment 6 completed memory tests based on these 20 related lists. Otherwise, the design, procedure, and data analysis were identical to those used in Experiment 5.

Results

Figure 6 presents both the empirical and simulated results, depicting the proportions of correct recall, intralist errors, omissions, false memories (critical lures), and extralist errors for younger and older adults across serial positions.

Strict Correct. As in Experiment 5, younger adults showed higher accuracy in recalling items in their original positions (M = .37, SD = .15) compared to older adults (M = .26, SD = .15). The results from the analysis of variance confirmed these trends. There was a main effect of age group, $F(1, 79) = 10.39$, $\eta_p^2 = .116$, $BF_{10} = 8.55$, and a strong main effect of serial position, $F(5, 395) = 239.50$, $\eta_p^2 = .752$, $BF_{10} > 10{,}000$. There was no interaction between age group and serial position, $F<1$, $\eta_p^2 = .005$, $BF_{01} > 10{,}000$.

Intralist Errors. Younger adults made fewer intralist errors (M = .23, SD = .08) than older adults (M = .27, SD = .13) at the descriptive level. However, the results from the analysis of variance showed no main effect of age group, $F(1, 79) = 2.71$, $\eta_p^2 = .033$, $BF_{01} = 4.760$, but a strong main effect of serial position, $F(5, 395) = 46.76$, $\eta_p^2 = .372$, $BF_{10} > 10{,}000$. There was also no clear evidence for an interaction between age group and serial position, $F(5, 395) = 2.61$, $\eta_p^2 = .032$, $BF_{10} = 1.179$.

Omission Errors. Omission rates were similar between younger adults (M = .11, SD = .15) and older adults (M = .14, SD = .18). There was no main effect of age group, F<1, η_p^2 = .009, BF_{01} = 4.75, and a strong main effect of serial position, F(5, 395) = 43.28, η_p^2 = .354, BF_{10} > 10,000. There was also clear evidence for an interaction between age group and serial position, F(5, 395) = 1.40, η_p^2 = .017, BF_{10} > 10,000.

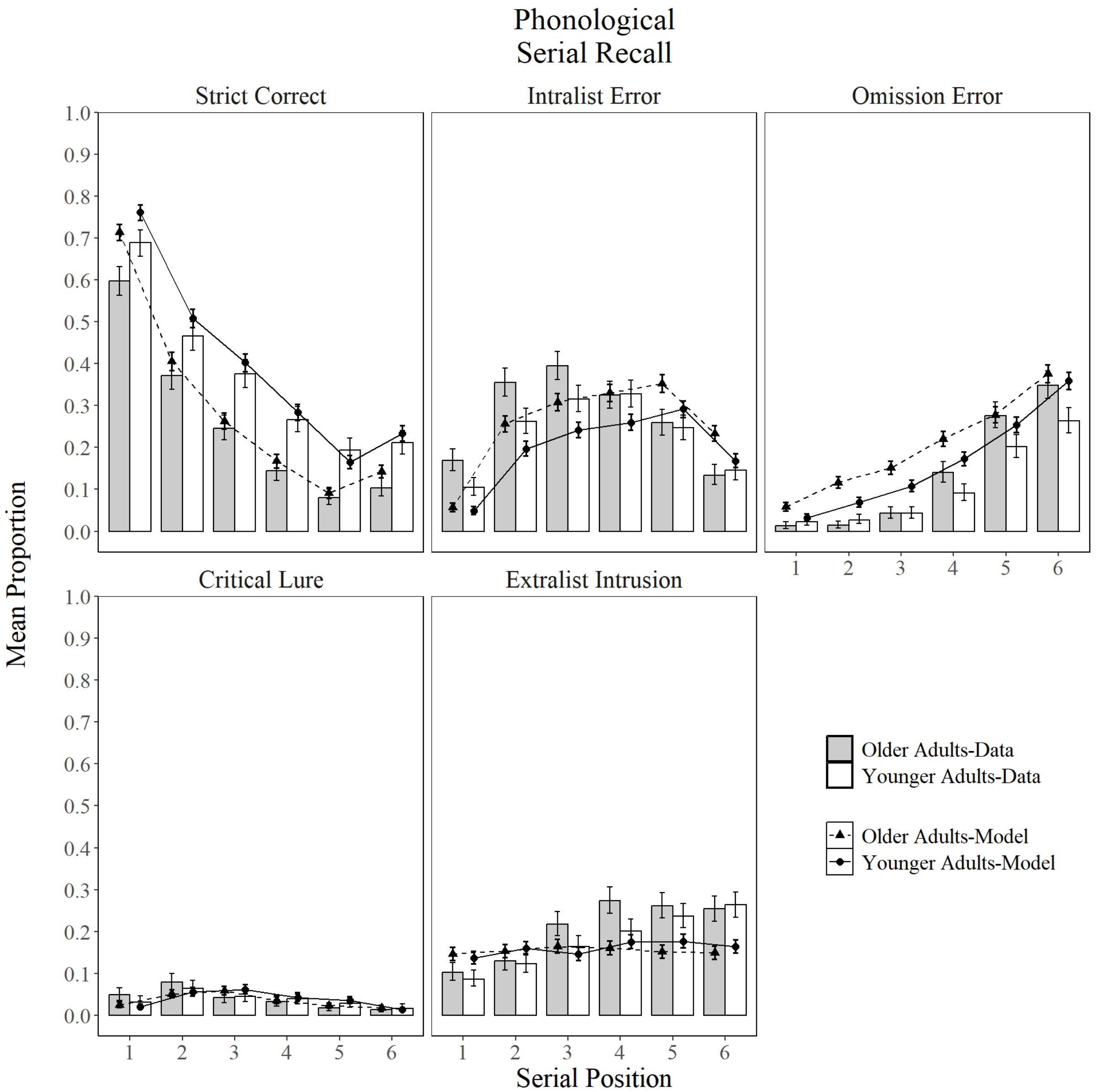
False Memories (Critical Lures). The rate of critical-lure recall was comparable between younger (M = .037, SD = .014) and older (M = .039, SD = .015) adults. The analysis of variance showed no main effect of age group, F<1, η_p^2 = .003, BF_{01} = 22.77, and a main effect of serial position, F(5, 395) = 14.90, η_p^2 = .159, BF_{10} > 10,000. There was no interaction between age group and serial position, F(5, 395) = 1.22, η_p^2 = .015, BF_{01} = 3105.64

Extralist Errors. Younger adults produced a similar number of extralist intrusions (M = .18, SD = .12) as older adults (M = .21, SD = .15). The results from the analysis of variance showed no main effect of age group, F<1, η_p^2 = .011, BF_{01} = 7.346, a strong main effect of serial position, F(5, 395) = 24.59, η_p^2 = .237, BF_{10} > 10,000, and no interaction between age group and serial position, F(5, 395) = 1.23, η_p^2 = .015, BF_{01} = 111.967.

Empirical Summary

As observed in Experiment 5, younger adults outperformed older adults in correct recall. However, the two groups performed similarly overall on intralist errors, omissions, false memories, and extralist intrusions. Once again, we found an interaction for omissions, with older adults making more errors toward the end of the lists. Next, we assess whether the eCFM can account for these findings.

Figure 6. *Empirical results (bars) and computational results (lines) showing the mean proportions of correct recall, intralist errors, omission errors, false memories (critical lures), and extralist errors for younger and older adults as a function of serial position for Experiment 6.*



Note. *Error bars represent 95% credible intervals.*

eCFM Simulation of Experiment 6

We now assess the ability of the eCFM (Guitard et al., 2025a, 2025b, 2025c) to simulate the key empirical patterns presented in Figure 6. As in Experiment 5, the simulations employed representations tailored to the demands of the task, in this case, phonological representations.

Simulation Lexicon. The phonological lexicon (containing only the phonological representations of the words) was refined using the same procedure as in Experiment 5. The final lexicon comprised 41,005 words, each represented by a unique 300-dimensional vector, following the approach inspired by Parrish (2017).

Simulation Parameters. As before, we conducted 100 simulations using 20 phonologically related lists per run, approximating a sample size of 100 participants. To capture the performance of young and older adults, we set to L = 0.23 and d = 0.24 for younger adults, and L = 0.21 and d = 0.20 for older adults. These values were slightly lower than those used in Experiment 5, reflecting the increased difficulty of the phonologically related lists. The remaining parameters were held constant across age groups and matched those from Experiment 5: g = 0.03, s = 0.01, τ = 3, and T = 0.35.

Simulation Results. As shown in Figure 6, the model captures many of the key empirical patterns, with a better fit for younger adults (R^2 = .91, MAD = .04) compared to older adults (R^2 =.79, MAD = .05). Specifically, the model successfully reproduces the serial position curve, marked by better recall of early list items and recency effect. It also captures the intralist error pattern, with increased positional confusion occurring in the middle of the list, and it reflects the gradual rise in omissions across serial positions. Some discrepancies remain, particularly for older adults, where the model underpredicts intralist errors at positions 2 and 3 and slightly overpredicts correct recall. It also produces more omissions than observed in the empirical data for both age groups. While the model closely aligns with the observed pattern for critical lures, it tends to underpredict extralist errors. Despite these minor deviations, the model broadly captures the overall trends in the data.

Discussion

In a serial recall task for short lists of phonologically related items, younger and older adults produced comparable levels of false recall, but younger adults outperformed older adults in correct recall. By incorporating phonological representations into the eCFM, we were able to replicate many of the key patterns and age-related differences. These differences were accounted for by reducing the *L* (global fidelity) and *d* (positional precision) parameters for older adults, as in serial recall of semantically similar word lists (Experiment 5). Collectively, the results of Experiments 5 and 6 (serial recall), in comparison with the matched stimulus types in Experiments 1 and 2 (order reconstruction), highlight a more general reduction in the precision of order/context representations with aging that extends across task formats, but age differences in the overall fidelity of item representations (i.e., how many lexical and order features are used to represent an item) for semantically and phonologically similar words are task-dependent. These insights were only made possible by jointly modelling age differences in temporal order reconstruction and serial recall across these stimulus types with the same computational model.

Experiment 7

We turn now to an empirical and computational investigation of true and false serial recall, and age differences therein, for lists of orthographically similar items. We previously observed that the eCFM could successfully capture key behavioral metrics of order reconstruction for these items (Experiment 3), including the stability of order reconstruction across age groups. Now we evaluate its ability to capture patterns of serial recall differences between young and older adults. A key goal of this modelling is to reveal whether the demand to retrieve both the items and their order would manifest age differences in the fidelity of item representations and/or in the precision with which order is represented for orthographically similar nonwords, as it does for semantically (Experiment 5) and phonologically (Experiment 6) related words. Such a finding would be at odds with the enhanced fidelity of item representations and preserved positional discriminability among older adults when the task of memory is simply to reorder a small subset of available items into their original positions (Experiment 3).

Method

Participants. The sample size, recruitment procedure, and inclusion criteria for Experiment 7 mirrored those used in previous experiments. Participants from earlier studies were excluded. A total of 81 participants were recruited via Prolific, consisting of 41 younger adults and 40 older adults. One additional younger adult was included due to slight over-recruitment. In the younger group, the mean age was 22.46 years (SD = 2.44), with 18 participants identifying as female and 23 as male. In the older adult group, the mean age was 69.25 years (SD = 3.61), with 29 participants identifying as female and 11 as male.

Materials, Design, Procedure, and Data Analysis. The materials, design, procedure, and data analysis in Experiment 7 closely followed those of Experiments 5 and 6, with one notable modification. Instead of phonologically or semantically related words, participants studied the 20 orthographically related nonword lists previously used in Experiment 3. Each list consisted of six nonwords that shared orthographic similarity with an unpresented critical lure (see Appendix C). All participants completed a serial recall memory task based on these 20 lists.

Results

Figure 7 summarizes the empirical and computational data, displaying the proportions of correct recall, intralist errors, omissions, false memories (critical lures), and extralist errors for younger and older adults across serial positions.

Strict Correct. As in Experiments 5 and 6, under strict scoring criteria, younger adults demonstrated higher accuracy (M = .27, SD = .26) than older adults (M = .13, SD = .12). The analysis of variance confirmed this pattern, revealing a main effect of age group, $F(1, 79) = 10.10$, $\eta_p^2 = .113$, $BF_{10} = 12.86$, and a strong main effect of serial position, $F(5, 395) = 74.65$, $\eta_p^2 = .486$, $BF_{10} > 10{,}000$. There was no interaction between age group and serial position, $F<1$, $\eta_p^2 = .009$, $BF_{01} > 10{,}000$.

Intralist Errors. Younger adults made a comparable number of intralist errors (M = .17, SD = .08) to older adults (M = .19, SD = .11). The analysis of variance showed no main effect of age group, $F<1$, $\eta_p^2 = .007$, $BF_{01} = 11.31$, and a strong main effect of serial position, $F(5, 395) = 22.73$, $\eta_p^2 = .223$, $BF_{10} > 10{,}000$. There was no interaction between age group and serial position, $F<1$, $\eta_p^2 = .005$, $BF_{01} > 10{,}000$.

Omission Errors. As in Experiment 6, omission rates were similar across groups, with younger adults (M = .12, SD = .19) and older adults (M = .08, SD = .18). The analysis of variance showed no main effect of age group, $F<1$, $\eta_p^2 = .008$, $BF_{01} = 3.87$, and a strong main effect of serial position, $F(5, 395) = 16.94$, $\eta_p^2 = .177$, $BF_{10} > 10{,}000$. There was an interaction between age group and serial position, $F<1$, $\eta_p^2 = .010$, $BF_{10} = 187.27$.

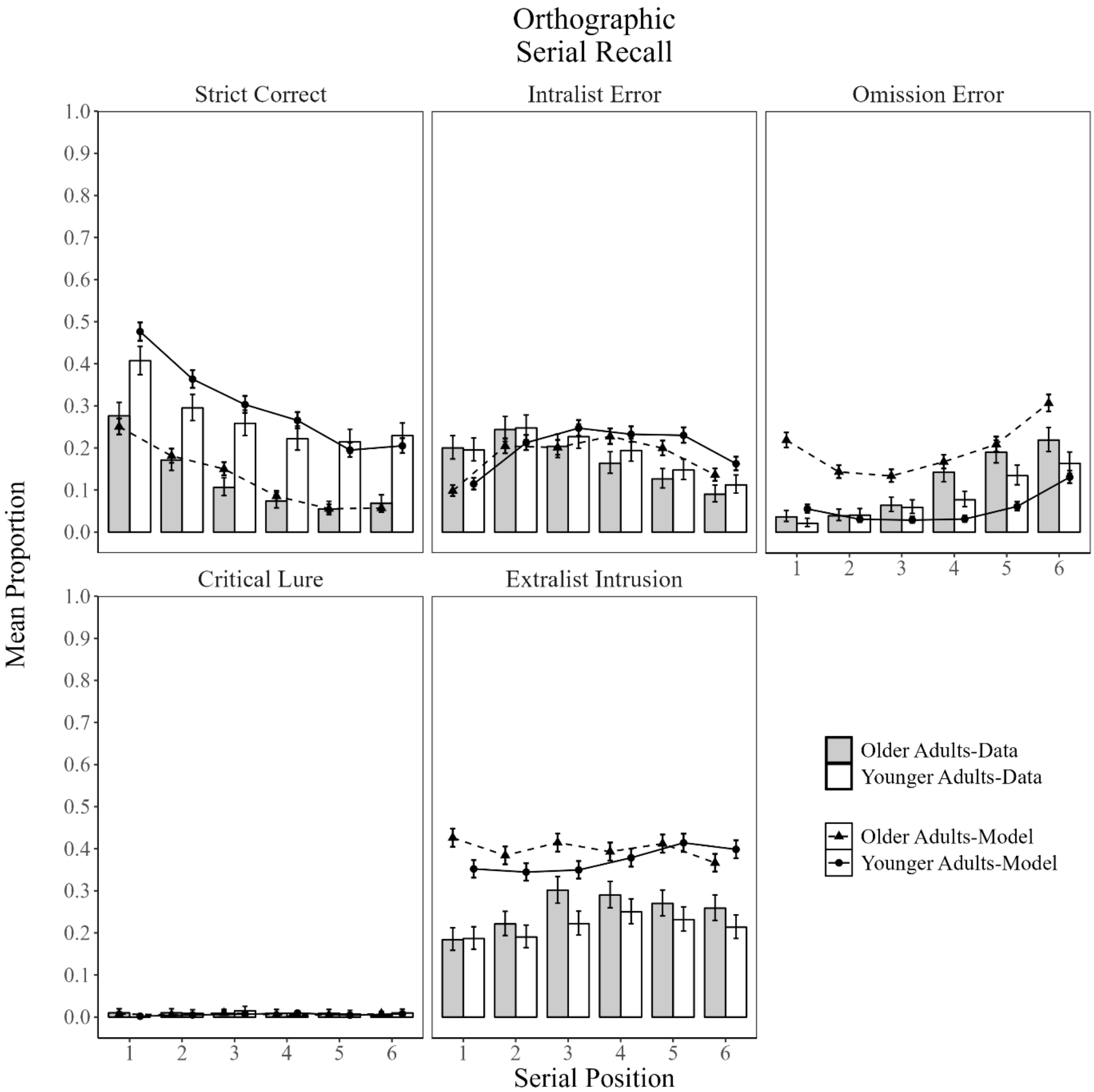
False Memories (Critical Lures). The frequency of critical lure recall remained low and consistent across age groups, with younger adults (M = .007, SD = .009) and older adults (M = .008, SD = .010). The analysis of variance showed no main effect of age group, $F<1$, $\eta_p^2 = .004$, $BF_{01} = 20.61$, and no main effect of serial position, $F(5, 395) = 1.39$, $\eta_p^2 = .017$, $BF_{01} > 10{,}000$, and no interaction between age group and serial position, $F(5, 395) = 2.29$, $\eta_p^2 = .028$, $BF_{01} = 478.54$.

