The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 2026, **81**(5), gbag038
https://doi.org/10.1093/geronb/gbag038
Published: 15 March 2026
Special Issue: Decision Neuroscience and Aging | Research Article

Multistep inference across the human lifespan can be improved with individualized memory interventions

Sharon M. Noh, PhD[1], Keiland W. Cooper, PhD[2], Shuheng Guo, BS[1], Dale Zhou, PhD[1,2], Craig E. L. Stark, PhD[2,3], Aaron M. Bornstein, PhD[1,3,*]

[1]Department of Cognitive Sciences, The University of California, Irvine, Irvine, California, United States
[2]Department of Neurobiology and Behavior, The University of California, Irvine, Irvine, California, United States
[3]Center for the Neurobiology of Learning and Memory, The University of California, Irvine, Irvine, California, United States
[*]Address correspondence to: Aaron M. Bornstein, PhD. E-mail: aaron.bornstein@uci.edu

Decision Editor: Karolina M. Lempert, PhD (Psychological Sciences Section)

Abstract

Objectives: **Effective goal-directed decision making relies on memory and planning–processes that are known to decline with age. We tested the hypothesis that these declines stem from a common mechanism by focusing on *mnemonic discrimination*, a measure of memory precision that shows unique vulnerability to age-related decline.**

Methods: **We used a latent learning task that measures the ability to learn and make judgments about associations among interconnected stimuli, assessing performance across the adult life span. This task allows us to measure multistep inference judgments that reflect how individuals organize relational structure, previously shown to capture the internal model-construction processes that support model-based planning. In Experiment 1, we examined relationships between judgment performance and memory precision. In Experiment 2, we tested whether a "blocked" learning schedule designed to reduce memory interference by separating overlapping objects could improve performance for individuals with weaker memory abilities.**

Results: **Across the life span, both younger and older adults showed evidence of successful latent learning and inference, but variability in judgment performance was explained by mnemonic discrimination ability. In Experiment 2, mnemonic discrimination interacted with training condition: intermixed training benefited those with high memory precision, whereas blocked training benefited those with low memory precision. We also implemented artificial neural network simulations, which reproduced these qualitative patterns.**

Discussion: **These findings suggest that age-related declines in complex judgments stem from declines in mnemonic discrimination and demonstrate that individualized, memory-based training interventions can improve learning and reasoning processes that support goal-directed planning, offering a promising approach to preserving decision-making abilities across the life span.**

Keywords: Individual differences, Latent learning, Memory precision, Training intervention, Goal-directed planning

Humans across the life span must often make decisions with lasting consequences for themselves and others. Increasingly, older adults occupy positions of power at the highest levels of government and major corporations, making their decision-making abilities particularly consequential for constituents and employees (Akhtar, 2019; Hall & Hickey, 2022; Ingraham, 2014; Schnoor, 2020). Yet aging is associated with declines in several cognitive abilities, including decision making (Craik & Salthouse, 2011; Hess et al., 2015; Raz, 2000). In particular, older adults show impairments in complex decision making requiring multistep associations and judgments, such as forward planning (Drummond & Niv, 2020; Eppinger & Bruckner, 2015; Eppinger et al., 2013). However, the mechanism by which multistep inference and planning decline with age is unclear, limiting efforts to develop effective interventions to reverse age-related cognitive decline. One challenge is that existing behavioral assays do not distinguish whether age-related declines arise from the judgment process itself or from difficulties learning the associations needed for such judgments, as in model-based planning (Feher da Silva & Hare, 2020; Konovalov & Krajbich, 2020).

Many real-world decisions require inferring structure from disparate experiences to form flexible, goal-directed plans. For instance, consider a student, Alice, preparing to present at an out-of-town conference. She books a 4 PM Friday flight and

Received: 15 August 2025. **Accepted:** 6 February 2026

plans to call a taxi at noon to allow ample time for travel. Even if this is her first conference, she can engage in multistep planning by assembling relevant knowledge from related experiences: after many delayed flights, she opts to arrive the night before her presentation (as opposed to the morning of); she anticipates heavier Friday afternoon traffic; and she chooses a taxi over driving due to past struggles finding airport parking.

Recent work in younger adults suggests that such goal-directed planning depends on inferring structure from the environment. Rmus et al. (2022) developed a task that measures how participants learn the associative structure among image pairs (edges) arranged along a latent "graph." Participants implicitly learned complex associative structures (*latent structure learning*) from randomized exposures to individual graph edges. This knowledge was assessed via shortest path judgments requiring multistep inference. Importantly, performance on the judgment task predicted greater use of model-based planning on a goal-directed decision-making task. A major strength of the task developed by Rmus et al. (2022) is its sensitivity to both the degree of latent structure learning and the ability to measure performance across varying associative distances. One open question, however, is *how* structured learning occurs, and what type of knowledge representations support multistep inference judgments.

One reasonable assumption is that memory may influence how individuals learn latent structures from experience, as model-based planning involves hippocampal contributions and effective memory search (Bornstein & Daw, 2013; Doll et al., 2015; Vikbladh et al., 2019). In the task by Rmus et al., an accurate representation of individual associations is critical for learning the graph. More specifically, memory *precision* may be especially important for forming representations that enable efficient planning. For example, remembering Friday afternoon traffic patterns (rather than Monday afternoon or Friday evening) allows Alice to make more accurate predictions for her trip. In episodic memory, high precision can be achieved through pattern separation, in which competing information is encoded as distinct neural patterns (Bakker et al., 2008; Yassa & Stark, 2011). Pattern separation reduces or resolves memory interference, enabling discrimination of similar events (Lacy et al., 2011; Poppenk et al., 2013), but this ability declines with age (Burke et al., 2010; Stark et al., 2013; Toner et al., 2009; Yassa & Stark, 2011). Declines in pattern separation may explain older adults' increased susceptibility to interference and related memory failures (Campbell et al., 2010; Wilson et al., 2006). We therefore asked whether age-related deficits in multistep planning stem from impaired latent structure learning due to reduced precision and greater interference.

If older adults' increased susceptibility to memory interference undermines their ability to learn the structures necessary for multistep planning, training interventions aimed at bolstering encoding may improve multistep inference judgments needed for complex decision-making tasks. In support of this idea, work on episodic memory has shown that manipulating learning sequences or temporally separating overlapping associations can reduce memory interference and bias formation of distinct neural representations to support associative inference (Zeithamova & Preston, 2017; Zhou et al., 2023).

What is the nature of the representations that support structural inference?

Research suggests that associative inference can be supported by at least two kinds of neural representations (Eichenbaum, 2001; Eichenbaum, 2017; Poppenk et al., 2013; Schlichting et al., 2015; Zhou et al., 2023). In standard associative inference tasks, participants learn A–B pairs, then B–C pairs, and are later tested on their knowledge of the indirect association between A and C, despite never having been presented with an A–C pair. If participants encode A–B and B–C as separate episodes during learning via pattern separation, they may form localist (orthogonalized) representations of each pair (Kumaran & McClelland, 2012; Zhou et al., 2023) and A–C inference requires effortful retrieval and recombination at test (e.g., "A paired with B, and B paired with C, so A is associated with C"). Although localist representations yield high precision and resist interference, they may be less efficient for making inferences due to the additional effort required for information retrieval and recombination. (Zeithamova et al., 2012; Zhou et al., 2023).

Alternatively, encountering B–C may reactivate A–B, allowing integration of new (-C) information into an updated A–B–C representation (Morton et al., 2017; Schlichting & Preston, 2015; Zeithamova et al., 2012). These distributed representations support rapid inference but are prone to interference and false memories (Bowman et al., 2021; McCloskey & Cohen, 1989; Zhou et al., 2023), making them potentially challenging for individuals such as older adults who are already vulnerable to interference. If age-related declines in memory precision cause inference failures, then a learning method that reduces memory interference should improve latent structure learning and judgments.