Extralist Errors. Younger adults showed slightly fewer extralist intrusions (M = .22, SD = .13) than older adults (M = .25, SD = .13). The analysis of variance showed no main effect of age group, $F(1, 79) = 1.79$, $\eta_p^2 = .022$, $BF_{01} = 5.46$, a strong main effect of serial position, $F(5, 395) = 8.72$, $\eta_p^2 = .099$, $BF_{10} > 10{,}000$, and no interaction between age group and serial position, $F(5, 395) = 1.37$, $\eta_p^2 = .017$, $BF_{01} = 1511.10$.

Empirical Summary

In line with Experiments 5 and 6, younger adults outperformed older adults in correct serial recall. However, performance across error types, including intralist errors, omissions, false memories, and extralist intrusions, was largely comparable between the groups overall. Once again, omissions were more common toward the end of the list for older adults. In the following section, we examine whether the eCFM can capture these empirical outcomes.

Figure 7. *Empirical results (bars) and computational results (lines) showing the mean proportions of correct recall, intralist errors, omission errors, false memories (critical lures), and extralist errors for younger and older adults as a function of serial position for Experiment 7.*



Note. *Error bars represent 95% credible intervals.*

eCFM Simulation of Experiment 7

We now examine the capacity of the eCFM (Guitard et al., 2025a, 2025b, 2025c) to reproduce the core empirical patterns observed in Figure 7. As in previous simulations, we used representations suited to the task demands, specifically, orthographic representations in this case.

Simulation Lexicon. The orthographic lexicon was expanded to include the nonwords used in the experiment. Unlike previous simulations, no filtering criteria were applied as they were not relevant to the nonwords. The final lexicon contained 53,642 entries (words and nonwords), each encoded as a unique 300-dimensional vector. These representations were developed using an approach inspired by SERIOL and SERIOL2 models (Whitney, 2001; Whitney & Marton, 2013; Guitard et al., 2025b).

Simulation Parameters. We conducted 100 simulations, each involving 20 orthographically related lists, approximating a sample size of 100 participants. To capture the performance of young and older adults in Experiment 7, we set $L = 0.15$ and $d = 0.27$ for younger adults and $L = 0.11$ and $d = 0.23$ for older adults. Additionally, the parameter g was reduced from Experiments 5 and 6 to 0.015 in both age groups to account for the increased task difficulty and flatter serial position curves observed with orthographically similar nonwords. All other parameters were kept consistent with those used in Experiments 5 and 6: $s = 0.01$, $\tau = 3$, and $T = 0.35$.

Simulation Results. As shown in Figure 7, overall, while the model misfit some aspects of people's performance, it did successfully capture some of the key empirical patterns, with a slightly better fit for younger adults ($R^2 = .80$, MAD = .06) than for older adults ($R^2 = .73$, MAD = .06). In particular, the model accurately reproduced the serial position curve for strict scoring and intralist errors, with only minor discrepancies. This aspect of the model's fit is especially important given the close correspondence between the model-estimated and observed true recall rates across age groups, showing that the model can successfully account for the behavioral metric on which age differences were manifest. For omissions, the model tended to produce more errors than observed, especially at early serial positions, a pattern that was more pronounced among older adults. Regarding false memories, both the model and participants produced very few critical lure intrusions. The model also reflected the descriptively higher rate of extralist errors in older adults compared to younger adults, although it slightly overestimated the overall frequency of these errors for both younger and older adults. Despite these small deviations, the model captured the general trends in the data and aligned well with the observed age-related differences.

Discussion

Results of Experiment 7 with serial recall of orthographically related lists aligned with those of the previous serial recall experiments. Younger and older adults showed similarly low levels of false memory, but younger adults continued to outperform older adults in terms of correct recall. The eCFM captured some of the main patterns in the data, in particular patterns of correct serial recall. These differences were attributed to reduced global fidelity (L) in conjunction with less distinctive order representations (d) among older adults.

Summary of Serial Recall

The results of the serial recall experiments (Experiments 5-7) were clear and consistent. Across all experiments, young and older adults produced comparable rates of false recall, defined as either the recall of critical lures (unstudied words or nonwords associated with the lists) or extralist intrusions. At the same time, younger adults consistently exhibited a clear memory advantage in correct recall based on strict scoring. These patterns emerged reliably across experiments where items in a list were semantically (Experiment 5), phonologically (Experiment 6), or orthographically (Experiment 7) related. Importantly, these findings were well captured by eCFM, which reproduced the key empirical patterns by embedding task-relevant representations.

To account for the age differences in true recall, the eCFM required that items were represented with fewer lexical and order features among older adults ($L_{\text{old}} < L_{\text{young}}$) and that the features used to represent order were more similar across different serial positions ($d_{\text{old}} <$

d_{young}). The consistency in these representational impairments across the serial recall experiments reinforces both earlier computational models attributing age differences in serial recall to degraded representations (Neath & Surprenant, 2007; Surprenant et al., 2006) and more widespread claims that episodic memories become fuzzier with aging (Brainerd & Reyna, 2015; Greene & Naveh-Benjamin, 2023). When evaluated alongside the order reconstruction results for matched stimulus types (Experiments 1-3), the finding that age-related impairments in serial recall were manifest not only as a reduction in the discriminability of order representations (the main driver of age differences in order reconstruction, at least for semantically or phonologically related lists) but also in the overall fidelity of item representations highlights a task-dependent nature to age-related memory representational changes. That these representational changes are more extensive when the task of memory is to recall previously seen items in their exact order, with no explicit retrieval cues, aligns with both the environmental support hypothesis of aging (Craik, 1983) and the specificity principle of memory (Surprenant & Neath, 2009). Accordingly, the extent of age-related reductions in the fidelity of memory depends on how demanding it is to retrieve detailed information and what specific information must be retrieved.

Having now established that the eCFM can successfully capture true and false memories on tasks of order reconstruction and serial recall across multiple representational modalities (semantic, phonological, orthographic, and even visual) and age groups, we turn next to the first generalization of the eCFM to recognition memory.

Recognition

Because eCFM is built on the core principles of MINERVA 2 (Hintzman, 1984, 1986), we next sought to extend the model to recognition memory, a domain where MINERVA 2 has been successfully used in prior research (e.g., Arndt & Hirshman, 1998;

Chang et al., 2025; Hintzman, 1984, 1986; Jamieson et al., 2022; Reid & Jamieson, 2022; Reid et al., 2023, 2025). Although these prior successes might suggest that the eCFM should naturally perform well in recognition tasks, the present simulations serve a different purpose: to test whether the eCFM's structured lexical representations can support the same core retrieval principles across multiple representational domains and age groups. In doing so, we build directly on the framework and implementation established by Reid et al. (2025), extending their approach to a new context and demonstrating how representational structure can influence recognition performance across the lifespan.

Extending the eCFM to recognition has the added advantage of formally bridging recognition, order reconstruction, and serial recall in a computational modelling framework for the first time. These implications include the potential of the model to identify common processes underlying true and false memories across a diversity of tasks, which each place unique demands on memory, and representational modalities. Another key implication is in the eCFM's potential to advance a general processing account of age-related differences in memory that can reveal which components of memory are sensitive to representational change with age as a function of differing task demands. This aligns our approach closely with Benjamin's DRYAD framework (Benjamin, 2010), which, like the eCFM, builds on the core principles of MINERVA 2 to explain age differences in recognition. However, unlike DRYAD, the eCFM incorporates embedded representations rather than engineered ones (i.e., orthogonal representations chosen to maximize DRYAD's ability to parsimoniously explain age differences in memory with the minimal number of assumptions). The embedded lexical representations within the eCFM offer a different avenue for modeling how experience shapes memory performance while substantially reducing experimenter degrees of freedom. This is so because the representations are not selectively engineered but rather derived from a real lexicon, inducing substantial variability in how features relate across items based on actual experiences.

In old/new recognition tasks, participants typically study longer lists of items (e.g., 40 to 60 words). Unlike serial recall or order reconstruction, the memory test presents items one at a time, mixing previously studied (old) items with entirely new ones, and participants indicate whether each item was presented during the study phase. False memory can be examined by introducing new items that are highly similar to studied ones and assessing how often participants mistakenly endorse these lures as "old" (see Chang & Brainerd, 2021 for a review). Lure similarity induces substantial false recognitions among young and older adults alike (Greene et al., 2022; Roediger & McDermott, 1995), though these effects are typically heightened for older adults (for reviews, see Devitt & Schacter, 2016; Schacter et al., 1997). In the next series of experiments, we extend our investigation by systematically studying age differences in true and false recognition across DRM-style lists of semantically (Experiment 8), phonologically (Experiment 9), orthographically (Experiment 10), or visually (Experiment 11) similar items.

Experiment 8

In Experiment 8, we investigated recognition memory for words belonging to 10 different semantic categories, all associated with an unpresented critical lure, that were intermixed in a long study list. Participants studied a total of 60 words, with six each from 10 different semantic categories, presented in a random order.[6] Following a 6-second parity judgment task, participants completed a recognition test in which they were presented with 140 words: 60 studied words, the 10 critical lures, and 70 completely new and unrelated words. The goal was to examine age-related differences in true and false recognition between

[6] In effect, our procedure blends multiple DRM-type lists in one long study phase.

younger and older adults and to evaluate whether the eCFM could accurately simulate the observed patterns.

Method

Participants. The sample size, recruitment method, and inclusion criteria for Experiment 8 followed the procedures used in the previous experiments, while excluding from eligibility individuals who had participated in any earlier experiment. A new group of 84 participants was recruited via Prolific, comprising 42 younger adults and 42 older adults. Due to slight over-recruitment, four extra participants, two younger and two older adults were included. In the younger adult group, the mean age was 22.55 years (SD = 1.79), with 25 participants identifying as female, 16 as male, and 1 preferring not to disclose their gender. Among the older adults, the mean age was 68.17 years (SD = 3.77), with 28 participants identifying as female and 14 as male.

Materials and Design. The stimuli used in Experiment 8 were the same as those from Experiments 1 and 5 (see Appendix A). As in Experiment 1, the 20 related lists were divided into two sets of 10 lists (A and B), with half of the participants receiving A and half B as study items (60 words). At test, participants were presented with the same 60 studied words, 10 critical lures (each semantically related to six studied words), and 70 unrelated new words. Participants who studied the A lists received all the items from the B lists as unrelated new test probes, and vice versa.

Procedure. The procedure involved a similar format as the earlier experiments, with a study phase, a period of interpolated activity (the parity judgment task), and a test phase, but with the following modifications. Participants completed only a single trial in which they studied 60 words, presented sequentially in the center of their computer screen at a rate of one word per second (1000 ms on, 0 ms off). The 60 studied words were comprised of six words each from 10 semantic categories, all presented in a random order. As in the earlier experiments, the study phase was immediately followed by a 6-second parity judgment task.

Following the parity task, participants began the recognition test. Words were presented one at a time at the center of the screen, accompanied by the instructions: "Press the Z key for *Old*" (displayed in green on the bottom left) and "Press the M key for *New*" (displayed in red on the bottom right). Once a response was made, the next word appeared automatically. This continued until all 140 words (60 studied words, 10 critical lures, and 70 unrelated new words) were presented in a random order. As in previous experiments, participants were not allowed to go back and revise their responses, and they were informed about the upcoming memory test prior to the study phase.

Data Analysis. All recognition data, along with R Markdown files for analyses and modeling, are available on the OSF page linked to this project.

Scoring. Across Experiments 8 to 11, we analyzed the proportion of "old" responses as a function of test probe type (old, new, related). “Old” responses were classified as *hits* to old items and *false alarms* to new and related (critical lure) items. In line with previous studies (e.g., Roediger & McDermott, 1995), we predicted a higher rate of false alarms to critical lures than unrelated new words, reflecting increased false recognition for related items. We also predicted a higher rate of false alarms to the lures among older adults based on previous findings (e.g., Norman & Schacter, 1997), though it is possible that the mixing of multiple DRM-type lists into one long study phase could attenuate these age differences.

Statistical Analysis. Statistical analyses followed the same approach as in previous experiments. We conducted an ANOVA with probe condition (new, related, old) and age group (younger vs. older adults) as factors. Planned comparisons were further examined using both frequentist and Bayesian t-tests.

Results

Figure 8 displays the empirical and simulation results for the mean proportions of "old" responses for old words (studied items), new words (unstudied and unrelated items), and related lures (unstudied items, each semantically related to six studied words) for younger and older adults.

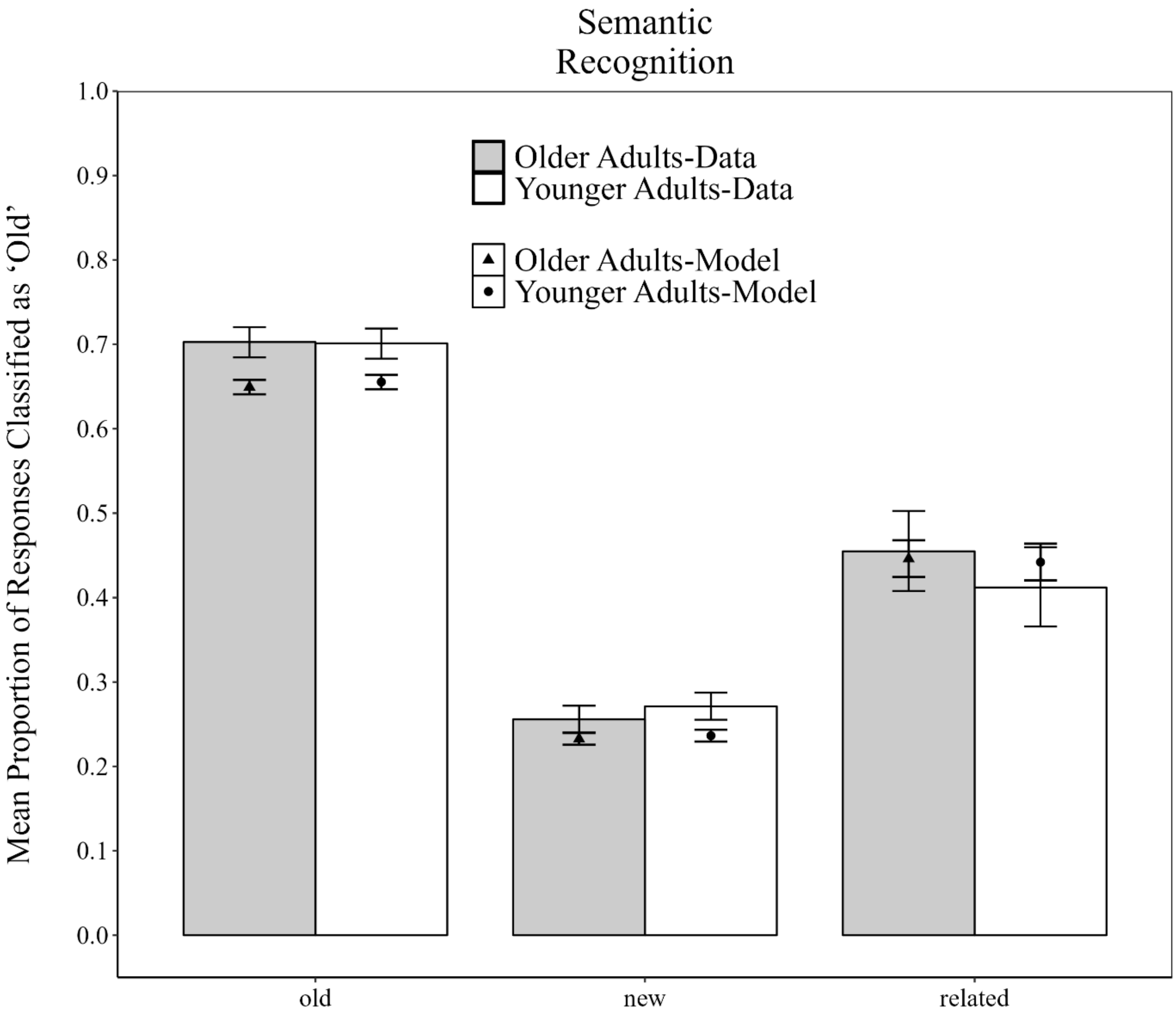
At the descriptive level, the hit rate was similar between younger adults (*M* = .70, *SD* = .14) and older adults (*M* = .70, *SD* = .17). Likewise, false alarm rates for new words were comparable between younger (*M* = .27, *SD* = .13) and older adults (*M* = .26, *SD* = .15). False recognition of critical lures was descriptively slightly lower for younger adults (*M* = .41, *SD* = .17) compared to older adults (*M* = .45, *SD* = .22).

The results of the analysis of variance (ANOVA) with age group (younger vs. older adults) and probe condition (new, related, old) revealed no main effect of age group, $F < 1$, $\eta_p^2 = .002$, $BF_{01} = 13.99$, nor an interaction between age group and probe condition, $F(2, 164) = 1.02$, $\eta_p^2 = .01$, $BF_{01} = 57.31$. However, there was a clear main effect of probe condition, $F(2, 164) = 223.73$, $\eta_p^2 = .73$, $BF_{10} > 10{,}000$. Post hoc Bayesian t-tests confirmed that the proportion of “old” responses differed credibly across all probe conditions (new vs. related, new vs. old, and old vs. related), all BFs > 10,000.

Empirical Summary

Overall, the results were clear. Younger and older adults produced comparable rates of false memory, specifically in the tendency to endorse related lures as “old” during the recognition test. Performance across all probe conditions (old, new, related) was highly similar between the two age groups, with no credible age-related differences. In the following section, we evaluate whether the eCFM can successfully account for these key empirical patterns.

Figure 8. *Empirical results (bars) and computational results (points) showing the mean proportions of "old" responses for old words (studied words), new words (unstudied and unrelated words), and related lures (unstudied words, each semantically related to six studied words) for younger and older adults in Experiment 8.*



Note. *Error bars represent 95% credible intervals.*

eCFM Simulation of Experiment 8

We now evaluate the ability of the eCFM (Guitard et al., 2025a, 2025b, 2025c) to reproduce the core empirical patterns observed in Figure 15. Here, we extend the model to capture recognition memory. Because the eCFM is grounded in the principles of MINERVA 2 (Hintzman, 1984, 1986), this extension to recognition is seamless and conceptually straightforward.

As described in Equation 4, to model recognition, the underlying architecture of the eCFM remains nearly identical to that used in previous simulations of order reconstruction and serial recall, with only a few adjustments. Recognition memory tasks typically exhibit minimal serial position effects, as output curves tend to be relatively flat particularly for long lists of items like ours (see e.g., Gionet et al., 2024). Consequently, we applied a uniform encoding strength across all studied items, and parameters related to differential encoding, specifically, *g* and *d*, were not needed in these simulations.[7] (They could, however, be introduced in future extensions if special serial position effects in recognition tasks emerge, but we omit them here to avoid spuriously attributing true and false recognitions to a process that, while realized in the model, might not operate in recognition memory at all.)

In our earlier order reconstruction and serial recall simulations, the model selected the word with the highest cosine similarity to the echo either from a re-presented subset (reconstruction) or from the entire lexicon (serial recall). In recognition, consistent with the original MINERVA 2 framework (Hintzman, 1984, 1986), no selection among candidates is necessary. Instead, the probe item contacts all memory traces in parallel, each trace is activated based on its cosine similarity to the probe, and these activations are then summed across all traces to yield a single *familiarity* value for each probe (see Equation 4). This familiarity score is used to make a recognition decision. Recognition decisions were then determined by comparing each familiarity against a threshold *T*. If the familiarity exceeded *T*, the model responded "old"; otherwise, it responded "new." Thus, the threshold *T* serves as a decision criterion and to reflect an ideal observer based on the actual proportion of old items

[7] Although this removes age differences in positional precision from being considered in recognition as they were in order reconstruction and serial recall, the minimal role that the exact order of items plays on tests of recognition justifies this omission. Otherwise, the model could risk attributing the source of age differences in recognition memory to a process (changes in the discriminability of order encoding) that might actually play no role at all in these differences.

at test (60 old items out of 140 total), we set T to the 42.86th percentile of the familiarity distribution, following the approach used by Reid et al. (2023, 2025).

Simulation Representations. For recognition simulations in Experiment 8, we used the same semantic representations as in Experiments 1 and 5. Each studied and probe word was represented as a unique 300-dimensional vector based on LSA (Landauer & Dumais, 1997; Guitard et al., 2025b).

Simulation Parameters. For each experiment, we conducted 100 simulation runs, each involving 60 studied words followed by a 140-item recognition test (60 old, 10 related, 70 unrelated probes), corresponding to an effective sample size of 100 participants per age group. Because there were minimal age-related differences observed in Experiment 8, the same encoding fidelity parameter setting was used for both younger and older adults (L = 0.17), and τ (controlling the influence of similarity at retrieval) was fixed at 3 in both age groups as in our earlier simulations. No additional parameters (e.g., g, d, s) were needed in this set of simulations, aside from setting the threshold T as described above.

Simulation Results. As shown in Figure 8, the model successfully captures the key empirical patterns, demonstrating strong fits for both younger adults (R^2= .97, MAD = .04) and older adults (R^2=.99, MAD = .03). Although there were minor discrepancies, such as a slight underestimation of hit rates for old words, the model accurately reproduced the overall structure of responses across all probe types.

Discussion

When younger and older adults were tasked with recognizing which words were previously studied in a recognition test with old (studied) words, unrelated new words, and semantically related critical lures, they exhibited comparable levels of performance across probe types (old, new, and related). Critically, by embedding semantic representations within the eCFM, we were able to accurately capture all major patterns of memory performance across age groups. These findings evidence that the eCFM, originally developed for order reconstruction (Experiments 1–4) and serial recall (Experiments 5–7), can be seamlessly extended to model recognition memory as well. This extension underscores the generalizability of our framework across multiple memory paradigms.

In our results, we observed a small age-related difference at the descriptive level in false recognition. The finding of no reliable age-related difference in false recognition of the critical lures in Experiment 8 is at odds with many previous studies reporting higher rates of false recognition of semantically related items among older adults (Dennis et al., 2007; Norman & Schacter, 1997; Rankin & Kausler, 1979; Smith, 1975). However, our results are not statistically convincing and may instead reflect the influence of mixing multiple DRM-type lists within a single extended study phase, with items from each list presented randomly throughout. Typically, when shown multiple items consecutively from the same semantic category, older adults engage in more gist-based processing than younger adults do, focusing on and retaining the shared meaning of the items, thereby increasing their sensitivity to false recognition of similar items consistent with this shared meaning (Devitt & Schacter, 2016; Tun et al., 1998). Intermixing items from different categories in our procedure might have attenuated age differences in gist-based processing. Because there was no age difference in true or false recognition in our data, the eCFM did not require any differences in representational fidelity (L) between young and older adults. Yet, we will soon revisit how the eCFM could account for the typical finding of higher false recognition in older adults in procedures like ours (see section “Summary of Recognition”).

Experiment 9

Experiment 9 extends our empirical and modelling investigation of age differences in true and false recognition to phonologically related materials.

Method

Participants. The sample size, recruitment procedures, and inclusion criteria were identical to those used in previous experiments. As before, individuals who had participated in earlier studies were not eligible for this experiment. A new group of 80 participants was recruited via Prolific, comprising 40 younger adults and 40 older adults. In the younger group, the mean age was 23.10 years (SD = 1.67), with 24 participants identifying as female, 15 as male, and one participant preferring not to disclose their gender. In the older group, the mean age was 69.5 years (SD = 4.32), with 27 participants identifying as female and 13 as male.

Materials, Design, Procedure, and Data Analysis. The experimental procedure and analysis approach mirrored those of Experiment 8, with the exception of the materials. Here, the stimuli were drawn from the phonologically related lists used in Experiments 2 and 6 (see Appendix B), with half of the participants receiving set A (the first 10 related lists) and half receiving set B (the second 10 related lists) as study lists. As in Experiment 8, participants completed a single study phase (60 words) followed by a recognition test with 140 items (60 old studied words, 70 unrelated new words, and 10 phonologically related new words).

Results

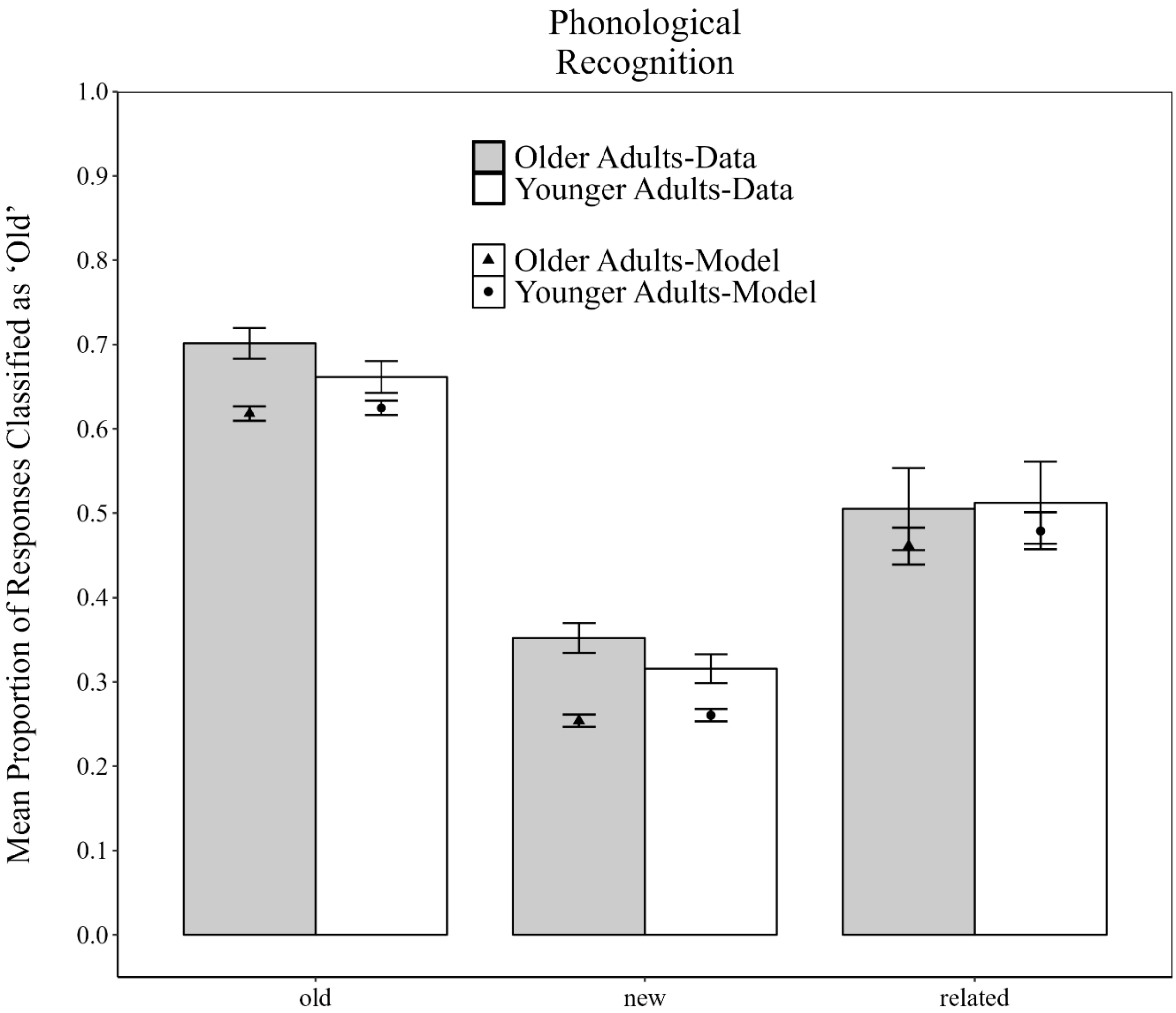
Figure 9 presents the mean proportions of "old" responses for old words (studied items), new words (unstudied and unrelated items), and phonologically related lures for both younger and older adults and the results from our simulations. At the descriptive level, the proportion of old responses to studied words was slightly lower for younger adults (M = .66, SD = .15) compared to older adults (M = .70, SD = .16). False alarm rates to new words were similar across groups, with younger adults at (M = .32, SD = .20) and older adults at (M = .35, *SD* = .21). False recognition rates for phonologically related lures were high and nearly identical between younger (*M* = .51, *SD* = .19) and older adults (*M* = .50, *SD* = .24).

The ANOVA results closely paralleled those of Experiment 8. There was no main effect of age group, $F < 1$, $\eta_p^2 = .005$, $BF_{01} = 9.27$, nor an interaction between age group and probe condition, $F < 1$, $\eta_p^2 = .01$, $BF_{01} = 118.59$, but there was a strong main effect of probe condition, $F(2, 156) = 144.30$, $\eta_p^2 = .65$, $BF_{10} > 10{,}000$. Post-hoc Bayesian t-tests confirmed that the proportions of "old" responses differed across all probe types (new vs. related, new vs. old, and old vs. related), all BF > 10,000.

Empirical Summary

Results of Experiment 9 with phonologically similar items align with the recognition findings for semantically similar items in Experiment 8. Both younger and older adults showed increased false recognitions for the critical lures than the unrelated items. Performance across all probe types (old, new, and related) was highly similar between the two age groups, with no credible age-related differences. In the next section, we examine whether the eCFM can successfully model these empirical patterns.

Figure 9. *Empirical results (bars) and computational results (points) showing the mean proportions of "old" responses for old words (studied words), new words (unstudied and unrelated words), and related lures (unstudied words, each phonologically related to six studied words) for younger and older adults in Experiment 9.*



Note. *Error bars represent 95% credible intervals.*

eCFM Simulation of Experiment 9

Here, we evaluate the ability of the eCFM (Guitard et al., 2025a, 2025b, 2025c) to reproduce the core empirical patterns observed in Figure 9.

Simulation Representations. For the recognition simulations in Experiment 9, we used the same phonological representations as in Experiments 2 and 6. Each studied and probe word was represented as a unique 300-dimensional vector based on Parrish's method (Parrish, 2017; Guitard et al., 2025b).

Simulation Parameters. As in Experiment 8, we conducted 100 simulation runs for Experiment 9, each involving 60 studied words followed by a 140-item recognition test (60 old words, 10 phonologically related new lures, and 70 unrelated new words), corresponding to an effective sample size of 100 participants per age group. Given that minimal age-related differences were observed in Experiment 9, the same encoding parameter ($L = 0.04$) was applied for both younger and older adults, and τ was fixed at 3. No additional parameters (e.g., g, d, s) were required in these simulations, aside from setting the decision threshold T, relative to the the familiarity distribution as described in Experiment 8.

Simulation Results. As shown in Figure 9, the model successfully captures the main empirical patterns, demonstrating good fits for both younger adults ($R^2 = .99$, MAD = .04) and older adults ($R^2 =.98$, MAD = .08). Although there were minor discrepancies, such as a slight underestimation of hit rates for old words, particularly for older adults, and a slightly lower false alarm rate for the unrelated new probes, these deviations were relatively small. Despite these minor omissions, the model captured the overall structure and key aspects of recognition performance across both age groups.

Discussion

Experiment 9 replicated and extended the true and false recognition results for semantically similar items in Experiment 8 to phonologically similar items. Younger and older adults committed comparably high rates of false alarms to phonologically related lures. Indeed, recognition performance was comparable between age groups across all probe types (old, new, and related). Importantly, by embedding phonological representations into the eCFM, the model was able to replicate the key patterns of memory performance observed across both age groups. Here again, the model identified no representational difference in item fidelity among young and older adults in a test of recognition memory. As noted earlier, we will soon revisit (see section "Summary of Recognition") how the eCFM could account for an age difference in false recognition in procedures like ours, given that such differences have been reported previously for phonologically similar items (Budson et al., 2003; Watson et al., 2001).

Experiment 10

In Experiment 10, we extend our systematic investigation of similarity effects on false recognition to the orthographic domain.

Method

Participants. The sample size, recruitment methods, and inclusion criteria for Experiment 10 followed the same procedures as in previous experiments. As before, individuals who had participated in any earlier studies were excluded from participation. A new group of 80 participants was recruited through Prolific, evenly divided between 40 younger and 40 older adults. Among the younger adults, the mean age was 22.75 years (SD = 1.93), with 19 participants identifying as female, 19 as male, and two preferring not to disclose their gender. In the older adult group, the mean age was 68.62 years (SD = 3.14), with 22 participants identifying as female and 18 as male.

Materials, Design, Procedure, and Data Analysis. The experimental procedure and analysis strategy were consistent with those used in Experiments 8 and 9 but with different materials. In this experiment, stimuli were drawn from the orthographically related nonword lists used in Experiments 3 and 7, as detailed in Appendix C. Half of the participants studied the first 10 related lists (set A), and half studied the second 10 related lists (set B), with the unstudied lists in each condition serving as the unrelated items in the recognition test. Participants completed a single study phase, in which they were presented with 60 nonwords, followed by a recognition test comprising 140 items: 60 studied nonwords (old), 70 unrelated new nonwords, and 10 orthographically related new nonwords.