Current study

In the present study, we tested whether age-related declines in memory precision affect performance on a multistep associative inference task (the "graph" task; Rmus et al., 2022). Unlike standard associative inference tasks that use independent triads to test inference (e.g., AB–BC → AC; Carpenter et al., 2021; Schlichting et al., 2015; Zhou et al., 2023), our task requires participants to integrate many overlapping associations into a coherent cognitive graph and to estimate relative distances within that graph. These representational operations have been shown to predict the degree to which individuals rely on model-based planning strategies (Rmus et al., 2022; Yoo et al., 2024). In Experiment 1, participants studied randomly presented object pairs drawn from an underlying graph of 12 nodes (objects) and 16 edges (object pairs). At test, they judged relative distances between object pairs (considered to be a fundamental computation for high-level planning; Huang et al., 2019) assessing their ability to mentally navigate associations and make shortest-path judgments. Although the task does not involve overt goal-directed action or planning, the multistep inference judgments directly probe the internal model-construction processes that support goal-directed planning. Accordingly, we interpret performance on our inference judgment test as indexing the fidelity of internal relational models that underlie overt planning behavior. We expected performance deficits with age, specifically driven by poor memory precision, so participants also completed the Mnemonic Discrimination Task to independently measure *mnemonic discrimination*, a sensitive behavioral index of pattern separation ability that tracks age-related decline (Stark et al., 2019) and predicts decision making beyond chronological age (Noh et al., 2023). In Experiment 2, we tested the prediction that separating overlapping edge pairs during learning (blocked training) would reduce working memory load and improve multistep inference judgment performance for older adults or those with low memory precision (Schlichting et al., 2015). To confirm the relationship between training condition and memory precision, we simulated task performance using variants of artificial neural network (ANN) models that differed only in their internal representational capacity and compared the model outputs to our behavioral findings.

Method

Experiments 1 and 2 were identical except for the learning-phase presentation order (intermixed vs blocked). Data were collected concurrently, with participants randomly assigned to the intermixed condition (Experiment 1) or blocked condition (Experiment 2). Although the two training conditions were implemented under a single overarching protocol with random assignment of participants to conditions, we present the work as two experiments to reflect their distinct conceptual aims. Experiment 1's task closely followed Rmus et al. (2022) but included a life span sample with older adults, was administered online, and incorporated an additional measure of memory specificity. Thus, Experiment 1 uses the original intermixed schedule (Rmus et al., 2022) to replicate prior findings and to test whether age or mnemonic discrimination better explains multistep inference performance across the adult life span. Experiment 2 implements a temporally separated schedule to test the hypothesis that reducing interference would preferentially benefit low-precision learners. Anticipating that older adults might find the original task difficult, we implemented the blocked sequence in Experiment 2, based on prior work in associative memory suggesting that temporally separating overlapping associations might reduce cognitive load and improve performance in populations with lower memory capacity (Schlichting et al., 2015). Note that the "blocked" manipulation used here (i.e., separating overlapping edges across time over the course of training) is consistent with similar manipulations of "blocking" used in episodic memory (Beukers et al., 2024; Schlichting et al., 2015; Zhou et al., 2023) and is somewhat distinct from how "blocking" is used in the category learning literature (Brunmair & Richter, 2019; Kang & Pashler, 2012; Noh et al., 2024, 2016). Methods are described jointly below, with procedural differences between the two experiments noted explicitly.

Participants

Sample size was based on prior work with similar designs ($N = 81$; Rmus et al., 2022), with a ∼25% increase to offset higher noise in our online data (relative to the supervised in-person data collected by Rmus et al., 2022). We aimed for ∼100 participants per experiment post-exclusion. A total of 219 participants (112 women, 107 men; ages 19–84, mean[*SD*] age = 55.7[14.2]) were recruited online (see online supplementary material; Online Data Collection): 113 in Experiment 1 (59 women, 54 men, ages 22–84, mean[*SD*] age = 56.7[13.8]), and 106 in Experiment 2 (53 women/53 men, ages 19–79, mean[*SD*] age = 54.8[14.7]). Participants received monetary compensation and completed a tutorial plus a rotation-detection screener (10 practice trials; ≥70% accuracy required within two attempts to ensure attention to task instructions) to be eligible to participate in the full task. All protocols were classified as an "Exempt Online Survey" by the Institutional Review Board of the University of California, Irvine.

Procedure

Mnemonic discrimination task

Participants completed the MST (Figure 1) as an independent measure of memory ability to assess whether mnemonic discrimination explained individual differences in graph task performance.

Encoding phase

Participants viewed object images and made indoor/outdoor judgments (cover task). No MST images overlapped with those used in the graph task, and participants were unaware that their memory would later be tested.

Mnemonic discrimination test

On a surprise discrimination test, participants saw object images that were identical (old), similar (lures), or novel (foils) relative to objects shown during encoding, and judged each object as "old," "similar," or "new." A Lure Discrimination Index (LDI) was calculated for each participant: p('sim'|lure) - p('sim'|foil). Higher LDI values indicate better ability to classify lures as "similar" relative to foils, reflecting greater mnemonic discrimination and better memory encoding precision (Stark et al., 2019).

Structural inference "graph" task

Study phase

Participants viewed and were told to memorize a series of object pairs (Figure 2A), each presented for 1 s in random order. Participants provided a rotation judgment on each trial to ensure attention. No information about the underlying structure was given, though pairs were drawn from a hidden 12-node, 16-edge graph (Figure 2B). In Experiment 1, all 16 unique edge pairs were repeated 44 times (704 trials total) in a random sequence (Figure 2C, "intermixed"). Experiment 2 grouped object pairs into 4 mini-blocks, each with 4 unique, non-overlapping pairs (Figure 2C, "blocked"). This separation of potentially confusing (overlapping) edges across time aimed to reduce memory interference and improve learning of the graph structure.

Following the study phase, participants completed two different tests to measure how well participants can *use* their knowledge of the graph structure (Figure 3A, *Judgment Test*), as well as how well they *learned* the graph structure (Figure 3B, *Graph Reconstruction Test)*.

Judgment test

After completing the learning phase, participants completed a relative distance judgment task. This task measured participants' ability to use learned information for structural inference. Participants were asked to judge which of two objects (left vs

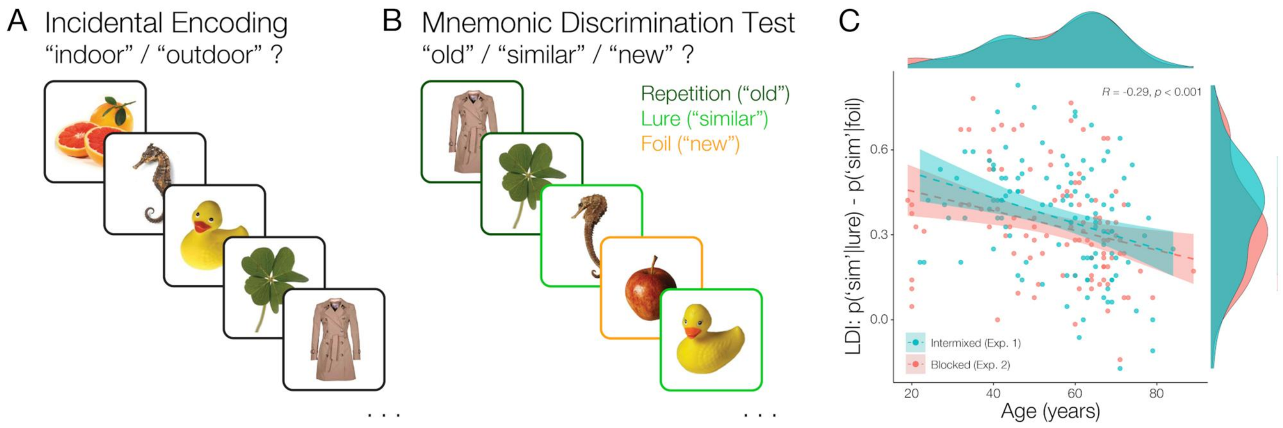

Figure 1 The mnemonic discrimination task. (A) Participants view a sequence of objects during an incidental encoding phase in which participants are asked to classify each object as an indoor or outdoor object. (B) In a surprise discrimination test, participants view a series of objects and are asked to determine whether each object is an "old," "new," or "similar" item relative to what was shown during the encoding phase. (C) Relationship between chronological age and LDI across participants in both experiments. Shaded bands indicate 95% confidence intervals around the best fit regression line for each experiment (dashed lines). LDI = Lure Discrimination Index.