Results

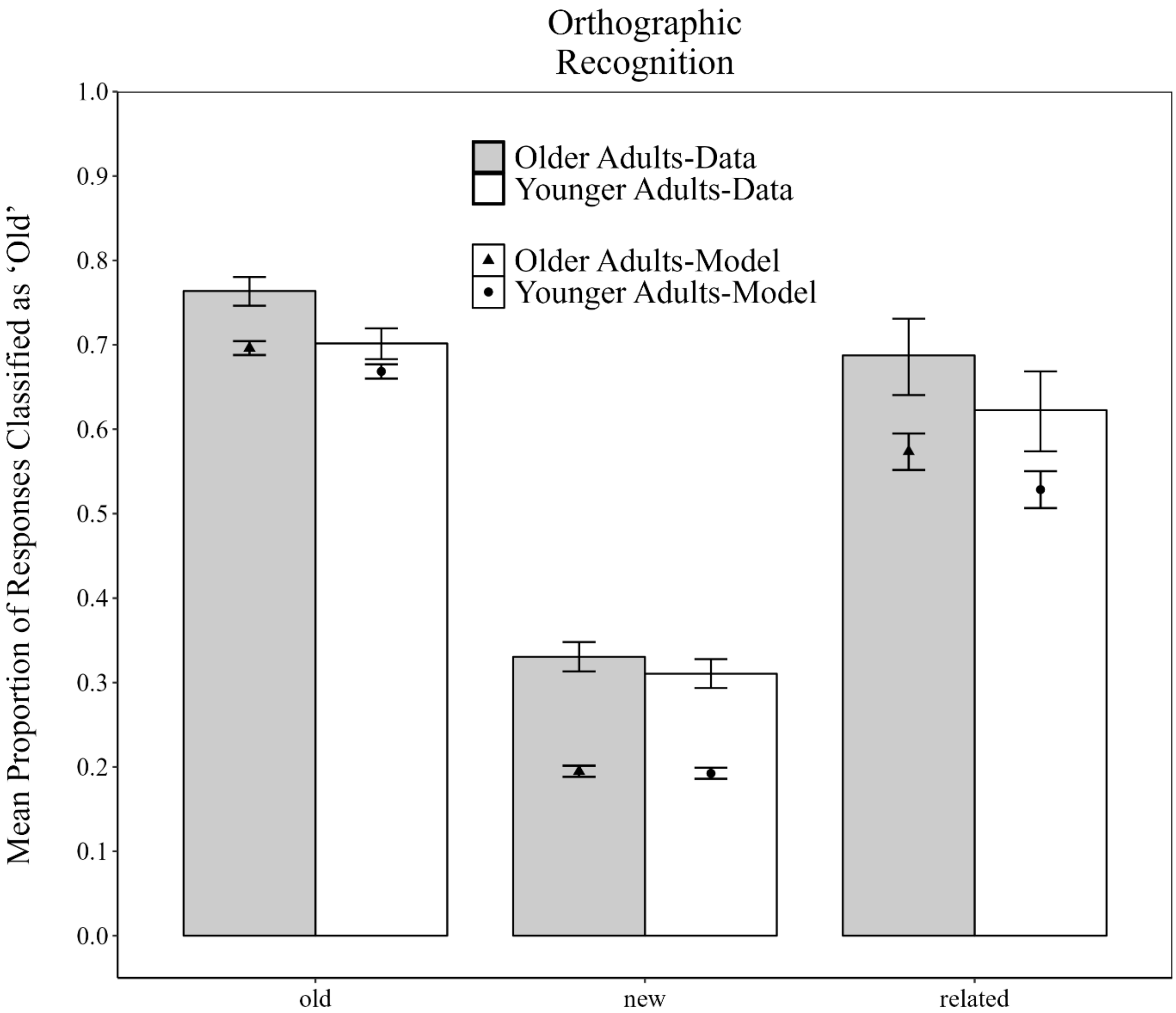
Figure 10 shows the mean proportions of "old" responses for old nonwords (studied items), new unrelated nonwords, and orthographically related lures, separately for younger and older adults and the results from our simulations. At the descriptive level, younger adults showed a slightly lower proportion of old responses to studied nonwords (M = .70, SD = .17) compared to older adults (M = .76, SD = .16). False alarm rates to unrelated new nonwords were similar between younger adults (M = .31, SD = .16) and older adults (M = .33, SD = .15). False recognition rates for orthographically related lures were descriptively somewhat lower for younger adults (M = .62, SD = .21) than for older adults (M = .69, SD = .18) and were extremely high in both groups.

Despite these small descriptive trends, the ANOVA revealed no credible main effect of age group, $F(1, 78) = 3.37$, $\eta_p^2 = .04$, $BF_{01} = 3.48$, nor an interaction between age group and probe condition, $F < 1$, $\eta_p^2 = .01$, $BF_{01} = 13.22$. A robust main effect of probe condition was observed, $F(2, 156) = 173.27$, $\eta_p^2 = .69$, $BF_{10} > 10{,}000$. As in previous experiments, post-hoc Bayesian t-tests further confirmed that the proportion of “old” responses differed across all probe types (new vs. related, new vs. old, and old vs. related), all BF > 598.29.

Empirical Summary

Overall, both younger and older adults exhibited elevated false recognition rates when the unstudied items were orthographically related to studied nonwords. Although small age-related differences were evident descriptively, the statistical analyses provided stronger support for the absence of any meaningful age effects. Consistent with our previous findings, performance across all probe types (old, new, and related) was comparable between age groups. In the next section, we evaluate whether the eCFM can successfully account for these empirical patterns.

Figure 10. *Empirical results (bars) and computational results (points) showing the mean proportions of "old" responses for old nonwords (studied nonwords), new nonwords (unstudied and unrelated nonwords), and related lures (unstudied nonwords, each orthographically related to six studied nonwords) for younger and older adults in Experiment 10.*



Note. *Error bars represent 95% credible intervals.*

eCFM Simulation of Experiment 10

Here, we evaluate the ability of the eCFM (Guitard et al., 2025a, 2025b, 2025c) to reproduce the core empirical patterns observed in Figure 10.

Simulation Representations. For the recognition simulations in Experiment 10, we employed the same orthographic representations used in Experiments 3 and 7. Each studied and probe nonword was represented as a unique 300-dimensional vector, based on orthographic coding principles from the SERIOL and SERIOL2 frameworks (Whitney, 2001; Whitney & Marton, 2013).

Simulation Parameters. As in Experiments 8 and 9, we conducted 100 simulation runs for Experiment 10. Each run involved 60 studied nonwords followed by a recognition test comprising 140 probes (60 old nonwords, 10 orthographically related new lures, and 70 unrelated new nonwords), corresponding to an effective sample size of 100 participants per age group. Given the minimal age-related differences observed empirically in Experiment 10, we applied the same encoding parameter (L = 0.09) for both younger and older adults and kept τ fixed at 3. No additional parameters (e.g., g, d, s) were needed for these simulations, aside from setting the decision threshold T based on the quantile of the familiarity distribution, as previously described in Experiment 8.

Simulation Results. As shown in Figure 10, the model closely captured the main empirical patterns, yielding strong fits for both younger adults (R^2= .99, MAD = .08) and older adults (R^2 =.99, MAD = .11). Although the model slightly underpredicted the number of false alarms to unrelated new nonwords and, to a lesser extent, to related lures, it successfully reproduced the overall structure of recognition performance across both age groups. The model even produced a slight apparent advantage for older adults; however, this difference likely reflects random variation within the simulations and mirrors the minor, non-systematic variations observed in the empirical data.

Discussion

In Experiment 10, younger and older adults completed a recognition memory task involving nonwords following a brief parity judgment task. The test phase included studied items (old), unrelated new nonwords, and orthographically related lures. Consistent with the patterns observed for semantic (Experiment 8) and phonological (Experiment 9) materials, participants were more likely to falsely recognize orthographically related lures as "old" compared to unrelated new nonwords. No significant age-differences in true or false recognition were observed. By incorporating orthographic representations into the eCFM, we were able to accurately reproduce the key patterns of recognition performance across both younger and older adults. As in the previous recognition experiments, there were no age differences in the fidelity of item representations (L) in the eCFM.

Experiment 11

Experiment 11 builds on our previous recognition experiments by testing true and false recognition, and the eCFM’s ability to capture these findings, for visually abstract images, like those used in Experiment 4 on order reconstruction.

Method

Participants. The sample size, recruitment procedure, and inclusion criteria were consistent with those used in prior experiments, including the exclusion of eligibility of any individual who participated in any of those experiments. A new sample of 81 participants was recruited via Prolific, comprising 41 younger adults and 40 older adults. One additional younger adult was included due to unintentional over-recruitment. Among the younger adults, the mean age was 23.73 years (SD = 1.47), with 27 identifying as female, 11 as male, and three opting not to disclose their gender. In the older adult group, the mean age was 68.90 years (SD = 3.13), with 27 identifying as female and 13 as male.

Materials, Design, Procedure, and Data Analysis. The experimental procedure and analysis approach mirrored those of Experiments 8–10, with the exception of the stimuli. In this experiment, visual stimuli were drawn from the image sets used in Experiment 4, detailed in Appendix C. During the study phase, participants viewed 60 images, with half of the participants studying the first 10 lists of related items (set A) and half the second 10 lists (set B), while items from the unstudied lists for a given participant were presented as unrelated test probes. The subsequent recognition test comprised 140 images: 60 studied (old) images, 70 unrelated new images, and 10 new images that were visually similar to studied items.

Results

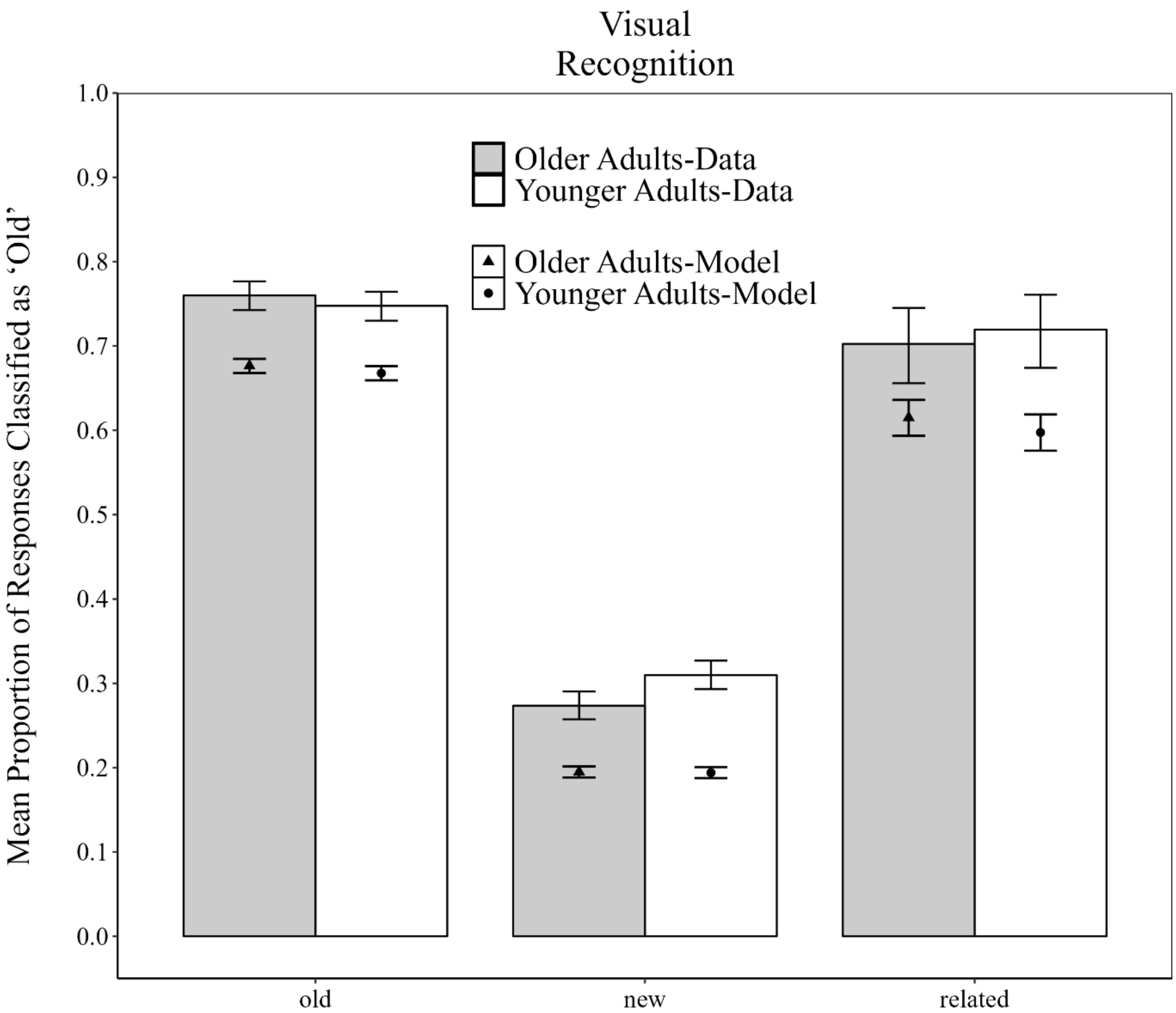
Figure 11 displays the mean proportions of "old" responses for old images (studied items), new unrelated images, and visually related lures, separately for younger and older adults and the results from our simulations. Descriptively, performance was similar across age groups, with comparable hit rates between young adults (M = 0.75, SD = 0.14) and older adults (M = 0.76, SD = 0.15). False alarm rates to unrelated new images were also comparable (younger adults: M = 0.31, SD = 0.17; older adults: M = 0.27, SD = 0.15). Strikingly, false recognition rates for visually related lures were extremely high in both groups (younger adults: M = 0.72, SD = 0.17; older adults: M = 0.70, SD= 0.18), almost indistinguishable from responses to studied items.

ANOVA results revealed no credible main effect of age group, $F < 1$, $\eta_p^2 = .03$, BF_{01} = 11.94, nor an interaction between age group and probe condition, $F < 1$, $\eta_p^2 = .01$, BF_{01} = 4.81. A clear main effect of probe condition was observed, $F(2, 158) = 393.02$, $\eta_p^2 = .83$, BF_{10} > 10,000. Post hoc Bayesian t-tests further confirmed that all probe types differed from one another (new vs. related, new vs. old, and old vs. related), with BF > 5.25 in all comparisons.

Empirical Summary

Participants in both age groups showed very high false recognition rates for visually related lures, nearly equivalent to their correct recognition rates for studied items, highlighting the strong influence of visual similarity on recognition memory. Consistent with our earlier experiments, there were no reliable age-related differences across any probe type. In the next section, we assess whether the eCFM can successfully account for these empirical outcomes.

Figure 11. *Empirical results (bars) and computational results (points) showing the mean proportions of "old" responses for old images (studied images), new images (unstudied and unrelated images), and related lures (unstudied images, each visually related to six studied images) for younger and older adults in Experiment 11.*



Note. *Error bars represent 95% credible intervals.*

eCFM Simulation of Experiment 11

In this section, we evaluate whether the eCFM (Guitard et al., 2025a, 2025b) can account for the core empirical patterns observed in Figure 11.

Simulation Representations. For the recognition simulations in Experiment 11, we employed the same visual representations used in Experiment 4. Each studied and probe image was encoded as a unique 300-dimensional vector, capturing visual similarity within a high-dimensional feature space.

Simulation Parameters. Consistent with the approach used in Experiments 8–10, we conducted 100 simulation runs for Experiment 11. Each run involved a study phase with 60 images followed by a recognition test consisting of 140 probes: 60 previously studied images (old), 10 visually related lures, and 70 unrelated new images. This setup was designed to mirror the empirical procedure and simulate a sample size of 100 participants per age group. Given the lack of age-related differences observed in the behavioral data, we applied the same encoding rate parameter for both age groups (L = 0.04), while retaining the setting $\tau = 3$ and setting the familiarity threshold T to the 42.86th percentile of the familiarity distribution, as described in Experiment 8. As with our previous recognition simulations, no additional parameters (e.g., g, d, s) were required.

Simulation Results. Overall, as shown in Figure 11, the model closely replicated the observed data, yielding excellent fits for both younger adults (R^2 = .99, MAD = .11) and older adults ($R^2 \approx 1.00$, MAD = .08). Although the model slightly underpredicted the proportion of "old" responses across all probe types, it successfully captured the overall structure and relative differences among the three conditions. These results further demonstrate the eCFM's flexibility in generalizing across different stimulus types, including visual information, and across both age groups.

Discussion

Consistent with the earlier recognition experiments (Experiments 8 -10), results of Experiment 11 with visually abstract images revealed higher rates of false recognition of visually related lures than unrelated items. These heightened false recognitions were observed among young and older adults at a comparable rate and were nearly as high as true recognition of studied images. No age differences in true or false recognition of visually abstract images were detected. By incorporating visual representations into the eCFM, we successfully replicated the key patterns of recognition performance across both age groups. We turn now to a summary of the recognition results, where we interrogate how the eCFM could account for the more standard finding in the literature, wherein older adults commit *more* false recognitions than younger adults (for a review, see Devitt & Schacter, 2016).

Summary of Recognition

As in our earlier analysis of order reconstruction (Experiments 1-4) and serial recall (Experiments 5-7), in our systematic investigation of true and false recognition across lists of semantically (Experiment 8), phonologically (Experiment 9), orthographically (Experiment 10), and visually (Experiment 11) related items, we showed that a model that incorporates structured lexical representations – the eCFM (Guitard et al., 2025a, 2025b, 2025c) – could successfully capture the performance of young and older adults across these various stimulus types. Critically, the model retained most of the same architecture as it used to simulate order reconstruction and serial recall, allowing it to easily generalize to recognition the same basic processing assumptions it uses to explain memory on these other tasks. The main difference was that, in modeling recognition, the eCFM did not assign a unique positional representation to each item in the list, as it does in order reconstruction and serial recall and, instead, computed trace activation based on the similarity of the test word presented and the representation of the words stored into memory. This was motivated by the relatively constant rate of recognition accuracy across serial positions in long study lists like the ones employed in Experiments 8-11 (e.g., Gionet et al., 2024). Nonetheless, the eCFM could be easily extended to model serial position effects in recognition in future applications, as we retained the positional features in the model but rendered them, in the present simulations, functionally inert.

Across the recognition experiments, we observed comparable rates of true and false recognition among young and older adults. False recognitions were often extremely high in both age groups, especially for visually related lures in Experiment 11, where these items were judged as "old" equally as often as the studied items were. The lack of age differences in true or false recognition was a bit perplexing, given that many previous studies have reported such differences in procedures like ours (e.g., Abadie et al., 2021; Budson et al., 2003; Dennis et al., 2007; Koutstaal & Schacter, 1997; Norman & Schacter, 1997; Tun et al., 1998). However, unlike those earlier procedures, we intermixed items from multiple categories in a long study phase, potentially attenuating age differences in gist-based processing presumed to underlie older adults' heightened tendency to false recognition of critical lures in DRM procedures where items from a specific category are all presented consecutively (cf., Devitt & Schacter, 2016). To be clear, young and older adults likely adopted a similar degree of gist-based processing in our recognition procedures, allowing both age groups to recognize the types of items they studied (e.g., animals, or a general visual pattern) that resulted in their over-generalizing to recognizing unstudied items that were consistent with those categories. This is evident both in the very high false recognition rates in both groups (especially in Experiment 11), which were always higher for critical lures than unrelated items, and by the fact that the eCFM successfully captured the performance of young and older adults alike by assuming a common representational fidelity, or learning rate (L), applied to both age groups for a given stimulus type.

Still, the question of whether the eCFM could account for an age difference in false recognition, particularly one that selectively applies to critical lures but not unrelated items, remains unaddressed from our simulations. Given that such patterns of age differences in false recognition have been reported in previous empirical studies whose procedures share much in common with ours, we next asked how the eCFM might produce such a pattern and whether it would arise from a change in representational fidelity L or a different process.

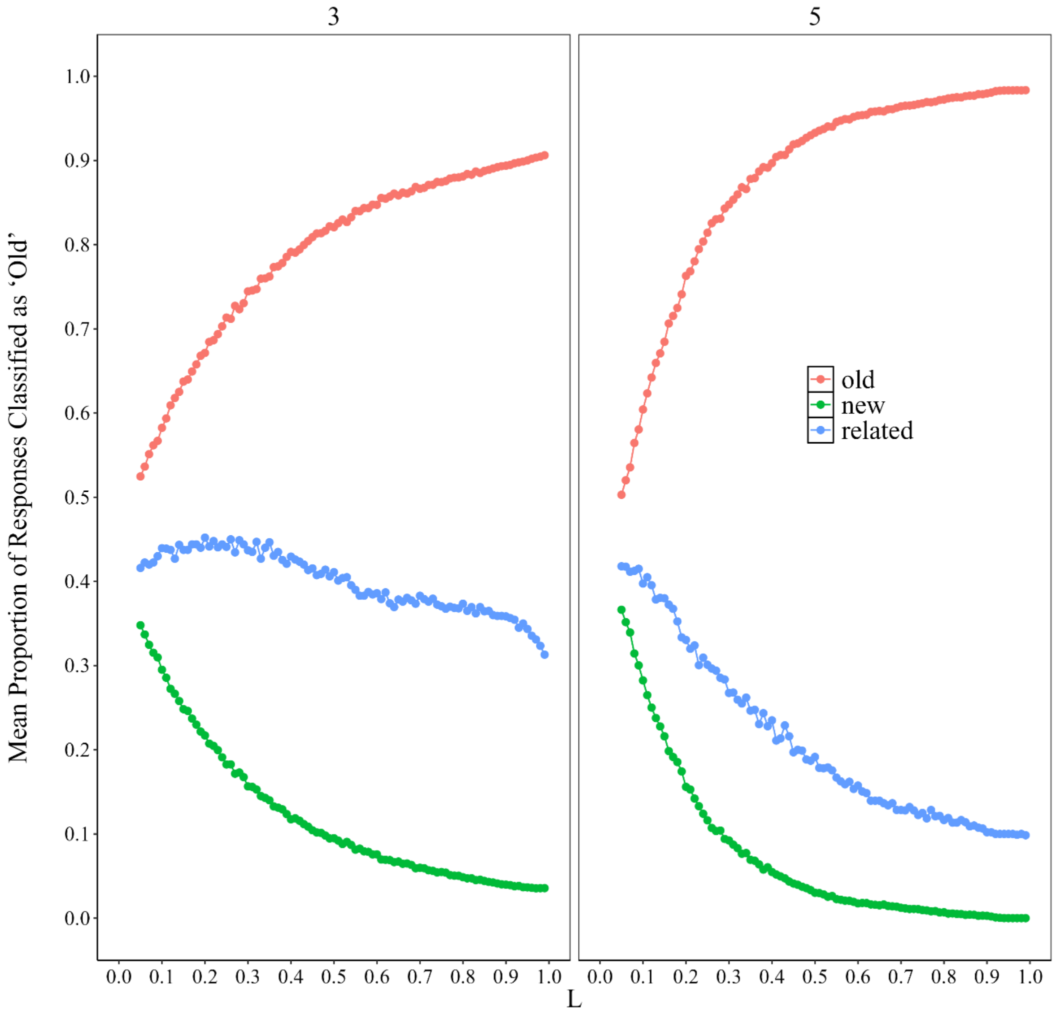
Age-Related Differences in Recognition. Although our previous simulations established that the eCFM reproduces overall recognition patterns, the model had not yet been tested on whether it can account for age-related differences in false recognition. To address this question, in this section we briefly illustrate how two parameters may account for such age-related differences: the learning probability (L) and the retrieval sensitivity (τ).

The learning probability (L), as described in the Introduction, determines the likelihood that a feature is encoded during study; lower values yield sparser and less faithful representations. The retrieval sensitivity (τ) governs the model's responsiveness to similarity: higher τ amplifies strong similarity signals, whereas lower τ makes retrieval more tolerant to weaker matches.

To examine their influence and potential to capture age-related differences, we systematically varied L between 0 and 1 for two levels of τ (3 and 5). The results, shown in Figure 12, illustrate the model's predicted proportion of "old" responses to studied, related, and unrelated items using semantic vectors. These simulations were conducted with the materials and model parameters described in Experiment 8, but the pattern is consistent across materials and representational types. As L increased, recognition accuracy improved for studied items, while false alarms to unrelated items declined, reflecting more distinctive encoding. Responses to related lures, however, showed a non-monotonic trend: at intermediate L values, false recognition peaked, consistent with partial overlap among encoded features. Increasing τ accentuated this differentiation, amplifying the advantage for old items and the relative vulnerability of related lures.

These simulations suggest that age differences in false recognition can naturally emerge from reductions in representational fidelity (L), from altered retrieval sensitivity (τ), or from a combination of both. Together, these parameters provide a simple, yet informative account of why older adults may show elevated false recognition for related items yet maintain discrimination between studied and unrelated words.

Figure 12. Simulation of the mean proportion of responses classified as "old" across varying values of L for $\tau = 3$ (left panel) and $\tau = 5$ (right panel), shown separately for old, new, and semantically related lures.



General Discussion

Here, we embraced Newell's (1973) call for general computational frameworks to study cognition by demonstrating how a model, the eCFM, combining instance theories of episodic memory (e.g., Hintzman, 1986, 1988) with a lexicon derived from distributional semantic (Landauer & Dumais, 1997), phonological (Parrish, 2017), orthographic (Whitney, 2001; Whitney & Marton, 2013), or visual (Bingham & Mannila, 2001) techniques can apply the same set of principles across tasks, materials, and age groups. Specifically, we built on the recent success of the eCFM in explaining order reconstruction and serial recall in younger adults (Guitard et al., 2025a, 2025b, 2025c), as well as the success of MINERVA-OPS in modeling recognition in young adults (Reid et al., 2025), by extending these approaches to older adults and to recognition memory across lists of semantically, phonologically, orthographically, and visually related materials. Our systematic investigation of true and false memory across 11 new empirical studies of order reconstruction (Experiments 1-4), serial recall (Experiments 5-7), and item recognition (Experiments 8-11), through the lens of a common computational modeling framework, afforded rich new insights into shared processes across these tasks that have often been studied separately. In the ensuing, we summarize these new insights and evaluate what this new modeling tells us about the source of age differences in adult memory performance, with broader implications for theories of memory and cognition more generally.

On Building a General Framework for Memory Theory

Perhaps more than any domain of cognition, research on human memory has often proceeded in highly specialized fashion. Although there are certainly exceptions (e.g., Brown et al., 2007; Hurlstone et al., 2014; Hedayati et al., 2022; Kahana, 2020; Jamieson et al., 2022; Oberauer & Lin, 2024; Spens & Burgess, 2024), much of what we have come to know about human memory has been derived from research focused on distinct tasks (e.g., free or serial recall, item recognition, source memory), domains (e.g., working memory, long-term memory), or representational modalities (e.g., visual memory, semantic memory), with limited cross-talk across these specialized concentrations. This is especially common in cognitive aging research (e.g., Salthouse, 2000), where attempts to integrate findings across disparate tasks, domains, or modalities within a common computational modelling framework have been rare (for an exception, see Healey & Kahana, 2016). Even in the mainstream memory literature, computational models have typically been formalized to explain the dynamics of a single type of task, like recognition (for reviews, see Clark & Gronlund, 1996; Humphreys et al., 2024) or serial recall (see, for example, Hurlstone, 2024; Hurlstone et al., 2014).

Yet, as our systematic empirical and modelling investigation shows, there is much more that unites different tasks of memory – like serial recall and item recognition – than there is that divides them. Empirically, we observed consistent and high rates of false memories for items that shared salient features with studied items, regardless of task format, representational modality, or age group. That is, in tasks of order reconstruction, serial recall, and item recognition, participants exhibited higher rates of false memory – falsely reconstructing, recalling, or recognizing an item that was never encountered before in the experiment – for unstudied items that were similar versus dissimilar to studied items. This elevated false memory for similar versus dissimilar items was observed when similarity was defined semantically (Experiments 1, 5, and 8), phonologically (Experiments 2, 6, and 9), orthographically (Experiments 3, 7, and 10), or visually (Experiments 4 and 11). Furthermore, it was consistently obtained for young and older adults alike across all task types and representational modalities. This was a striking finding, given the different emphasis that these tasks place on what needs to be remembered. Whereas order reconstruction emphasizes the need to remember the precise order of items with little to no need to reconstruct memory for the items themselves (Guitard et al., 2021; Guitard & Cowan, 2023), item recognition demands one remember the items but not their order (Gionet et al., 2024; Grenfell-Essam et al., 2017), and serial recall requires one remember both item and order information when no cues are provided for either (Waugh, 1961). Yet, despite these differing emphases, all tasks showed clear and consistent evidence of similarity effects on false memory.

Critically, we were able to simulate the performance of young and older adults alike on each task type, and across representational modalities, with the eCFM. By embedding structured representations into a model rooted in MINERVA 2 (Hintzman, 1984, 1986), we showed that the eCFM can reproduce the gradients of true and false memory in serial recall, order reconstruction, and item recognition, applying the same basic processing principles across task types. Furthermore, the structured representations allowed the eCFM to successfully capture similarity effects in specific representational domains (semantic, phonological, orthographic, or visual) without the need to engineer feature vectors in an attempt to simulate these effects. Instead, the eCFM learned each item's lexical representation through established techniques (Bingham & Mannila, 2001; Landauer & Dumais, 1997; Parrish, 2017; Whitney, 2001; Whitney & Marton, 2013) that allowed it to pick up on both strong and subtle inter-item similarities through stored associations among items in the model's lexicon (its "long-term memory"). This substantially mitigated experimenter degrees of freedom, while imbuing the model with structured representations that allowed it to perform like a human when confronted with lists of similar items. It also aligned the eCFM with contemporary trends in computational models of both recognition (Chang et al., 2025; Jamieson et al., 2018; Johns & Jones, 2010; Johns et al., 2012; Jones & Mewhort, 2007; Osth & Zhang, 2023; Reid & Jamieson, 2023) and recall (Kimball et al., 2007; Lohnas, 2024; Mewhort et al., 2018; Polyn et al., 2009; Sirotin et al., 2005), while innovating on these models in applying the same principles to both task types (and to order reconstruction).

A key insight of the modelling was its ability to provide insight on the conditions and processes underlying age-related differences in true and false memory across these diverse tasks and representational modalities. We revisit this insight in the next section, where we discuss the theoretical implications of our modelling for understanding age differences in memory. First, however, we conclude this section by briefly commenting on what it means for memory theory more broadly now that we have shown that order reconstruction, serial recall, and item recognition can all be modeled within a common computational framework. The main implication of this finding is that the same underlying processes are at play on widely different tasks – those that emphasize memory for order (order reconstruction), for items (item recognition), or for both (serial recall). Although previously such a claim might have been taken at face value, the work presented formalizes this assumption within a computational model that makes explicit how these latent processes give rise to observed performance on these different tasks (cf., Oberauer & Lewandowsky, 2019). By having formalized these assumptions and verified that relatively the same basic processes are involved in each of these tasks, we have positioned the eCFM as a candidate model for identifying commonalities across other tasks of memory (e.g., free recall, associative recognition) or cognition more generally. Furthermore, the success of the eCFM in explaining performance across these diverse tasks presents a general framework for understanding memory, one that is formal and can therefore be extended to other tasks with the potential to identify boundaries where this framework might break down. But for now, we have shown that the eCFM constitutes a more general model of memory and a promising tool for re-connecting and integrating different foci of memory research.

On Explaining Age Differences in Memory

One of the key innovations of our approach was to apply the eCFM to data from young *and* older adults across multiple tasks, thereby allowing us to identify potential common processes underlying age-related declines in performance on these disparate tasks. Age differences have been detected to varying degrees on serial recall (Golomb et al., 2008; Maylor et al., 1999; Naveh-Benjamin et al., 2007), tests of temporal order (Cabeza et al., 2000), and item recognition (Fraundorf et al., 2019; Rhodes et al., 2019). Although many theories have been advanced to explain these differences (Salthouse, 2016), efforts to formalize theories of aging in computational models have been quite limited. When models have been leveraged to identify processes underlying age differences in memory, they have typically been constrained to single tasks. For example, there have been attempts to model age differences in serial recall (Neath & Surprenant, 2007; Maylor et al., 1999; Surprenant et al., 2006) or in recognition (Benjamin, 2010; Buchler & Reder, 2007; Stephens & Overman, 2018), with rare attempts to jointly model age differences across tasks (but see Healey & Kahana, 2016).

Our approach presents a productive challenge to these prior modelling applications in several key ways. First, we jointly modeled age differences in tasks that place different demands on what needs to be remembered, namely those that emphasize remembering items (item recognition), their orders (order construction), or both (serial recall). The magnitude of age differences are typically larger on tasks of recall than recognition (Rhodes et al., 2019), a finding we confirmed in our empirical investigations, where age differences were detected most noticeably in serial recall but not in item recognition. The eCFM was able to account for the more substantial age-related declines in serial recall accuracy (Experiments 5-7), along with the more modest ones in temporal order reconstruction (Experiments 1-4), and the stability of performance in item recognition across age groups (Experiments 8-11), while also identifying potential mechanisms that could account for age differences in recognition that were not present in our data.

Second, the eCFM used structured representations, improving upon earlier attempts to explain age differences in memory with engineered representations in MINERVA 2 (e.g., Benjamin, 2010) or similar models (e.g., Stephens & Overman, 2018). As noted earlier, the use of structured representations allowed the eCFM to naturally pick up on inter-item similarities through established associations in the model's lexicon, thereby providing a principled way of explaining older adults' false memories across various representational modalities. This is an important feature of our analysis because it affords the eCFM with the ability to explain age differences in true and false memory through a host of mechanisms including, for example, that these differences could arise from the exact ways in which inter-item similarities are learned based on the size or composition of the lexicon. Although we did not investigate this possibility here, it constitutes a potential avenue for extending the eCFM to studying the effects of development and aging jointly, as by, for instance, gradually building up the model's lexicon (its long-term memory) through repeated experience, allowing it to accumulate new experiences each time (for a similar approach, see Johns et al., 2012).

Third, the eCFM identified both shared and unique mechanisms of age-related memory change across different tasks. Appealing to theoretical claims that episodic memories become "fuzzier" or less precise in older adulthood (Brainerd & Reyna, 2015; Greene & Naveh-Benjamin, 2023), we sought to simulate age differences in true and false memory by tuning parameters of the eCFM that affect the representational fidelity of its learned experiences (for similar approaches in other models, see Benjamin, 2010; Neath & Surprenant, 2007; Stephens & Overman, 2018; Surprenant et al., 2006). There are two parameters that control different aspects of the model's representational precision. The first,

L, governs the overall fidelity with which the model learns each item, dictating how many lexical and order features of each item are faithfully stored in memory. The second, d, dictates the precision of order representations specifically, by determining the degree to which the features used to represent the position of each item in a list repeat across positions. Because it is possible that an age difference could be based solely on a change in L or in d or through a joint change in both parameters, there were multiple ways in which a representational impairment could give rise to age-related memory decline, providing a more nuanced perspective to the claim that episodic memories become fuzzier with age (Greene & Naveh-Benjamin, 2023). Namely, that independent of retention, representations of order are more distinct for younger than older adults.