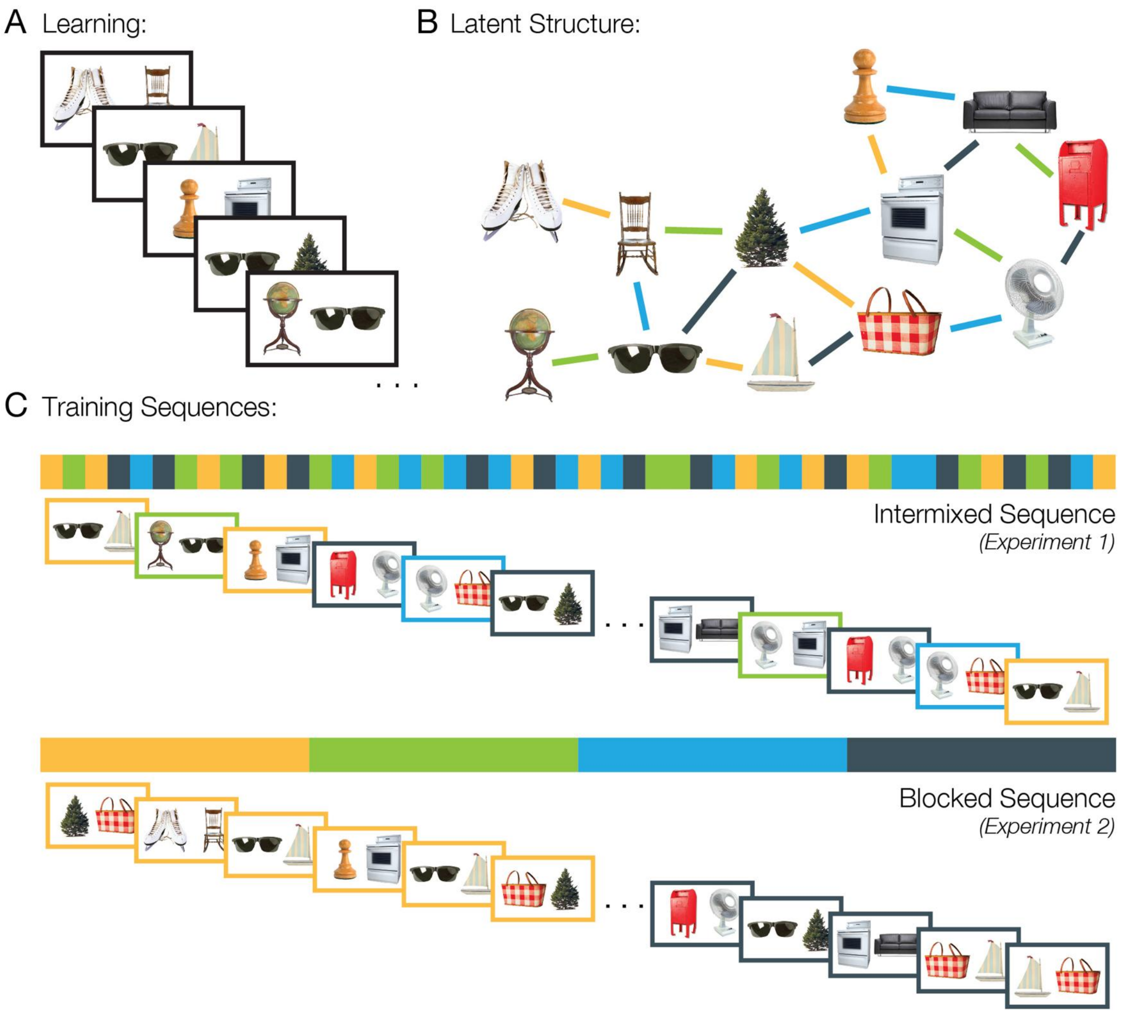

Figure 2 Structural inference task learning phase. Participants learn individual edges (A) drawn from a latent structure made up of several overlapping edges (B). In Experiment 1, edges are drawn randomly and presented in an intermixed fashion (C, intermixed), whereas in Experiment 2, overlapping edges are separated in time across different mini-blocks (C, blocked).

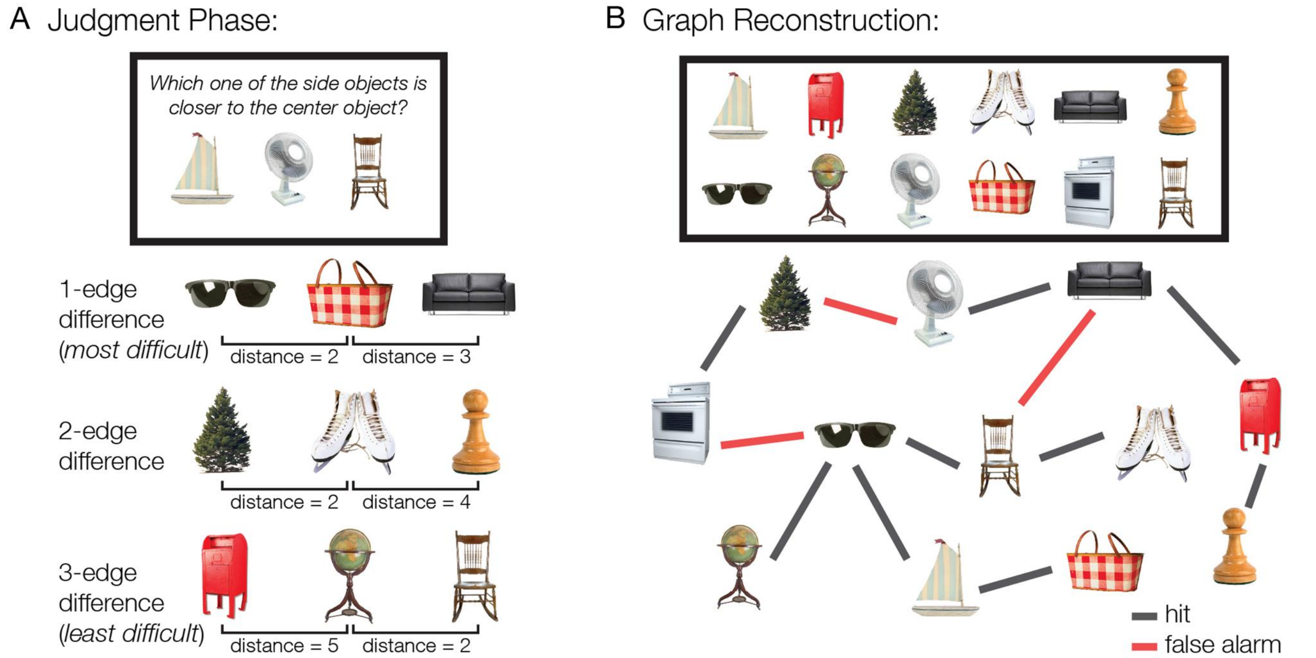

Figure 3 Structural inference task test phases. (A) After the learning phase, structural inference-based judgments were assessed for each participant. Participants were presented with three objects and asked whether the object on the left or right was closer to the center object, based on the associations they learned in the previous study phase. Judgment phase trials varied in difficulty based on the difference between choice options. The most difficult trials were ones in which options differed by an associative distance of 1 (distance 2 vs 3, 3 vs 4, or 4 vs 5), whereas the easiest trials were ones in which choice options differed by an associative distance of 3 (distance 2 vs 5). (B) After the judgment phase, participants were asked to reconstruct the graph to the best of their knowledge by placing all studied objects on a "canvas" on their screen and connecting objects only if they had been directly paired together during the study phase. Correctly drawn connections were classified as "hits," whereas incorrectly drawn connections were classified as "false alarms." Participants were required to place all objects on the canvas, and each object had to be connected to at least one other object to complete this phase.

right) was closer to a central object, based on indirect relationships learned previously (i.e., graph distance). Responses were made using keyboard buttons, across 204 trials with up to 10 s per trial. Each unique object served as a central node 17 times. Choice options were randomly selected with the following constraints: (1) neither option was directly paired with the central node during the study, and (2) the shortest path length between the central node and the two options (associative distance) was not equal. Trial difficulty varied based on the difference in associative distance between choice options and the reference node (ranging from 1 to 3). Accuracy was calculated within each difficulty bin.

Because online data are unsupervised and noisier, judgment test data were screened for outliers prior to analysis using a two-step reaction time (RT)–based procedure, as RTs were not analyzed directly (see online supplementary material; *RT-based Filters* for additional details). Filtered judgment-phase data were analyzed using linear mixed-effects models (R nlme package; Pinheiro et al. 2017, with difficulty [distDiff = 1, 2, 3; 1= most difficult]), age, LDI, and sequence (blocked vs intermixed) specified as fixed effects, and participant as a random effect.

Graph reconstruction test

Following the judgment test, participants completed a graph reconstruction task to assess explicit knowledge of the learned structure. All 12 study-phase objects were displayed above a blank canvas, and participants arranged them by clicking and dragging each object onto the canvas. Participants were asked to place and link objects that were directly paired during the study by selecting two objects sequentially; linked items were connected with a straight line. All objects had to be placed on the canvas and linked to at least one other object before submission.