Indeed, our modelling revealed that on tests of order reconstruction for familiar word stimuli, including lists of semantically or phonologically related items, age differences in correct reconstruction were attributed to a reduction in d (positional precision) with age, with no concomitant reduction in L (global fidelity). That is, when the task of memory was to reconstruct the order of items provided as retrieval cues, and those items were familiar words, older adults retained a similar proportion of features of each item in memory as younger adults did ($L_{old} = L_{young}$), but more of the same features were used to represent the order of each item in the list ($d_{old} < d_{young}$), blurring their memories for which items appeared in which positions. However, for abstract visual items, there was an age-related reduction in both the overall fidelity of the items ($L_{old} < L_{young}$) and in their positional discriminability ($d_{old} < d_{young}$), in line with evidence of more degraded perceptual representations among older adults (Neath & Surprenant, 2007). Meanwhile, when the task of memory was to reconstruct both the items and their orders with no cues provided (i.e., serial recall), there were consistent age-related declines in both global fidelity and order precision (i.e., $L_{old} < L_{young}$; $d_{old} < d_{young}$), regardless of stimulus type. Thus, on serial recall tasks, age differences in the precision of memory are more expansive, affecting both how well each item is represented in memory and how discriminable the features that represent the position of each item are. Collectively, the simulation results of the order reconstruction and serial recall experiments support the principle that older adults' memories are generally less precise (Greene & Naveh-Benjamin, 2023), while showing that the degree to which an age-related reduction in the precision of memory manifests is, in part, task-dependent because different tasks do and do not rely on those representations for remembering.

Because we detected no age differences in our recognition experiments, it is perhaps unsurprising that parameters affecting the representational fidelity of the model's learned experiences did not differ between age groups in explaining true and false recognition. Although age differences in item recognition are typically smaller than those in recall (Rhodes et al., 2019), they are nonetheless often reported. It is possible that the use of online samples attenuated such differences in the present experiments (but see Greene & Naveh-Benjamin, 2022). However, this explanation seems unlikely given that clear age differences were still observed in serial recall and order reconstruction. A more plausible account is that intermixing lists of semantically (Experiment 8), phonologically (Experiment 9), orthographically (Experiment 10), and visually (Experiment 11) related items reduced reliance on category-specific gist, thereby minimizing group differences in gist-based processing (cf. Tun et al., 1998). Consistent with this interpretation, both young and older adults exhibited elevated false alarm rates to related relative to unrelated lures, and the eCFM captured this pattern using equivalent estimates of global fidelity (L) across age groups.

Nevertheless, the model remains capable of producing age-related differences in recognition by adjusting the learning probability (L) and retrieval sensitivity (τ). As demonstrated in our simulations (Figure 12), modest reductions in L (or alterations in τ) are sufficient to reproduce the selective increase in false recognition for related lures often observed in older adults. Thus, much like DRYAD (Benjamin, 2010), the eCFM, grounded in the same family of instance-based similarity principles, provides a mechanistically transparent framework for understanding both the stability and variability of recognition memory across age.

Summary and Future Directions

In summary, our aim was to offer researchers an articulate and flexible framework for modeling memory behavior. We applied the model to order reconstruction, serial recall, and recognition. Although these tasks are relatively simple, they provide rich empirical constraints that can serve as a foundation for refining and extending general models of memory. There are many additional avenues to explore, including free recall, which has been extensively modeled by Healey and Kahana (2016), and integrated accounts of serial and free recall recently advanced by Lohnas (2024). The eCFM could also be readily extended to cued recall, source memory, and other paradigms discussed in recent reviews (e.g., Jamieson et al., 2022).

In future work, we aim to broaden the model's generalization across tasks and across the human lifespan, from early development to older adulthood. The model's representational architecture makes it particularly well-suited to generate developmental predictions grounded in age-related changes to lexical structure and learning efficiency. While such applications remain speculative, they represent a natural next step for testing the boundaries of a general account of memory.

Before concluding, it is worth noting that the present work shares substantial conceptual overlap with several contemporary modeling approaches, including DRYAD (Benjamin, 2010), the frameworks advanced by Reid and Jamieson (2022; Reid et al., 2025), as well as recent developments by Chang et al. (2025). Similarities also exist with the Feature

Model (Neath & Surprenant, 2007), the OSCAR model (Brown et al., 2000; Maylor et al., 1999), and the SIMPLE model (Brown et al., 2007; Surprenant et al., 2006). Although our focus throughout this paper has been on highlighting distinctions between these approaches and the eCFM, we also view these efforts as converging toward a shared theoretical goal: developing general, process-level accounts of how memory representations are formed, maintained, and distorted across tasks and age groups. Each model articulates this goal in a different formal language, yet together they represent complementary attempts to build a integrated theory of memory. Developing a comprehensive model that spans tasks, materials, and developmental stages is a major challenge—but one well worth pursuing. Our work contributes to this objective by advancing Newell's (1973) vision of a general theory of memory, providing a formal and extensible framework for linking diverse memory phenomena within a common theoretical architecture.

Conclusion

Human memory is a rich and complex domain that demands principled frameworks capable of generating precise and generalizable predictions across tasks, materials, and populations. In this work, we took a step toward that ambitious goal with the eCFM (Guitard et al., 2025a, 2025b, 2025c). Across eleven empirical and computational investigations, the model successfully reproduced key features of memory performance across distinct representational modalities (orthographic, phonological, semantic, and visual), task demands (reconstruction, recall, recognition), and age groups (younger and older adults). Although this work represents an initial step, it highlights the power of integrating empirical precision with computational formalization to bridge longstanding divides in memory research. We hope these efforts inspire continued progress toward a more unified and generalizable science of human memory.

References

Abadie, M., Gavard, E., & Guillaume, F. (2021). Verbatim and gist memory in aging. *Psychology and Aging, 36*(8), 891–901. https://doi.org/10.1037/pag0000635

Achlioptas, D. (2003). Database-friendly random projections: Johnson–Lindenstrauss with binary coins. *Journal of Computer and System Sciences, 66*(4), 671–687. https://doi.org/10.1016/S0022-0000(03)00025-4

Arndt, J., & Hirshman, E. (1998). True and false recognition in MINERVA2: Explanations from a global matching perspective. *Journal of Memory and Language, 39*(3), 371–391. https://doi.org/10.1006/jmla.1998.2581

Balota, D. A., Cortese, M. J., Duchek, J. M., Adams, D., Roediger, H. L., McDermott, K. B., & Yerys, B. E. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer's type. *Cognitive Neuropsychology, 16*(3–5), 361–384. https://doi.org/10.1080/026432999380834

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods, 39*(3), 445–459. https://doi.org/10.3758/BF03193014

Benjamin, A. S. (2010). Representational explanations of "process" dissociations in recognition: The DRYAD theory of aging and memory judgments. *Psychological Review, 117*(4), 1055–1079. https://doi.org/10.1037/a0020810

Benjamin, A. S., Diaz, M., Matzen, L. E., & Johnson, B. (2012). "Tests of the DRYAD theory of the age-related deficit in memory for context: Not about context, and not about aging": Correction to Benjamin, Diaz, Matzen, and Johnson (2012). *Psychology and Aging, 27*(4), 824. https://doi.org/10.1037/a0031162

Bhatarah, P., Ward, G., Smith, J., Hayes, L. (2009). Examining the relationship between free recall and immediate serial recall: Similar patterns of rehearsal and similar effects of word length, presentation rate, and articulatory suppression. *Memory & Cognition,* 37, 689–713. https://doi.org/10.3758/MC.37.5.689

Bingham, E., & Mannila, H. (2001). Random projection in dimensionality reduction: Applications to image and text data. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 245–250). Association for Computing Machinery. https://doi.org/10.1145/502512.502546

Bailey, H., Dunlosky, J., & Hertzog, C. (2009). Does differential strategy use account for age-related deficits in working-memory performance? *Psychology and Aging, 24*(1), 82-92 https://doi.org/10.1037/a0014078

Ballardini, N., Yamashita, J. A., & Wallace, W. P. (2008). Presentation duration and false recall for semantic and phonological associates. *Consciousness and Cognition*, 17(1), 64-71. https://doi.org/10.1016/j.concog.2007.01.008

Bayen, U. J., Phelps, M. P., & Spaniol, J. (2000). Age-related differences in the use of contextual information in recognition memory: a global matching approach. *The journals of gerontology. Series B, Psychological sciences and social sciences*, *55*(3), P131–P141. https://doi.org/10.1093/geronb/55.3.p131

Brysbaert, M., New, B. & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods ,* 44, 991–997. https://doi.org/10.3758/s13428-012-0190-4

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics, 5*, 135–146. https://doi.org/10.1162/tacl_a_00051

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory.* Oxford University Press.

Brainerd, C. J., & Reyna, V. F. (2015). Fuzzy-trace theory and lifespan cognitive development. *Developmental Review, 38,* 89–121. https://doi.org/10.1016/j.dr.2015.07.006

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107*(1), 127–181. https://doi.org/10.1037/0033-295X.107.1.127

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review, 114*(3), 539–576. https://doi.org/10.1037/0033-295X.114.3.539

Buchler, N. E. G., & Reder, L. M. (2007). Modeling age-related memory deficits: A two-parameter solution. *Psychology and Aging, 22*(1), 104–121. https://doi.org/10.1037/0882-7974.22.1.104

Budson, A. E., Sullivan, A. L., Daffner, K. R., & Schacter, D. L. (2003). Semantic versus phonological false recognition in aging and Alzheimer's disease. *Brain and Cognition, 51*(3), 251–261. https://doi.org/10.1016/S0278-2626(03)00030-7

Cabeza, R., Anderson, N. D., Houle, S., Mangels, J. A., & Nyberg, L. (2000). Age-related differences in neural activity during item and temporal-order memory retrieval: A position emission tomography study. *Journal of Cognitive Neuroscience, 12*(1), 197–206. https://doi.org/10.1162/089892900561832

Caplan, J. B., & Guitard, D. (2024a). A feature-space theory of the production effect in recognition.*Experimental Psychology, 71*(1), 64–82. https://doi.org/10.1027/1618-3169/a000611

Caplan, J. B., & Guitard, D. (2024b). Stimulus duration and recognition memory: An attentional subsetting account. *Journal of Memory and Language, 139*, 104556. https://doi.org/10.1016/j.jml.2024.104556

Caplan, J. B., & Guitard, D. (2025). Inverted list-strength effects in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* Advance online publication. https://doi.org/10.1037/xlm0001489

Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. *Psychology and Aging, 20*(4), 718–721. https://doi.org/10.1037/0882-7974.20.4.718

Chang, M., & Brainerd, C. J. (2021). Semantic and phonological false memory: A review of theory and data. *Journal of Memory and Language*, *119*, 104210. https://doi.org/10.1016/j.jml.2020.104210

Chang, M., Johns, B. T., & Brainerd, C. J. (2025). True and false recognition in MINERVA2: Integrating fuzzy-trace theory and computational memory modeling. *Psychological Review.* Advance online publication. https://doi.org/10.1037/rev0000541

Cox, G. E., Kachergis, G., Recchia, G., & Jones, M. N. (2011). Toward a scalable holographic word-form representation. *Behavior Research Methods*, *43*, 602-615. https://doi.org/10.3758/s13428-011-0125-5

Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society B: Biological Sciences, 302*(1110), 341-359. https://doi.org/10.1098/rstb.1983.0059

Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 191-211). Plenum Publishing.

Criss, A. H., & Shiffrin, R. M. (2005). List discrimination in associative recognition and implications for representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1199–1212. https://doi.org/10.1037/0278-7393.31.6.1199

Cyr, V., Poirier, M., Yearsley, J. M., Guitard, D., Harrigan, I., & Saint-Aubin, J. (2022). The production effect over the long term: Modeling distinctiveness using serial positions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(12), 1797–1820. https://doi.org/10.1037/xlm0001093

Darby, K. P., & Sederberg, P. B. (2022). Transparency, replicability, and discovery in cognitive aging research: A computational modeling approach. *Psychology and Aging, 37*(1), 10–29. https://doi.org/10.1037/pag0000665

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The “Small World of Words” English word association norms for over 12,000 cue words. *Behavior Research Methods, 51*(3), 987–1006. https://doi.org/10.3758/s13428-018-1115-7

Deese. J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17–22. https://doi.org/10.1037/h0046671

Dennis, N. A., Kim, H., & Cabeza, R. (2008). Age-related differences in brain activity during true and false memory retrieval. *Journal of Cognitive Neuroscience, 20*(8), 1390–1402. https://doi.org/10.1162/jocn.2008.20096

Devitt, A. L., & Schacter, D. L. (2016). False memories with age: Neural and cognitive underpinnings. *Neuropsychologia, 91*, 346-359. https://doi.org/10.1016/j.neuropsychologia.2016.08.030

Dienes, Z. (1992). Connectionist and memory-array models of artificial grammar learning. Cognitive Science, 16, 41–79. https://doi.org/10.1016/0364-0213(92)90017-O

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fawcett, J. M., Baldwin, M. M., Whitridge, J. W., Swab, M., Malayang, K., Hiscock, B., Drakes, D. H., & Willoughby, H. V. (2023). Production improves recognition and reduces intrusions in between-subject designs: An updated meta-analysis. *Canadian Journal of Experimental Psychology / Revue canadienne de psychologie expérimentale, 77*(1), 35–44. https://doi.org/10.1037/cep0000302

Fraundorf, S. H., Hourihan, K. L., Peters, R. A., & Benjamin, A. S. (2019). Aging and recognition memory: A meta-analysis. *Psychological Bulletin, 145*(4), 339–371. https://doi.org/10.1037/bul0000185

Golomb, J. D., Peele, J. E., Addis, K. M., Kahana, M. J., & Wingfield, A. (2008). Effects of adult aging on utilization of temporal and semantic associations during free and serial recall. *Memory & Cognition, 36*(5), 947–956. https://doi.org/10.3758/MC.36.5.947

Gionet, S., Guitard, D., & Saint-Aubin, J. (2024). The interaction between the production effect and serial position in recognition and recall. *Experimental Psychology, 71*(5), 259–277. https://doi.org/10.1027/1618-3169/a000623

Greene, N. R., & Naveh-Benjamin, M. (2020). A specificity principle of memory: Evidence from aging and associative memory. *Psychological Science, 31*(3), 316-331. https://doi.org/10.1177/0956797620901760

Greene, N. R., & Naveh-Benjamin, M. (2022). Adult age differences in specific and gist associative episodic memory across short- and long-term retention intervals. *Psychology and Aging, 37*(6), 681–697. https://doi.org/10.1037/pag0000701

Greene, N. R., & Naveh-Benjamin, M. (2023). Adult age-related changes in the specificity of episodic memory representations: A review and theoretical framework. *Psychology and Aging, 38*(2), 67–86. https://doi.org/10.1037/pag0000724

Greene, N. R., & Naveh-Benjamin, M. (2024). The time course of encoding specific and gist episodic memory representations among young and older adults. *Journal of*

Experimental Psychology: General, 153(6), 1671-1697. https://doi.org/10.1037/xge0001589

Greene, N. R., Chism, S., & Naveh-Benjamin, M. (2022). Levels of specificity in episodic memory: Insights from response accuracy and subjective confidence ratings in older adults and in younger adults under full or divided attention. *Journal of Experimental Psychology: General, 151*(4), 804–819. https://doi.org/10.1037/xge0001113

Greene, N. R., Guitard, D., & Naveh-Benjamin, M. (2025). Process explanations of age-related changes in memory specificity. *Psychology & Aging*.

Grenfell-Essam, R., Ward, G., & Tan, L. (2017). Common modality effects in immediate free recall and immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(12), 1909–1933. https://doi.org/10.1037/xlm0000430

Guitard, D., & Cowan, N. (2023). The tradeoff between item and order information in short term memory does not depend on encoding time. *Journal of Experimental Psychology: Human Perception and Performance, 49*(1), 51–70. https://doi.org/10.1037/xhp0001074

Guitard, D., Saint-Aubin, J., & Cowan, N. (2021). Asymmetrical interference between item and order information in short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(2), 243–263. https://doi.org/10.1037/xlm0000956

Guitard, D., Saint-Aubin, J., & Cowan, N. (2022). Tradeoffs between item and order information in short-term memory. *Journal of Memory and Language, 122*, 104300. https://doi.org/10.1016/j.jml.2021.104300

Guitard, D., Saint-Aubin, J., Reid, J. N., & Jamieson, R. K. (2025a). An embedded computational framework of memory: Accounting for the influence of semantic information in verbal short-term memory. *Journal of Memory and Language, 140*, 104573. https://doi.org/10.1016/j.jml.2024.104573

Guitard, D., Saint-Aubin, J., Reid, J. N., & Jamieson, R. K. (2025b). An embedded computational framework of memory: The critical role of representations in veridical and false recall predictions. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-025-02669-7

Guitard, D., Cowan, N., Saint-Aubin, J., Reid, J. N., & Jamieson, R. K. (2025c). Toward a comprehensive account of verbal memory: an embedded computational model across representational domains. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Hannagan, T., Dupoux, E. and Christophe, A. (2011), Holographic String Encoding. Cognitive Science, 35: 79-118. https://doi.org/10.1111/j.1551-6709.2010.01149.x

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory,* Vol. 22, pp. 193–225). Academic Press. https://doi.org/10.1016/S0079-7421(08)60041-9

Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review, 123*(1), 23–69. https://doi.org/10.1037/rev0000015

Hedayati, S., O'Donnell, R. E., & Wyble, B. (2022). A model of working memory for latent representations. *Nature Human Behaviour, 6*(5), 709–719. https://doi.org/10.1038/s41562-021-01264-9

Henson, R. N. A. (1998). Short-term memory for serial order: The Start–End Model. *Cognitive Psychology, 36*(2), 73–137. https://doi.org/10.1006/cogp.1998.0685

Hertzog, C., McGuire, C. L., & Lineweaver T. T. (1998). Aging, attributions, perceived control and strategy use in a free recall task. *Aging, Neuropsychlogy, and Cognition, 5*(2), 85-106. https://doi.org/10.1076/anec.5.2.85.601

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments & Computers, 16*(2), 96–101. https://doi.org/10.3758/BF03202365

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*(4), 411–428. https://doi.org/10.1037/0033-295X.93.4.411

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin, 140*(2), 339–373. https://doi.org/10.1037/a0034221

Hurlstone, Mark J., 'Serial Recall', in Michael J. Kahana, and Anthony D. Wagner (eds), *The Oxford Handbook of Human Memory, Two Volume Pack: Foundations and Applications*, Oxford Library of Psychology (2024; online edn, Oxford Academic, 18 July 2024), https://doi.org/10.1093/oxfordhb/9780190917982.013.28

Humphreys, M. S., Hockley, W. E., & Chalmers, K. A. (2024). Recognition memory: The probe, the returned signal, and the decision. *Psychonomic Bulletin & Review, 31*(2), 568–598. https://doi.org/10.3758/s13423-021-01955-4

Jamieson, R. K., & Mewhort, D. J. K. (2010). Applying an exemplar model to the artificial grammar task: String completion and performance on individual items. *The Quarterly Journal of Experimental Psychology*, 63(5), 1014–1039. https://doi.org/10.1080/17470210903267417

Jamieson, R. K., Avery, J. E., Johns, B. T., & Jones, M. N. (2018). An instance theory of semantic memory. *Computational Brain & Behavior, 1*(2), 119–136. https://doi.org/10.1007/s42113-018-0008-2

Jamieson, R. K., Johns, B. T., Vokey, J. R., & Jones, M. N. (2022). Instance theory as a domain-general framework for cognitive psychology. *Nature Reviews Psychology, 1*(3), 174–183. https://doi.org/10.1038/s44159-022-00025-3

Jennings, J. M., & Jacoby, L. L. (1997). An opposition procedure for detecting age-related deficits in recollection: Telling effects of repetition. *Psychology and Aging, 12*(2), 352-361. https://doi.org/10.1037/0882-7974.12.2.352

Jones, M. N. (2019). When does abstraction occur in semantic memory: Insights from distributional models. *Language, Cognition and Neuroscience, 34*(10), 1338–1346. https://doi.org/10.1080/23273798.2018.1431679

Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review, 17*(5), 662–672. https://doi.org/10.3758/PBR.17.5.662

Johns, B. T., & Jamieson, R. K. (2019). The influence of place and time on lexical behavior: A distributional analysis. *Behavior Research Methods, 51*(6), 2438–2453. https://doi.org/10.3758/s13428-019-01289-z

Johns, B. T., Jones, M. N., & Mewhort, D. J. K. (2012). A synchronization account of false recognition. *Cognitive Psychology, 65*(4), 486–518. https://doi.org/10.1016/j.cogpsych.2012.07.002

Johns, B. T., Jamieson, R. K., Crump, M. J. C., Jones, M. N., & Mewhort, D. J. K. (2020). Production without rules: Using an instance memory model to exploit structure in natural language. *Journal of Memory and Language, 115*, 104165. https://doi.org/10.1016/j.jml.2020.104165

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*, 1-37. https://doi.org/10.1037/0033-295X.114.1.1

Kahana, M. J. (2020). Computational models of memory search. *Annual Review of Psychology, 71,* 107–138. https://doi.org/10.1146/annurev-psych-010418-103358

Kahana, M. J., Diamond, N. B., & Aka, A. (2024). Laws of human memory. In M. J. Kahana & A. D. Wagner (Eds.), *The Oxford handbook of human memory: Foundations and applications* (online ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190917982.013.2

Kahana, M. J., Howard, M. W., Zaromb, F., & Wingfield, A. (2002). Age dissociates recency and lag recency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 530–540. https://doi.org/10.1037/0278-7393.28.3.530

Kahana, M. J., Zhou, F., Geller, A. S., & Sekuler, R. (2007). Lure similarity affects visual episodic recognition: Detailed tests of a noisy exemplar model. *Memory & Cognition, 35*(6), 1222–1232. https://doi.org/10.3758/BF03193596

Kimball, D. R., Smith, T. A., & Kahana, M. J. (2007). The fSAM model of false recall. *Psychological Review*, *114*, 954–993. https://doi.org/10.1037/0033-295X.114.4.954

Koutstaal, W. (2003). Older adults encode--but do not always use--perceptual details: intentional versus unintentional effects of detail on memory judgments. *Psychological Science*, *14*(2), 189–193. https://doi.org/10.1111/1467-9280.01441

Kouststaal, W., Reddy, C., Jackson, E. M., Prince, S., Cendan, D. L., & Schacter, D. L. (2003). False recognition of abstract versus common objects in older and younger adults: Testing the semantic categorization account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(4), 499–510. https://doi.org/10.1037/0278-7393.29.4.499

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language, 37*(4), 555–583. https://doi.org/10.1006/jmla.1997.2529

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. Discourse Processes, 25(2-3), 259–284. https://doi.org/10.1080/01638539809545028

Lawrence, M. A. (2016). *ez: Easy analysis and visualization of factorial experiments* (Version 4.4-0) [R package]. https://CRAN.R-project.org/package=ez

Levine, B., Svoboda, E., Hay, J. F., Winocur, G., & Moscovitch, M. (2002). Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychology and Aging, 17*(4), 677–689. https://doi.org/10.1037/0882-7974.17.4.677

Li, S.-C., Naveh-Benjamin, M., & Lindenberger, U. (2005). Aging Neuromodulation Impairs Associative Binding: Neurocomputational Account. *Psychological Science*, *16*(6), 445-450. https://doi.org/10.1111/j.0956-7976.2005.01555.x

Li, P., Hastie, T. J., & Church, K. W. (2006). Very sparse random projections. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 287–296). Association for Computing Machinery. https://doi.org/10.1145/1150402.1150436

Light, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology, 42,* 333–376. https://doi.org/10.1146/annurev.ps.42.020191.002001

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, 95(4), 492–527. https://doi.org/10.1037/0033-295X.95.4.492

Lohnas, L. J. (2024). A retrieved context model of serial recall and free recall. *Computational Brain & Behavior*.. https://doi.org/10.1007/s42113-024-00221-9

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological Review, 122*(2), 337–363. https://doi.org/10.1037/a0039036

Majerus, S. (2019). Verbal working memory and the phonological buffer: The question of serial order. *Cortex, 112*, 122–133. https://doi.org/10.1016/j.cortex.2018.04.016

Maylor, E. A., Vousden, J. I., & Brown, G. D. A. (1999). Adult age differences in short-term memory for serial order: Data and a model. *Psychology and Aging, 14*(4), 572–594. https://doi.org/10.1037/0882-7974.14.4.572

McRae, K., Cree, G. S., Seidenberg, M. S., & Mcnorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*(4), 547–559. https://doi.org/10.3758/BF03192726

Mewhort, D.J.K., Shabahang, K.D. & Franklin, D.R.J. (2018). Release from PI: An analysis and a model. *Psychonomic Bulletin & Rev*iew, *25*, 932–950. https://doi.org/10.3758/s13423-017-1327-3

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. https://arxiv.org/abs/1301.3781

Morey, R. D., & Rouder, J. N. (2024). *BayesFactor: Computation of Bayes factors for common designs* (Version 0.9.12-4.7) [R package]. https://CRAN.R-project.org/package=BayesFactor

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General, 128*(4), 403–415. https://doi.org/10.1037/0096-3445.128.4.403

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*(3), 251–269. https://doi.org/10.3758/BF03213879

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(5), 1170–1187. https://doi.org/10.1037/0278-7393.26.5.1170

Naveh-Benjamin, M., & Cowan, N. (2023). The roles of attention, executive function and knowledge in cognitive ageing of working memory. *Nature Reviews Psychology, 2*(3), 151–165. https://doi.org/10.1038/s44159-023-00149-0

Naveh-Benjamin, M., Cowan, N., Kilb, A., & Chen, Z. (2007). Age-related differences in immediate serial recall: Dissociating chunk formation and capacity. *Memory & Cognition, 35*(4), 724–737. https://doi.org/10.3758/BF03193310

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 402–407. https://doi.org/10.3758/BF03195588

Neisser, U. (1967). *Cognitive psychology.* Appleton-Century-Crofts.

Neath, I. (1993). Distinctiveness and serial position effects in recognition. *Memory & Cognition, 21*(5), 689–698. https://doi.org/10.3758/BF03197199

Neath, I. (1999). Computer simulations of global memory models. *Behavior Research Methods, Instruments, and Computers, 31*(1), 74–80. https://doi.org/10.3758/BF03207695

Neath, I., Saint-Aubin, J., & Surprenant, A. M. (2022). Semantic relatedness effects in serial recall but not in serial reconstruction of order. *Experimental Psychology, 69*(4), 196–209. https://doi.org/10.1027/1618-3169/a000557

Newell, A. (1973). You can't play 20 questions with nature and win. In W. G. Chase (Ed.), *Visual information processing* (pp. 283–308). Academic Press.

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition, 25*(6), 838-848. https://doi.org/10.3758/BF03211328

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(4), 700–708. https://doi.org/10.1037/0278-7393.14.4.700

Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review, 26*(5), 1596–1618. https://doi.org/10.3758/s13423-019-01645-2

Oberauer, K., & Lin, H.-Y. (2024). An interference model for visual and verbal working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 50*(6), 858–888. https://doi.org/10.1037/xlm0001303

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D. E., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin, 144*(9), 885–958. https://doi.org/10.1037/bul0000153

Osth, A. F., & Hurlstone, M. J. (2023). Do item-dependent context representations underlie serial order in cognition? Commentary on Logan (2021). *Psychological Review, 130*(2), 513–545. https://doi.org/10.1037/rev0000352

Osth, A. F., & Zhang, L. (2023). Integrating word-form representations with global similarity computation in recognition memory. https://doi.org/10.31234/osf.io/kemfw

Osth, A. F., Shabahang, K. D., Mewhort, D. J. K., & Heathcote, A. (2020). Global semantic similarity effects in recognition memory: Insights from BEAGLE representations and the diffusion decision model. *Journal of Memory and Language, 111,* Article 104071. https://doi.org/10.1016/j.jml.2019.104071

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*(4), 761–781. https://doi.org/10.1037/0033-295X.105.4.761-781

Parrish, A. (2017, October). Poetic sound similarity vectors using Phonetic Features [Paper presentation]. AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, Snowbird, Utah, USA. https://www.aaai.org/ocs/index.php/AIIDE/AIIDE17/paper/view/15879/15227

Peich, M.-C., Husain, M., & Bays, P. M. (2013). Age-related decline of precision and binding in visual working memory. *Psychology and Aging, 28*(3), 729–743. https://doi.org/10.1037/a0033236

Perlmutter, M. (1978). What is memory aging the aging of? *Developmental Psychology, 14*(4), 330-345. https://doi.org/10.1037/0012-1649.14.4.330

Piguet, O., Connally, E., Krendl, A. C., Huot, J. R., & Corkin, S. (2008). False memory in aging: effects of emotional valence on word recognition accuracy. *Psychology and Aging*, *23*(2), 307–314. https://doi.org/10.1037/0882-7974.23.2.307

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*, 129–156. https://doi.org/10.1037/a0014420

Popov, V., & Reder, L. M. (2020). Frequency effects on memory: A resource-limited theory. *Psychological Review, 127*(1), 1–46. https://doi.org/10.1037/rev0000161

Radvansky, G. A., Zwaan, R. A., Curiel, J. M., & Copeland, D. E. (2001). Situation models and aging. *Psychology and Aging, 16*(1), 145–160. https://doi.org/10.1037/0882-7974.16.1.145

Rankin, J. L., & Kausler, D. H. (1979). Adult age differences in false recognitions. *Journal of Gerontology, 34*(1), 58–65. https://doi.org/10.1093/geronj/34.1.58

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language, 50*(4), 408-424. https://doi.org/10.1016/j.jml.2003.11.002

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(2), 294–320. https://doi.org/10.1037/0278-7393.26.2.294

Reid, J. N., & Jamieson, R. K. (2022). A computational model of directed forgetting. *Canadian Journal of Experimental Psychology, 76*(2), 75–86. https://doi.org/10.1037/cep0000281

Reid, J. N., & Jamieson, R. K. (2023). True and false recognition in MINERVA 2: Extension to sentences and metaphors. *Journal of Memory and Language*, *129*, 104397. https://doi.org/10.1016/j.jml.2022.104397

Reid, J. N., Yang, H., & Jamieson, R. K. (2023). A computational account of item-based directed forgetting for nonwords: Incorporating orthographic representations in

MINERVA 2. *Memory & Cognition.* Advance online publication. https://doi.org/10.3758/s13421-023-01433-3

Reid, J. N., Guitard, D., & Jamieson, R. K. (2025). MINERVA OPS: A computational framework for the representation and recognition of orthographic, phonological, and semantic associates. *Psychonomic Bulletin & Review*.

Rhodes, S., Greene, N. R., & Naveh-Benjamin, M. (2019). Age-related differences in recall and recognition: A meta-analysis. *Psychonomic Bulletin & Review, 26*(5), 1529–1547. https://doi.org/10.3758/s13423-019-01649-y

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Roediger, H. L. III. (2008). Relativity of remembering: Why the laws of memory vanished. *Annual Review of Psychology, 59,* 225–254. https://doi.org/10.1146/annurev.psych.57.102904.190139

Roediger, H. L. & McDermott, K. B. (1995) Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814. https://doi.org/10.1037/0278-7393.21.4.803

Roodenrys, S., Guitard, D., Miller, L. M., Saint-Aubin, J., & Barron, J. M. (2022). Phonological similarity in the serial recall task hinders item recall, not just order. *British Journal of Psychology*, 113, 1100–1120. https://doi.org/10.1111/bjop.12575

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89*(1), 63–77. https://doi.org/10.1037/h0031185

Sahlgren, M., Holst, A., & Kanerva, P. (2008). Permutations as a means to encode order in word space. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), Proceedings of the 30th Annual Conference of the Cognitive Science Society (pp. 1300–1305). Austin: Cognitive Science Society.

Saint-Aubin, J., Yearsley, J. M., Poirier, M., Cyr, V., & Guitard, D. (2021). A model of the production effect over the short-term: The cost of relative distinctiveness. *Journal of Memory and Language, 118*, 104219. https://doi.org/10.1016/j.jml.2021.104219

Saint-Aubin, J., Poirier, M., Yearsley, J. M., & Guitard, D. (2024). The production effect becomes spatial. *Experimental Psychology, 71*(1), 14–32. https://doi.org/10.1027/1618-3169/a000609

Salthouse, T. A. (1988). Initiating the formalization of theories of cognitive aging. *Psychology and Aging, 3*(1), 3–16. https://doi.org/10.1037/0882-7974.3.1.3

Semon, R. (1923). *Mnemic psychology* (B. Duffy, Trans.). George Allen & Unwin. (Original work published 1909)

Schacter, D. L., Koutstaal, W., & Norman, K. A. (1997). False memories and aging. *Trends in Cognitive Sciences*, *1*(6), 229–236. https://doi.org/10.1016/S1364-6613(97)01068-1

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*(2), 145–166. https://doi.org/10.3758/BF03209391

Sekuler, R., McLaughlin, C., Kahana, M. J., Wingfield, A., & Yotsumoto, Y. (2006). Short-term visual recognition and temporal order memory are both well-preserved in aging. *Psychology and Aging, 21*(3), 632–637. https://doi.org/10.1037/0882-7974.21.3.632

Sirotin, Y. B., Kimball, D. R., & Kahana, M. J. (2005). Going beyond a single list: Modeling the effects of prior experience on episodic free recall. *Psychonomic Bulletin & Review, 12*(5), 787–805. https://doi.org/10.3758/BF03196773

Smith, A. D. (1975). Partial learning and recognition memory in the aged. *The International Journal of Aging and Human Development, 6*(4), 359–365. https://doi.org/10.2190/A241-Vu10-M52Q-KCDM

Spear, J., Reid, J. N., Guitard, D., & Jamieson, R. K. (2024). Directed Forgetting and the Production Effect: Assessing Strength and Distinctiveness. *Experimental Psychology*, *71*(5), 278–297. https://doi.org/10.1027/1618-3169/a000630

Spens, E., & Burgess, N. (2024). A generative model of memory construction and consolidation. *Nature Human Behaviour, 8*(4), 526–543. https://doi.org/10.1038/s41562-023-01799-z

Stark, S. M., Yassa, M. A., Lacy, J. W., & Stark, C. E. (2013). A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia, 51*(12), 2442–2449. https://doi.org/10.1016/j.neuropsychologia.2012.12.014

Stephens, J. D. W., & Overman, A. A. (2018). Modeling age differences in effects of pair repetition and proactive interference using a single parameter. *Psychology and Aging, 33*(1), 182–194. https://doi.org/10.10137/pag0000195

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolKit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology, 44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Surprenant, A. M., & Neath, I. (2009). *Principles of memory.* Psychology Press.