We used a Signal Detection Theory approach to analyze the reconstruction data (Carterette, 1967). We scored reconstructions as follows: "hits" for correctly linked study pairs, "false alarms" for incorrectly linked pairs, and the total number of edges drawn. Accuracy was computed in two ways: (1) sensitivity (d'), defined as "hit rate—false alarm rate," and (2) precision. Across the 12 nodes, there are a total of 66 possible edges that could be drawn if every object pair were connected. Of these, 16 edges are true edges in the underlying graph structure "hits", and 50 represent non-edges "false alarms". Thus, the hit rate is calculated as hits/16, and the false-alarm rate is calculated as false alarms/50, and d′ is computed as hit rate minus false-alarm rate. We also examined an additional measure of accuracy (precision), taken from machine learning (Blair, 1979), to complement d′. Although d′ separately considers hits and false alarms relative to the total possible true and false edges, precision instead evaluates the proportion of drawn edges that were correct (hits/total edges drawn). Precision is useful for comparing reconstruction accuracy across individuals who may differ in memory capacity. For example, an individual with high memory capacity may draw more edges overall, whereas a lower-capacity individual may draw fewer; precision helps interpret reconstruction accuracy in light of these differences. Thus, the precision metric adjusts performance based on the number of edges drawn, allowing for better comparisons across individuals with varying memory capacity.

Analyses were conducted in R using linear regression (lm). We separately modeled hits, false alarms, total edges, accuracy, and precision as a function of LDI, training sequence (blocked vs intermixed), and their interaction. For each analysis, the full model was specified as follows: DV $\sim$ LDI + Sequence + LDI×Sequence. If no significant interaction was found, the model was refitted without the interaction term to test for main effects (Grace-Martin, 2011).

Artificial neural network model

We implemented a set of simple five-layer feedforward autoencoders to simulate associative learning of the graph structure. The models learned to predict each object's corresponding pair (as presented during study) and were then used to infer representational structure within their hidden layers. As with our human participants, half of the models were trained via the intermixed sequence, and the other half were trained via the blocked sequence. Across models, we also varied the number of units in the hidden layers to simulate differences in memory capacity (fewer units = lower memory capacity). This approach allowed us to test whether representational capacity could qualitatively reproduce the observed human interaction between memory precision and learning sequence. Full model specification, optimization parameters, and training details are provided in Supplementary Material.

Results

Experiment 1: Better memory encoding ability improves structural inference-based judgments

Judgment test

We used a linear mixed-effects model to examine whether trial difficulty (relative distance difference between choice options) and chronological age predicted judgment accuracy (Figure 4): judgment accuracy $\sim$ age + distDiff + age $\times$ distDiff, random = $\sim$1|subject (adjusted R^2 = 0.23, AIC = −152.55). There was a main effect of age (β_{age} = −0.030, SE = 0.012, $t(110)$ = −2.57, p = .011), with older adults performing worse than younger adults. There was also a main effect of difficulty (β_{distDiff} = 0.032, SE = 0.009, $t(216)$ = 3.50, $p < .001$), with performance improving as relative distance increased. The age $\times$ difficulty interaction was significant ($\beta_{\text{age}\times\text{distDiff}}$ = −0.022, SE = 0.009, $t(216)$ = −2.41, p = .017) such that younger adults showed greater performance gains as trials got easier (i.e., greater relative distance between options).

Adding mnemonic discrimination ability (LDI) to the model revealed that it was a stronger predictor of judgment accuracy than age: judgment accuracy $\sim$ age + distDiff + LDI + age $\times$ distDiff + age $\times$ LDI + distDiff $\times$ LDI + age $\times$ distDiff $\times$ LDI, random = $\sim$1|subject (adjusted R^2 = 0.26, AIC = −133.98). When both age and LDI were included, the main effect of age (β_{age} = −0.021, SE = 0.012, $t(108)$ = −1.72, p = .088) and its interaction with difficulty ($\beta_{\text{age}\times\text{distDiff}}$ = −0.012, SE = 0.010, $t(214)$ = −1.24, p = .218) were no longer significant. Instead, higher LDI predicted better performance (β_{LDI} = 0.031, SE = 0.013, $t(108)$ = 2.42, p = .017), with a significant LDI $\times$ difficulty interaction ($\beta_{\text{LDI}\times\text{distDiff}}$ = 0.033, SE = 0.010, $t(214)$ = 3.34, p = .001) indicating that memory precision benefits emerged primarily in easier trials. Age $\times$ LDI and the three-way interaction were non-significant. Given LDI's predictive strength, we re-ran the model excluding age (Figure 5, intermixed condition; adjusted R^2 = 0.26, AIC = −164.19). This model confirmed significant main effects of LDI (β_{LDI} = 0.035, SE = 0.012, $t(110)$ = 2.99, p = .003) and difficulty (β_{distDiff} = 0.032, SE = 0.009, $t(216)$ = 3.50, $p < .001$) as well as a robust LDI $\times$ difficulty interaction ($\beta_{\text{LDI}\times\text{distDiff}}$ = 0.034, SE = 0.009, $t(216)$ = 3.84, $p < .001$), with higher LDI predicting greater gains in easier trials.

These results suggest that structural inference performance depends on the precision of encoded associations, with the largest differences observed in trials requiring comparisons across larger associative distances (i.e., easier trials). Notably, participants with low memory abilities performed at chance across all difficulty levels, suggesting minimal or no latent structure learning. We hypothesized that this may reflect greater susceptibility to memory interference, which might be mitigated by training on non-overlapping subsets one block at a time. Thus, Experiment 2 introduced the "blocked" training intervention to reduce interference for low-LDI participants during learning and potentially improve subsequent judgment performance.

Experiment 2: Memory-based inference can be improved via individualized training

To test whether training conditions modulate structural inference as a function of memory ability, we combined data from Experiment 1 (n = 113, intermixed sequence) and Experiment 2 (n = 106, blocked sequence) and fit a linear mixed-effects model (Figure 5): judgment accuracy $\sim$ LDI + distDiff + Sequence + LDI $\times$ distDiff + Sequence $\times$ LDI + Sequence $\times$ distDiff + LDI $\times$ Sequence $\times$ distDiff, random = $\sim$1|subject (adjusted R^2 = 0.21, AIC = −304.55). There was a significant 3-way interaction between mnemonic discrimination ability, difficulty, and learning sequence ($\beta_{\text{LDI}\times\text{distDiff}\times\text{sequence}}$ = 0.054, SE = 0.013, $t(422)$ = 4.10, $p < .001$), indicating that the relationship between mnemonic discrimination ability (LDI) and performance differed by training sequence and trial difficulty. Importantly, there was an interaction between learning sequence and mnemonic discrimination ability ($\beta_{\text{LDI}\times\text{sequence}}$ = 0.038, SE = 0.016, $t(214)$ = 2.26, p = .025): participants with low LDI benefitted from blocked training, whereas those with high LDI performed better with intermixed training. This crossover interaction was evident only in the easiest trials (*distDiff* = 3). Still, reducing memory interference via blocked training improved structural inference for individuals with weaker memory ability, enabling more accurate inference judgments when associative distances were large.

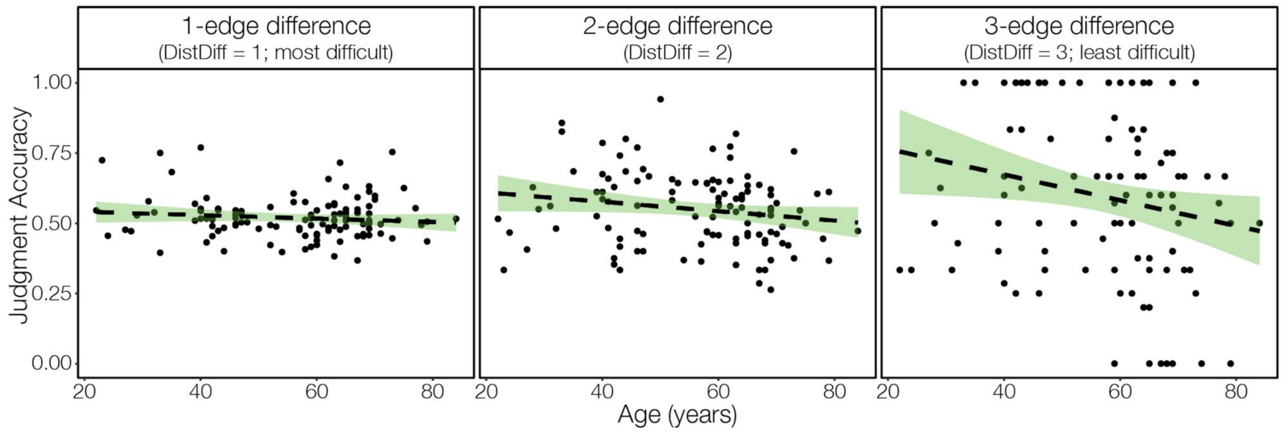

Figure 4 Behavioral results from experiment 1. Judgment accuracy as a function of chronological age and trial difficulty (distance difference between choice options). The green band indicates 95% confidence intervals around the best fit regression line.