Surprenant, A. M., Neath, I., & Brown, G. D. A. (2006). Modeling age-related differences in immediate memory using SIMPLE. *Journal of Memory and Language, 55*(4), 572–586. https://doi.org/10.1016/j.jml.2006.08.001

Trelle, A. N., Henson, R. N., Green, D. A. E., & Simons, J. S. (2017). Declines in representational quality and strategic retrieval processes contribute to age-related increases in false recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(12), 1883–1897. https://doi.org/10.1037/xlm0000412

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging, 13*(2), 230–241. https://doi.org/10.1037/0882-7974.13.2.230

Watson, J. M., Balota, D. A., & Sergent-Marshall, S. D. (2001). Semantic, phonological, and hybrid veridical and false memories in healthy older adults and in individuals with dementia of the Alzheimer type. *Neuropsychology, 15*(2), 254–267. https://doi.org/10.1037/0894-4105.15.2.254

Waugh, N. C. (1961). Free versus serial recall. *Journal of Experimental Psychology, 62*(5), 496–502. https://doi.org/10.1037/h0043891

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. Psychonomic Bulletin & Review, 8, 221–243. https://doi.org/10.3758/BF03196158

Whitney, C., & Marton, Y. (2013). The SERIOL2 model of orthographic processing. Retrieved from https://www.researchgate.net/publication/237065841_The_SERIOL2_Model_of_Orthographic_Processing

Yassa, M. A., Mattfeld, A. T., Stark, S. M., & Stark, C. E. (2011). Age-related memory deficits linked to circuit-specific disruptions in the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(21), 8873–8878. https://doi.org/10.1073/pnas.1101567108

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In T. A. Salthouse & F. I. M. Craik (Eds.), *Handbook of aging and cognition* (2nd ed., pp. 293-257). Lawrence Erlbaum.

Zhang, L., & Osth, A. F. (2024). Modelling orthographic similarity effects in recognition memory reveals support for open bigram representations of letter coding. *Cognitive Psychology, 148*, 101619. https://doi.org/10.1016/j.cogpsych.2023.101619

Appendix A

Semantically related words used in Experiment 1 and Experiment 5.

Version 1							
Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Lure	Condition
cabinet	paper	folder	drawer	document	misfiling	file	Related
fake	cheat	lie	crime	false	money	fraud	Related
golf	member	ball	dance	organization	house	club	Related
guitar	treble	drum	fish	music	boom	bass	Related
hold	tight	vise	chisel	tool	metal	clamp	Related
light	camera	bulb	bright	back	flood	flash	Related
methane	station	energy	stove	heat	liquid	gas	Related
oyster	seafood	shell	chowder	pearl	mussel	clam	Related
stop	pedal	car	clutch	accelerate	speed	brake	Related
window	crystal	cup	bottle	clear	jar	glass	Related
beef	pork	cook	turkey	oven	dinner	pink	Unrelated
bird	peace	white	beak	bar	feather	leaf	Unrelated
jet	air	fly	sky	travel	geometry	yard	Unrelated
mile	meter	inch	grass	stick	foot	plane	Unrelated
panther	pretty	purple	lemonade	rose	dress	roast	Unrelated
scared	fright	terror	anxiety	monster	snake	vest	Unrelated
suit	jacket	shirt	blouse	coat	pants	dove	Unrelated
throat	tie	collar	necklace	shoulder	long	fear	Unrelated
tonic	alcohol	vodka	drink	liquor	drunk	neck	Unrelated
tree	maple	branch	flower	fall	pot	gin	Unrelated
Version 2							
Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Lure	Condition
beef	pork	cook	turkey	oven	dinner	roast	Related
bird	peace	white	beak	bar	feather	dove	Related
jet	air	fly	sky	travel	geometry	plane	Related
mile	meter	inch	grass	stick	foot	yard	Related
panther	pretty	purple	lemonade	rose	dress	pink	Related
scared	fright	terror	anxiety	monster	snake	fear	Related
suit	jacket	shirt	blouse	coat	pants	vest	Related
throat	tie	collar	necklace	shoulder	long	neck	Related
tonic	alcohol	vodka	drink	liquor	drunk	gin	Related
tree	maple	branch	flower	fall	pot	leaf	Related
cabinet	paper	folder	drawer	document	misfiling	club	Unrelated
fake	cheat	lie	crime	false	money	clam	Unrelated
golf	member	ball	dance	organization	house	flash	Unrelated
guitar	treble	drum	fish	music	boom	gas	Unrelated
hold	tight	vise	chisel	tool	metal	fraud	Unrelated
light	camera	bulb	bright	back	flood	clamp	Unrelated
methane	station	energy	stove	heat	liquid	brake	Unrelated
oyster	seafood	shell	chowder	pearl	mussel	glass	Unrelated
stop	pedal	car	clutch	accelerate	speed	file	Unrelated
window	crystal	cup	bottle	clear	jar	bass	Unrelated

Note. Each row represents a list. In Experiment 1, half of the participants received Version 1 and the other half Version 2, with all lists from their assigned version. In Experiment 5, participants were tested only on the 20 related lists.

Appendix B

Phonologically related words used in Experiment 2 and Experiment 6.

Version 1							
Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Lure	Condition
barge	ban	bang	bash	bat	batch	badge	Related
birth	balm	barb	barn	bard	path	bath	Related
burp	chap	chop	churn	church	chip	chirp	Related
calf	cuff	cob	con	cop	cot	cough	Related
chase	check	choice	guess	yes	less	chess	Related
date	dot	doubt	hurt	shirt	dart	dirt	Related
deaf	den	deck	debt	dead	doth	death	Related
fig	cog	dog	hog	jog	log	fog	Related
fun	fuzz	budge	judge	forge	nudge	fudge	Related
jerk	choke	poke	soak	woke	yolk	joke	Related
chum	dumb	hum	rum	sum	mum	teeth	Unrelated
fat	feet	fought	fight	fit	soot	mesh	Unrelated
gas	goose	cease	piece	niece	lease	turf	Unrelated
mush	mash	marsh	mess	met	men	foot	Unrelated
noose	nerve	curse	purse	verse	worse	thief	Unrelated
seethe	sued	booth	soup	suit	soon	nurse	Unrelated
ship	shape	shop	shark	harp	carp	thumb	Unrelated
teach	team	tease	teethe	heath	wreath	soothe	Unrelated
terse	term	turn	tiff	tough	surf	geese	Unrelated
theme	sheaf	reef	leaf	beef	chief	sharp	Unrelated
Version 2							
Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Lure	Condition
chum	dumb	hum	rum	sum	mum	thumb	Related
fat	feet	fought	fight	fit	soot	foot	Related
gas	goose	cease	piece	niece	lease	geese	Related
mush	mash	marsh	mess	met	men	mesh	Related
noose	nerve	curse	purse	verse	worse	nurse	Related
seethe	sued	booth	soup	suit	soon	soothe	Related
ship	shape	shop	shark	harp	carp	sharp	Related
teach	team	tease	teethe	heath	wreath	teeth	Related
terse	term	turn	tiff	tough	surf	turf	Related
theme	sheaf	reef	leaf	beef	chief	thief	Related
barge	ban	bang	bash	bat	batch	cough	Unrelated
birth	balm	barb	barn	bard	path	dirt	Unrelated
burp	chap	chop	churn	church	chip	joke	Unrelated
calf	cuff	cob	con	cop	cot	badge	Unrelated
chase	check	choice	guess	yes	less	fudge	Unrelated
date	dot	doubt	hurt	shirt	dart	bath	Unrelated
deaf	den	deck	debt	dead	doth	fog	Unrelated
fig	cog	dog	hog	jog	log	death	Unrelated
fun	fuzz	budge	judge	forge	nudge	chess	Unrelated
jerk	choke	poke	soak	woke	yolk	chirp	Unrelated

Note. Each row represents a list. In Experiment 2, half of the participants received Version 1 and the other half Version 2, with all lists from their assigned version. In Experiment 6, participants were tested only on the 20 related lists.

Appendix C

Orthographically related nonwords used in Experiment 3 and Experiment 7.

Version 1							
Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Lure	Condition
bapte	dapte	vapte	swapte	prapte	blapte	hapte	Related
bazy	grazy	prazy	zazy	wazy	kazy	pazy	Related
creck	bleck	freck	breck	sneck	teck	weck	Related
fration	jation	kration	snation	plation	gration	kation	Related
fratop	satop	dratop	tratop	tatop	snatop	datop	Related
jique	nique	shique	plique	frique	wique	zique	Related
klalky	phalky	flalky	kralky	pralky	galky	valky	Related
kleagle	creagle	pleagle	greagle	preagle	sweagle	fleagle	Related
priffy	triffy	siffy	bliffy	ziffy	griffy	tiffy	Related
saise	vaise	snaise	paise	jaise	taise	zaise	Related
drigin	tigin	prigin	trigin	digin	higin	swowen	Unrelated
druss	zuss	nuss	kruss	gruss	kuss	preskto	Unrelated
dumme	grumme	pumme	wumme	swumme	krumme	kloom	Unrelated
krowen	towen	powen	trowen	clowen	frowen	kigin	Unrelated
phedi	nedi	kedi	bredi	hedi	kledi	grolor	Unrelated
sheskto	jeskto	greskto	meskto	kleskto	sneskto	vuter	Unrelated
snender	nender	hender	dender	wender	trender	pruss	Unrelated
snuter	pluter	bruter	huter	kluter	cruter	shedi	Unrelated
tolor	swolor	drolor	prolor	crolor	pholor	klender	Unrelated
troom	froom	kroom	floom	snoom	voom	klumme	Unrelated
Version 2							
Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Lure	Condition
drigin	tigin	prigin	trigin	digin	higin	kigin	Related
druss	zuss	nuss	kruss	gruss	kuss	pruss	Related
dumme	grumme	pumme	wumme	swumme	krumme	klumme	Related
krowen	towen	powen	trowen	clowen	frowen	swowen	Related
phedi	nedi	kedi	bredi	hedi	kledi	shedi	Related
sheskto	jeskto	greskto	meskto	kleskto	sneskto	preskto	Related
snender	nender	hender	dender	wender	trender	klender	Related
snuter	pluter	bruter	huter	kluter	cruter	vuter	Related
tolor	swolor	drolor	prolor	crolor	pholor	grolor	Related
troom	froom	kroom	floom	snoom	voom	kloom	Related
bapte	dapte	vapte	swapte	prapte	blapte	tiffy	Unrelated
bazy	grazy	prazy	zazy	wazy	kazy	datop	Unrelated
creck	bleck	freck	breck	sneck	teck	zique	Unrelated
fration	jation	kration	snation	plation	gration	zaise	Unrelated
fratop	satop	dratop	tratop	tatop	snatop	fleagle	Unrelated
jique	nique	shique	plique	frique	wique	weck	Unrelated
klalky	phalky	flalky	kralky	pralky	galky	kation	Unrelated
kleagle	creagle	pleagle	greagle	preagle	sweagle	valky	Unrelated
priffy	triffy	siffy	bliffy	ziffy	griffy	hapte	Unrelated
saise	vaise	snaise	paise	jaise	taise	pazy	Unrelated

Note. Each row represents a list. In Experiment 3, half of the participants received Version 1 and the other half Version 2, with all lists from their assigned version. In Experiment 7, participants were tested only on the 20 related lists.

Appendix D

Visually related images names used in Experiment 4.

Version 1							
Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Lure	Condition
r_cat1_1	r_cat1_2	r_cat1_3	r_cat1_4	r_cat1_5	r_cat1_6	r_cat1_Lure	Related
r_cat2_1	r_cat2_2	r_cat2_3	r_cat2_4	r_cat2_5	r_cat2_6	r_cat2_Lure	Related
r_cat3_1	r_cat3_2	r_cat3_3	r_cat3_4	r_cat3_5	r_cat3_6	r_cat3_Lure	Related
r_cat4_1	r_cat4_2	r_cat4_3	r_cat4_4	r_cat4_5	r_cat4_6	r_cat4_Lure	Related
r_cat5_1	r_cat5_2	r_cat5_3	r_cat5_4	r_cat5_5	r_cat5_6	r_cat5_Lure	Related
r_cat11_1	r_cat11_2	r_cat11_3	r_cat11_4	r_cat11_5	r_cat11_6	r_cat11_Lure	Related
r_cat12_1	r_cat12_2	r_cat12_3	r_cat12_4	r_cat12_5	r_cat12_6	r_cat12_Lure	Related
r_cat13_1	r_cat13_2	r_cat13_3	r_cat13_4	r_cat13_5	r_cat13_6	r_cat13_Lure	Related
r_cat14_1	r_cat14_2	r_cat14_3	r_cat14_4	r_cat14_5	r_cat14_6	r_cat14_Lure	Related
r_cat15_1	r_cat15_2	r_cat15_3	r_cat15_4	r_cat15_5	r_cat15_6	r_cat15_Lure	Related
r_cat6_1	r_cat6_2	r_cat6_3	r_cat6_4	r_cat6_5	r_cat6_6	r_cat16_Lure	Unrelated
r_cat7_1	r_cat7_2	r_cat7_3	r_cat7_4	r_cat7_5	r_cat7_6	r_cat17_Lure	Unrelated
r_cat8_1	r_cat8_2	r_cat8_3	r_cat8_4	r_cat8_5	r_cat8_6	r_cat18_Lure	Unrelated
r_cat9_1	r_cat9_2	r_cat9_3	r_cat9_4	r_cat9_5	r_cat9_6	r_cat19_Lure	Unrelated
r_cat10_1	r_cat10_2	r_cat10_3	r_cat10_4	r_cat10_5	r_cat10_6	r_cat20_Lure	Unrelated
r_cat16_1	r_cat16_2	r_cat16_3	r_cat16_4	r_cat16_5	r_cat16_6	r_cat6_Lure	Unrelated
r_cat17_1	r_cat17_2	r_cat17_3	r_cat17_4	r_cat17_5	r_cat17_6	r_cat7_Lure	Unrelated
r_cat18_1	r_cat18_2	r_cat18_3	r_cat18_4	r_cat18_5	r_cat18_6	r_cat8_Lure	Unrelated
r_cat19_1	r_cat19_2	r_cat19_3	r_cat19_4	r_cat19_5	r_cat19_6	r_cat9_Lure	Unrelated
r_cat20_1	r_cat20_2	r_cat20_3	r_cat20_4	r_cat20_5	r_cat20_6	r_cat10_Lure	Unrelated
Version 2							
Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Lure	Condition
r_cat6_1	r_cat6_2	r_cat6_3	r_cat6_4	r_cat6_5	r_cat6_6	r_cat6_Lure	Related
r_cat7_1	r_cat7_2	r_cat7_3	r_cat7_4	r_cat7_5	r_cat7_6	r_cat7_Lure	Related
r_cat8_1	r_cat8_2	r_cat8_3	r_cat8_4	r_cat8_5	r_cat8_6	r_cat8_Lure	Related
r_cat9_1	r_cat9_2	r_cat9_3	r_cat9_4	r_cat9_5	r_cat9_6	r_cat9_Lure	Related
r_cat10_1	r_cat10_2	r_cat10_3	r_cat10_4	r_cat10_5	r_cat10_6	r_cat10_Lure	Related
r_cat16_1	r_cat16_2	r_cat16_3	r_cat16_4	r_cat16_5	r_cat16_6	r_cat16_Lure	Related
r_cat17_1	r_cat17_2	r_cat17_3	r_cat17_4	r_cat17_5	r_cat17_6	r_cat17_Lure	Related
r_cat18_1	r_cat18_2	r_cat18_3	r_cat18_4	r_cat18_5	r_cat18_6	r_cat18_Lure	Related
r_cat19_1	r_cat19_2	r_cat19_3	r_cat19_4	r_cat19_5	r_cat19_6	r_cat19_Lure	Related
r_cat20_1	r_cat20_2	r_cat20_3	r_cat20_4	r_cat20_5	r_cat20_6	r_cat20_Lure	Related
r_cat1_1	r_cat1_2	r_cat1_3	r_cat1_4	r_cat1_5	r_cat1_6	r_cat11_Lure	Unrelated
r_cat2_1	r_cat2_2	r_cat2_3	r_cat2_4	r_cat2_5	r_cat2_6	r_cat12_Lure	Unrelated
r_cat3_1	r_cat3_2	r_cat3_3	r_cat3_4	r_cat3_5	r_cat3_6	r_cat13_Lure	Unrelated
r_cat4_1	r_cat4_2	r_cat4_3	r_cat4_4	r_cat4_5	r_cat4_6	r_cat14_Lure	Unrelated
r_cat5_1	r_cat5_2	r_cat5_3	r_cat5_4	r_cat5_5	r_cat5_6	r_cat15_Lure	Unrelated
r_cat11_1	r_cat11_2	r_cat11_3	r_cat11_4	r_cat11_5	r_cat11_6	r_cat1_Lure	Unrelated
r_cat12_1	r_cat12_2	r_cat12_3	r_cat12_4	r_cat12_5	r_cat12_6	r_cat2_Lure	Unrelated
r_cat13_1	r_cat13_2	r_cat13_3	r_cat13_4	r_cat13_5	r_cat13_6	r_cat3_Lure	Unrelated
r_cat14_1	r_cat14_2	r_cat14_3	r_cat14_4	r_cat14_5	r_cat14_6	r_cat4_Lure	Unrelated
r_cat15_1	r_cat15_2	r_cat15_3	r_cat15_4	r_cat15_5	r_cat15_6	r_cat5_Lure	Unrelated

Note. Each row corresponds to a list and half of the participants were tested on version 1 and the other half on version 2. All participants were tested on all lists of their respective version. The images are available at the OSF page associated with this manuscript.