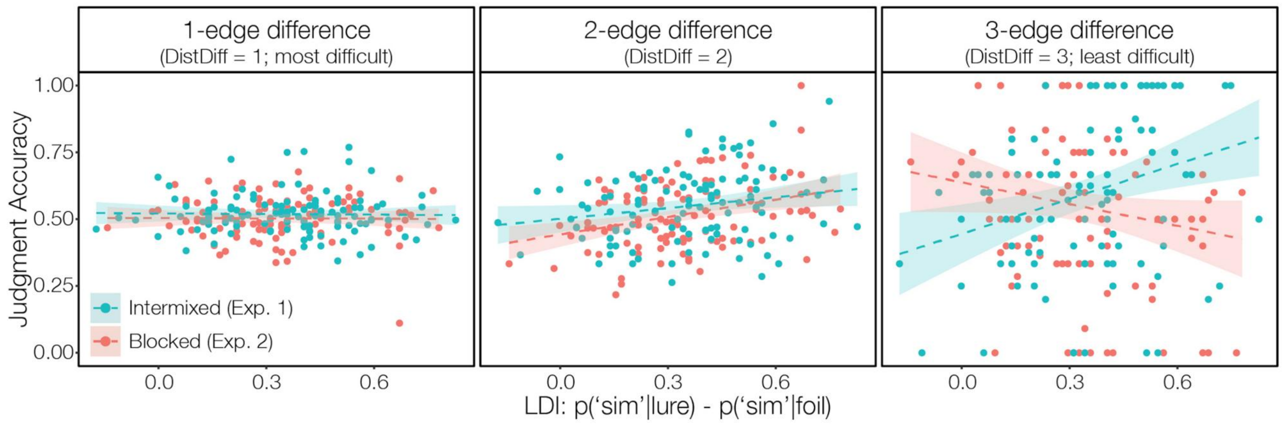

Figure 5 Behavioral results combining data from Experiment 1 and Experiment 2. Judgment accuracy as a function of trial difficulty (distance difference between choice options), mnemonic discrimination ability (LDI), and training condition (blocked vs intermixed learning sequence). Shaded bands indicate 95% confidence intervals around the best fit regression line for each training condition (dashed lines). LDI = Lure Discrimination Index.

ANN captures relationship between training sequence and memory capacity in distance-dependent judgment performance

Given the observed interaction between training sequence and mnemonic discrimination ability (LDI) in Experiments 1 and 2, we tested whether such effects could be reproduced by manipulating a neural network's representational capacity (Figure 6A). To test this, we implemented a five-layer feedforward autoencoder, which jointly optimizes encoding and decoding of sequential inputs (Blanco Malerba et al., 2024; Chandak et al., 2024), trained to associate paired items (edges) from the graph task (704 presentations; see online supplementary material: *Artificial Neural Network Model*).

To model variability in memory precision, we manipulated the number of units in the second and fourth layers (L2/L4; Figure 6A). Larger widths corresponded to more distinct, nonoverlapping representations of each node (analogous to higher LDI), whereas smaller widths simulated representational merging before latent structure extraction (analogous to lower LDI). Consistent with our behavioral results, there was a three-way interaction between layer width, difficulty, and training sequence ($\beta_{\text{width}\times\text{distDiff}\times\text{sequence}}$ = 0.0032, $SE = 0.0006$, $t(23996) = 5.58$, $p < .001$) (Figure 6B). Additionally, there was an interaction between training sequence and the models' layer width ($\beta_{\text{width}\times\text{sequence}}$ = 0.0147, $SE = 0.0008$, $t(11996) = 18.37$, $p < .001$) such that structural inference-based judgments were optimized by different training conditions (blocked vs intermixed learning sequences) depending on the neural network model's capacity (larger vs smaller hidden layer width) in a similar pattern to those seen in the human data (blocked > intermixed for low LDI, intermixed > blocked for high LDI). These interactions show that lower-capacity models benefited more from blocked training, whereas higher-capacity models benefited more from intermixed training – precisely the crossover pattern observed between memory precision and sequence in our human participants.

Graph reconstruction

To get a sense of the memory representations that participants may have formed during learning, we analyzed hits (correct edges), false alarms (incorrect edges), total edges drawn, and two accuracy metrics (sensitivity and precision) from the graph

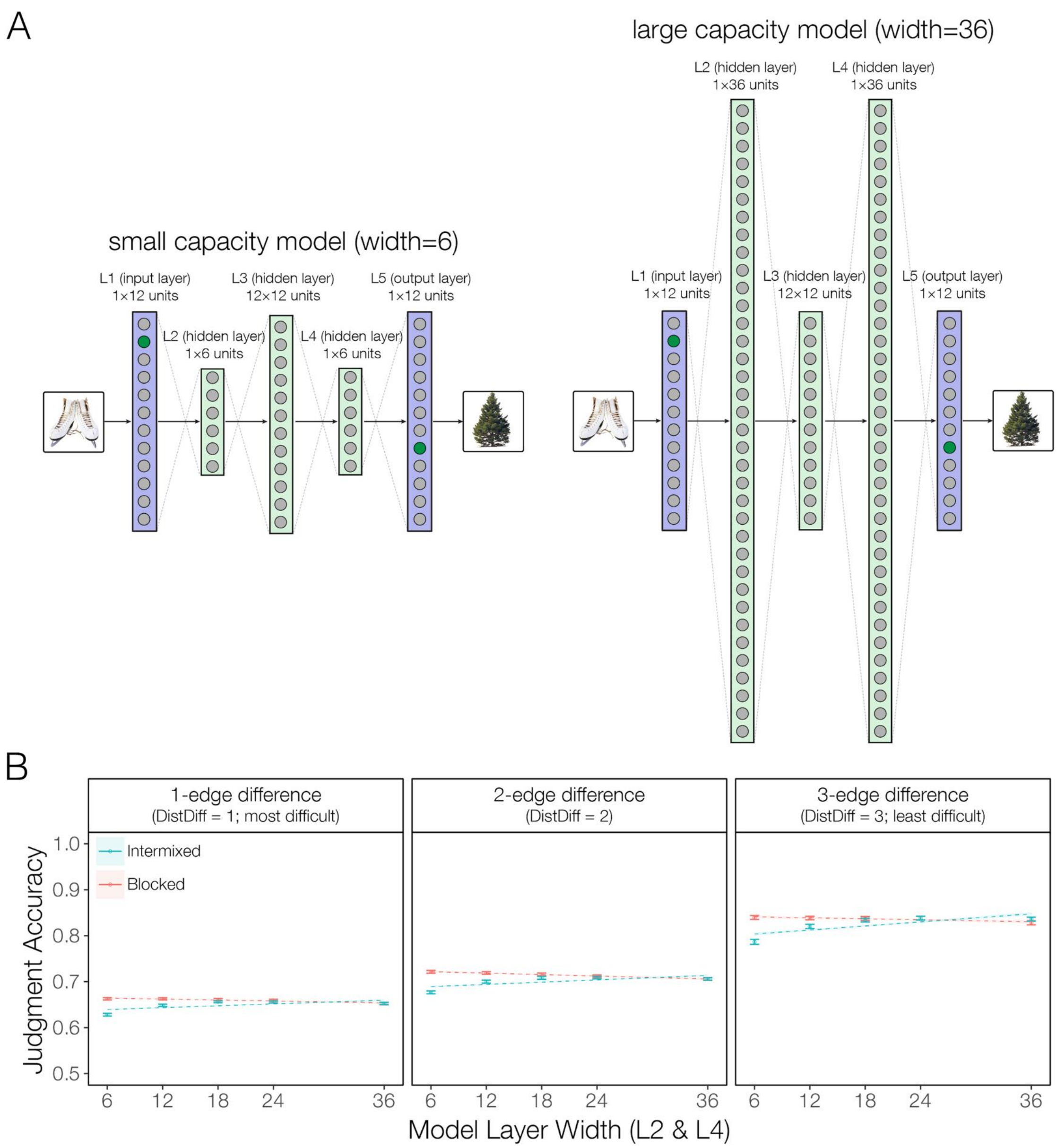

Figure 6 Neural network design and simulation results. (A) Schematic of two example artificial neural networks (the smallest and largest capacity models). To model memory capacities' effect on judgment accuracy, the width of the second and fourth layers was varied from 6 units (low-LDI; reflecting the hypothesized diminished capacity for separating inputs) to 36 units (high-LDI; allowing for sparse, highly separated representations of each input object). (B) Model judgment task results across varying layer widths and training conditions, matching the pattern observed in human participants with varying memory precision. Shaded bands indicate 95% confidence intervals around the best fit regression line for each training condition (dashed lines), and error bars around individual points reflect 95% confidence intervals around each mean. LDI = Lure Discrimination Index.

reconstruction phase as a function of mnemonic discrimination ability (LDI) and training sequence (blocked vs intermixed). Each of these five reconstruction metrics (hits, false alarms, total edges, sensitivity, and precision) was used as the outcome measure with LDI and sequence as independent predictors (outcome ∼ LDI + Sequence + LDI×Sequence). Across all five regressions, the LDI × Sequence interaction was not significant, so interaction terms were removed and models were refit to assess main effects; the results reported here are from those main-effects models. For hits, the linear regression (*hits ∼ LDI + Sequence*) was significant, adjusted R^2 = 0.03, $F(2, 212)$ = 4.09, p = .018, with a main effect of LDI, β = 0.65, SE = 0.23, $t(212)$ = 2.81, p = .005, indicating that participants with better memory ability produced more correct edges; there was no effect of training sequence (Figure 7A). For false alarms, the model (*false alarms ∼ LDI + Sequence*) was significant, adjusted R^2 = 0.03, $F(2, 212)$ = 4.07, p = .018, with a main effect of training sequence, β = 1.53, SE = 0.55, $t(212)$ = 2.77, p = .006, such that participants in the intermixed condition produced more false edges; LDI was not significant (Figure 7B). For total edges, the model (*total edges ∼ LDI + Sequence*) was significant, adjusted R^2 = 0.05, $F(2, 212)$ = 6.38, p = .002, with main effects of LDI, β = 0.79, SE = 0.33, $t(212)$ = 2.36, p = .019, and Sequence, β = 1.71, SE = 0.67, $t(212)$ = 2.55, p = .011, indicating that higher-LDI participants and those in the

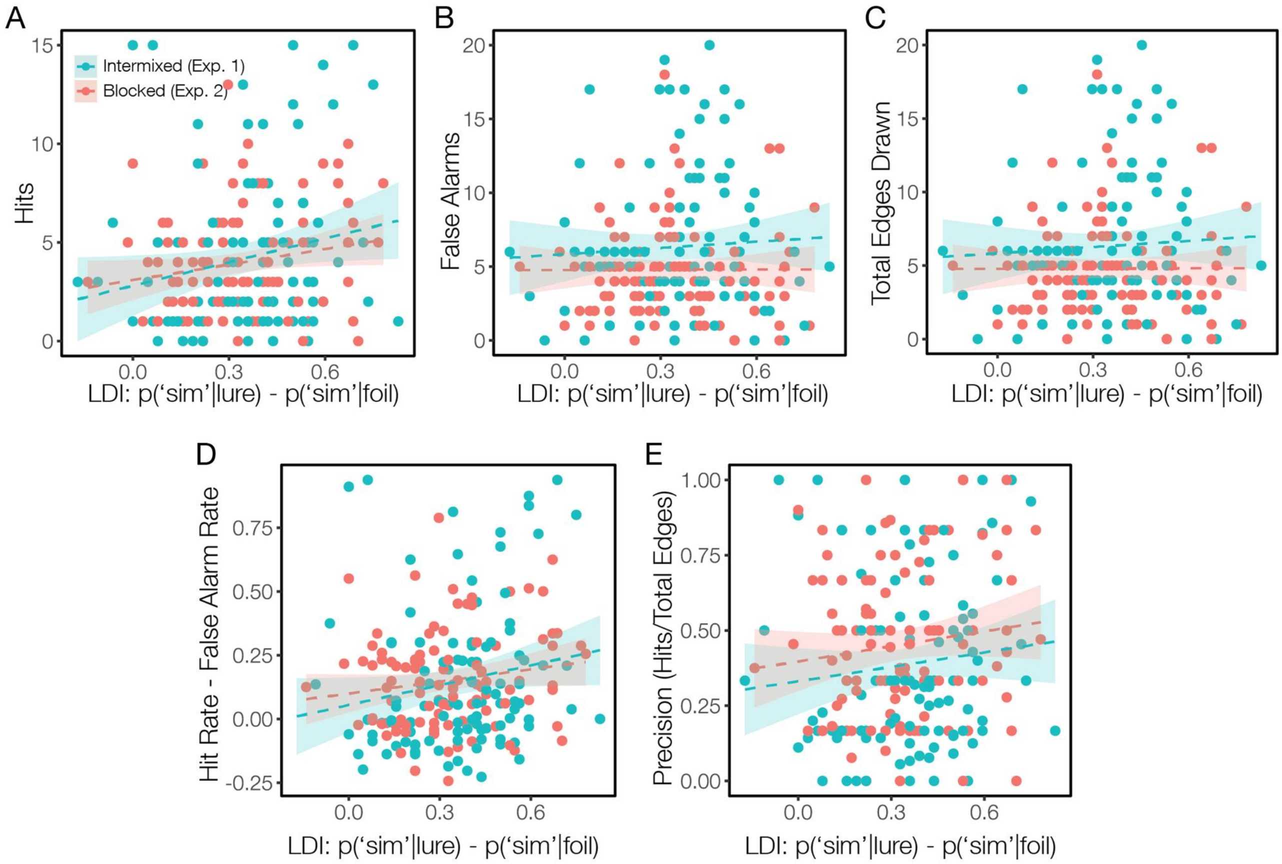

Figure 7 Behavioral results combining data from experiment 1 (blue) and experiment 2 (red) for the graph reconstruction phase. Shaded bands indicate 95% confidence intervals around the best fit regression line for each training condition (dashed lines). LDI = Lure Discrimination Index.

intermixed condition drew more edges overall (Figure 7C). We also assessed reconstruction accuracy using two metrics: sensitivity (proportion of hits—proportion of false alarms) and precision (number of hits divided by the total number of drawn edges). For sensitivity, the model (*sensitivity* $\sim$ *LDI* + *Sequence*) was significant, adjusted R^2 = .02, $F(2, 212) = 3.12$, $p = .046$, revealing a main effect of LDI, $\beta = 0.04$, $SE = 0.01$, $t(212) = 2.49$, $p = .014$, with higher-LDI participants showing better reconstruction accuracy (Figure 7D). For precision, which adjusts for the tendency of low-memory participants to draw fewer edges, the model (*precision* $\sim$ *LDI* + *Sequence*) was marginal, adjusted R^2 = .02, $F(2, 212) = 3.03$, $p = .050$, with trending effects of LDI, $\beta = 0.04$, $SE = 0.02$, $t(212) = 1.72$, $p = .088$, and Sequence, $\beta = -0.09$, $SE = 0.05$, $t(212) = -1.86$, $p = .065$, suggesting higher precision for participants with better memory ability and for those trained in the blocked condition (Figure 7E).

Discussion

The present study examined whether multistep inference judgments requiring latent structure learning are affected by age-related cognitive decline. In Experiment 1, we found that older adults performed worse on structural inference judgments, but these deficits were better explained by individual differences in mnemonic discrimination ability (LDI) than by chronological age. Experiment 2 demonstrated that structural inference can be improved through individualized training: intermixing overlapping associations benefited those with high LDI, while blocked training benefited those with low LDI. These behavioral findings were reinforced by ANN model simulations that varied in memory capacity to examine multistep inference performance under blocked and interleaved training conditions. Although the models are a simplified system, manipulating representational capacity in these networks successfully reproduced the sequence × memory precision interaction observed in human participants, thereby providing converging evidence for our hypotheses.

Collectively, our findings suggest that the way individuals encode and organize information interacts with memory ability to shape representations that vary in their effectiveness for supporting accurate inference. Matching training to an individual's representational tendencies may be critical for optimizing judgment performance. Although our task did not require participants to execute planned sequences of actions, the multistep inference judgments it elicited engage the representational mechanisms thought to support model-based planning (Rmus et al., 2022; Yoo et al., 2024). Specifically, the judgment test requires estimating which of two stimuli is "closer" within a learned associative graph, effectively querying the internal relational model used to chain together multiple steps of the structure. Prior work using similar graph-learning tasks has demonstrated that the integrity of these relational models predicts individuals' reliance on model-based strategies (Rmus et al., 2022). Thus, our findings offer insight into how age- and memory-related variability in representational fidelity may constrain the computations that underlie goal-directed planning.

In Experiment 1, evidence of latent structure learning emerged only in participants with high mnemonic discrimination ability. High-LDI individuals achieved above-chance inference performance, especially on easier trials, whereas low-LDI participants performed at chance. The lack of learning observed in low-LDI participants who performed our task may explain why certain populations such as older adults and individuals with mild cognitive impairments may exhibit deficits in goal-directed planning (Eppinger & Bruckner, 2015; Samanez-Larkin, 2013), as prior work shows that these populations are particularly vulnerable to deficits in memory precision (Yassa & Stark, 2011; Yassa et al., 2010). High-LDI participants also produced more accurate graph reconstructions across multiple metrics (number of correct edges, reconstruction accuracy measured using sensitivity and precision), but they also generated more false alarms and total edges. One explanation is that making inference judgments before reconstruction induced false memories by misattributing inferred information as directly experienced (Bowman et al., 2021). However, this order effect should have applied equally to all participants, making it unlikely to explain the selective increase in false alarms for high-LDI individuals. A more plausible account is that high-LDI individuals tend to form distributed representations during learning (Kumaran & McClelland, 2012; Zhou et al., 2023). According to parallel distributed processing models such as C-HORSE, distributed representations support flexible inference but are prone to false alarms (Zhou et al., 2023). Our data fit this prediction: high-LDI participants outperformed others on inference judgments but also produced more false edges.

In Experiment 2, we tested whether separating overlapping pairs in time (blocked training) would reduce memory interference and improve inference for low-LDI individuals. Indeed, blocked training improved structural inference for low-LDI participants but impaired performance for high-LDI participants, producing a crossover interaction. Visual inspection of graph reconstructions of the best performers in each condition (Supplementary Figure 1) suggests why: high-LDI participants in the intermixed condition tended to create highly interconnected, distributed graphs, whereas blocked training led them to form more fragmented, localist graphs. In contrast, low-LDI participants generally produced localized graphs, but blocked training encouraged more interconnections relative to intermixing. These patterns suggest that memory encoding ability may bias representational style—distributed for high-LDI, localist for low-LDI—and that training sequences aligning with these preferences yield better performance. This hints at the idea that blocked vs interleaved learning sequences can be leveraged to improve multistep inference and other forms of inferential reasoning as a function of one's memory capacity.

This framework may help reconcile conflicting findings in associative memory research. Some studies report that intermixing promotes distributed representations (Zhou et al., 2023), whereas others find it encourages more localist coding (Schlichting et al., 2015) with the reverse pattern for blocked learning. A key methodological difference is that Zhou et al.'s "blocked" condition still included intermixed trials within the same learning phase, likely negating interference-reduction benefits. This design may have disproportionately favored high-LDI participants while preventing low-LDI learners from reaching performance levels sufficient to influence overall effects. Supporting this, Zhou et al. excluded many low performers, yet performance remained low relative to Schlichting et al.

Recent neural network modeling shows that distributed representations are more likely to emerge when information is presented in an intermixed sequence (Zhou et al., 2023). Models with distributed coding predict that blocked learning increases memory interference as new information is acquired (McCloskey & Cohen, 1989; Zhou et al., 2023). Our behavioral data support this: the high-LDI intermixed group showed the best inference performance but also the highest false alarms during reconstruction—consistent with distributed coding—although the high-LDI blocked group exhibited low false alarms (Figure 7B) and high graph reconstruction precision (Figure 7E), suggesting more localist coding. However, this gain in precision came at the cost of impaired judgment performance (Figure 5). This tradeoff parallels evidence that temporally contiguous presentation facilitates integrative encoding (Pudhiyidath et al., 2022; Schapiro et al., 2013; Zeithamova & Preston, 2017; Zhou et al., 2023), whereas separating overlapping information can disrupt inference if it requires linking across longer temporal gaps (Zeithamova & Preston, 2017; Zhou et al., 2023). Visual inspection supports this: high-LDI blocked learners' graphs were less cohesive than those in the high-LDI intermixed group (Supplementary Figure 1), consistent with blocking fragmenting an otherwise distributed network into more localized subgraphs, making cross-representation inference harder.

For low-LDI learners—such as older adults with age-related decline—susceptibility to memory interference can severely impair learning and memory encoding in an intermixed schedule. With 16 overlapping pairs, interference may be so severe that learners can only focus on memorizing a small subset of pairs—similar to older adults' strategy of selectively encoding high-value items (Castel, 2007; Castel et al., 2002). Indeed, the top performers in the low-LDI intermixed condition reconstructed only the minimum six edges required to complete the task, suggesting encoding of a few localized subsets. Since our graph task likely benefits from distributed representations, the main challenge for low-LDI learners is overcoming interference during encoding (Kirwan & Stark, 2007; Pettigrew & Martin, 2014; Shimamura & Jurica, 1994). Blocked training may mitigate this by spacing overlapping pairs across time, allowing learners to consolidate subsets (e.g., A–B) before encountering related pairs (e.g., B–C). Previous studies in episodic memory have also suggested that stabilizing A–B representations prior to introducing overlapping B–C episodes can prevent memory interference and promote integration of the related episodes (Noh et al., 2026; Schlichting et al., 2015). Once initial representations are stabilized through repetition, new overlapping pairs can be integrated into existing knowledge (A–B–C) by pattern completing the overlapping memory traces (Schlichting et al., 2015). Consistent with this account, low-LDI participants in the blocked condition showed better inference judgments (Figure 5) and fewer false alarms (Figure 7B), indicating reduced interference. This pattern was also replicated with the ANN model simulations, showing that smaller-capacity models performed better under blocked (vs intermixed) training conditions. Visual reconstructions suggest blocked training promoted more integrated networks in low-LDI learners than intermixing, consistent with prior work showing that blocking overlapping pairs fosters integrated representations, while intermixing promotes pattern-separated representations (Schlichting et al., 2015).

Collectively, these results indicate that structural inference depends on both memory encoding ability and the alignment of training structure with representational tendencies. For high-LDI learners, intermixing supports distributed networks that facilitate inference; for low-LDI learners, blocking reduces interference and fosters integration. These findings suggest that some age-related deficits in model-based planning and decision making may stem from failures to form adequate latent structures. Because multistep inference over a cognitive graph provides a substrate for evaluating relational distances and future action paths, these findings have direct implications for models of goal-directed decision making. The present task therefore contributes to understanding how aging- and memory-related representational constraints shape the cognitive computations underlying planning, even in the absence of overt behavioral choice. Future work should test how disruptions in latent structure learning impair multistep planning (Harhen & Bornstein, 2023; Yoo et al., 2024), directly measure memory for individual associations, and use neuroimaging to track the formation of representations during learning. Given that successful inference can emerge from different representational formats, future studies should also examine their flexibility and limitations across different inference and planning contexts. Despite limitations, our results offer a mechanistic basis for individualized learning interventions to mitigate cognitive decline effects on decision making across the life span.

Supplementary material

Supplementary material is available at *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* online.

Funding

This work was supported by the National Institute of Aging at the National Institutes of Health (R21AG072673 and R01AG088306 to A.M.B., R01AG066683 and P30AG066519 to C.E. L.S., and F32AG072836 to S.M.N.); and the National Science Foundation (NSF DGE-1839285 to K.W.C.).

Conflicts of interest

None declared.

Author contributions

Sharon M. Noh and Keiland W. Cooper are co-first authors. Sharon M. Noh: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing–original draft, writing–review & editing, visualization, supervision, project administration, funding acquisition. Keiland W. Cooper: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing–review & editing, visualization, supervision. Shuheng Guo: methodology, software, validation, formal analysis, data curation, writing–review & editing, visualization. Dale Zhou: software, supervision. Craig E. L. Stark: conceptualization, methodology, resources, writing–review & editing, supervision, funding acquisition. Aaron M. Bornstein: conceptualization, methodology, validation, resources, writing–review & editing, supervision, project administration, funding acquisition.

Data availability

All de-identified data and the final versions of all R scripts used to generate each data figure, as well as Pytorch model code, are freely available at the *UCI CCNL GitHub repository* (https://github.com/uciccnl). This study was not preregistered.

Acknowledgments

S.M.N. and K.W.C. served as co-first authors for this study.

References

Akhtar, A. (2019, November 1). *Corporate America is seeing a spike in the age of CEOs being hired – and yes, they're overwhelmingly white men*. Business Insider. https://www.businessinsider.in/strategy/news/corporate-america-is-seeing-a-spike-in-the-age-of-ceos-being-hired-and-yes-theyre-overwhelmingly-white-men/articleshow/71856494.cms

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. L. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science (New York, N.Y.)*, *319*, 1640–1642. https://doi.org/10.1126/science.1152882

Beukers, A. O., Collin, S. H. P., Kempner, R. P., Franklin, N. T., Gershman, S. J., & Norman, K. A. (2024). Blocked training facilitates learning of multiple schemas. *Communications Psychology*, *2*, 28. https://doi.org/10.1038/s44271-024-00079-4

Blair, D. C. (1979). Information retrieval (2nd ed.). InVan Rijsbergen, C. J. (Ed.), 1979 (pp. 208). London: Butterworths.. *Journal of the American Society for Information Science*, *30*, 374–375. https://doi.org/10.1002/asi.4630300621

Blanco Malerba, S., Micheli, A., Woodford, M., & Azeredo da Silveira, R. (2024). Jointly efficient encoding and decoding in neural populations. *PLoS Computational Biology*, *20*, e1012240. https://doi.org/10.1371/journal.pcbi.1012240

Bornstein, A. M., & Daw, N. D. (2013). Cortical and hippocampal correlates of deliberation during model-based decisions for rewards in humans. *PLoS Computational Biology*, *9*, e1003387. https://doi.org/10.1371/journal.pcbi.1003387

Bowman, C. R., de Araujo Sanchez, M.-A., Hou, W., Rubin, S., & Zeithamova, D. (2021). Generalization and false memory in an acquired equivalence paradigm: The influence of physical resemblance across related episodes. *Frontiers in Psychology*, *12*, 669481. https://doi.org/10.3389/fpsyg.2021.669481

Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin*, *145*, 1029–1052. https://doi.org/10.1037/bul0000209

Burke, S. N., Wallace, J. L., Nematollahi, S., Uprety, A. R., & Barnes, C. A. (2010). Pattern separation deficits may contribute to age-associated recognition impairments. *Behavioral Neuroscience*, *124*, 559–573. https://doi.org/10.1037/a0020893

Campbell, K. L., Hasher, L., & Thomas, R. C. (2010). Hyper-binding: A unique age effect. *Psychological Science*, *21*, 399–405. https://doi.org/10.1177/0956797609359910

Carpenter, A. C., Thakral, P. P., Preston, A. R., & Schacter, D. L. (2021). Reinstatement of item-specific contextual details during retrieval supports recombination-related false memories. *NeuroImage*, *236*, 118033. https://doi.org/10.1016/j.neuroimage.2021.118033

Carterette, E. C. (1967). Toward a more rigorous psychophysics: Signal detection theory and psychophysics. In David M. Green, John A. Swets, 1966 (pp. 467). Wiley, New York. *Science*, *156*, 632–633. https://doi.org/10.1126/science.156.3775.632

Castel, A. D. (2007). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. In A. S. Benjamin, & B. H. Ross (Eds.), *Psychology of learning and motivation*. (Vol. *48*, pp. 225–270). Academic Press. https://doi.org/10.1016/S0079-7421(07)48006-9

Castel, A. D., Benjamin, A. S., Craik, F. I. M., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition*, *30*, 1078–1085. https://doi.org/10.3758/bf03194325

Chandak, S., Shah, P., Borkar, V. S., & Dodhia, P. (2024). Reinforcement learning in non-Markovian environments. *Systems & Control Letters*, *185,* 105751. https://doi.org/10.1016/j.sysconle.2024.105751

Craik, F. I. M., & Salthouse, T. A. (2011). *The handbook of aging and cognition* (3rd ed.) Psychology Press. https://play.google.com/store/books/details? id=YeJ4AgAAQBAJ

Doll, B. B., Shohamy, D., & Daw, N. D. (2015). Multiple memory systems as substrates for multiple decision systems. *Neurobiology of Learning and Memory*, *117*, 4–13. https://doi.org/10.1016/j.nlm.2014.04.014

Drummond, N., & Niv, Y. (2020). Model-based decision making and model-free learning. *Current Biology: CB*, *30*, R860–R865. https://doi.org/10.1016/j.cub.2020.06.051

Eichenbaum, H. (2001). The hippocampus and declarative memory: Cognitive mechanisms and neural codes. *Behavioural Brain Research*, *127*, 199–207. https://doi.org/10.1016/s0166-4328(01)00365-5

Eichenbaum, H. (2017). Memory: Organization and control. *Annual Review of Psychology*, *68*, 19–45. https://doi.org/10.1146/annurev-psych-010416-044131

Eppinger, B., & Bruckner, R. (2015). Chapter 4 – Towards a mechanistic understanding of age-related changes in learning and decision making: A neuro-computational approach. In T. M. Hess, J. Strough, & C. E. Löckenhoff (Eds.), *Aging and decision making* (pp. 61–77). Academic Press. https://doi.org/10.1016/B978-0-12-417148-0.00004-2

Eppinger, B., Walter, M., Heekeren, H. R., & Li, S.-C. (2013). Of goals and habits: Age-related and individual differences in goal-directed decision-making. *Frontiers in Neuroscience*, *7*, 253. https://doi.org/10.3389/fnins.2013.00253

Feher da Silva, C., & Hare, T. A. (2020). Humans primarily use model-based inference in the two-stage task. *Nature Human Behaviour*, *4*, 1053–1066. https://doi.org/10.1038/s41562-020-0905-y

Grace-Martin, K. (2011, April 2). *Testing and dropping interaction terms in regression and ANOVA models*. The Analysis Factor. https://www.theanalysisfactor.com/testing-and-dropping-interaction-terms/

Hall, M., & Hickey, W. (2022, September 14). *This one chart shows an unprecedented recent spike in age for Congress, posing a huge problem for American democracy*. Business Insider. https://www.businessinsider.com/congress-oldest-history-gerontocracy-lawmakers-2022-9

Harhen, N. C., & Bornstein, A. M. (2023). Overharvesting in human patch foraging reflects rational structure learning and adaptive planning. *Proceedings of the National Academy of Sciences of the United States of America*, *120*, e2216524120. https://doi.org/10.1073/pnas.2216524120

Hess, T. M., Strough, J., & Löckenhoff, C. (2015). *Aging and decision making: Empirical and applied perspectives*. Academic Press.

Huang, Z., Liu, F., & Su, H. (2019). Mapping state space using landmarks for universal goal reaching. *Advances in Neural Information Processing Systems*, *32*.

Ingraham, C. (2014, April 1). *Yes, Congress is getting older–but so are the rest of us*. The Washington Post. https://www.washingtonpost.com/news/wonk/wp/2014/04/01/yes-congress-is-getting-older-but-so-are-the-rest-of-us/

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*, 97–103. https://doi.org/10.1002/acp.1801

Kirwan, C. B., & Stark, C. E. L. (2007). Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *14*, 625–633. https://doi.org/10.1101/lm.663507

Konovalov, A., & Krajbich, I. (2020). Mouse tracking reveals structure knowledge in the absence of model-based choice. *Nature Communications*, *11*, 1893. https://doi.org/10.1038/s41467-020-15696-w

Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Review*, *119*, 573–616. https://doi.org/10.1037/a0028681

Lacy, J. W., Yassa, M. A., Stark, S. M., Muftuler, L. T., & Stark, C. E. L. (2011). Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *18*, 15–18. https://doi.org/10.1101/lm.1971111

McCloskey, M., & Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. H. Bower (Ed.), *Psychology of learning and motivation*. (Vol. *24*, pp. 109–165). Academic Press. https://doi.org/10.1016/S0079-7421(08)60536-8

Morton, N. W., Sherrill, K. R., & Preston, A. R. (2017). Memory integration constructs maps of space, time, and concepts. *Current Opinion in Behavioral Sciences*, *17*, 161–168. https://doi.org/10.1016/j.cobeha.2017.08.007

Noh, S. M., Bjork, R. A., & Preston, A. R. (2024). General knowledge and detailed memory benefit from different training sequences. *Journal of Applied Research in Memory and Cognition*, *13*, 329–341. https://doi.org/10.1037/mac0000193

Noh, S. M., Singla, U. K., Bennett, I. J., & Bornstein, A. M. (2023). Memory precision and age differentially predict the use of decision-making strategies across the lifespan. *Scientific Reports*, *13*, 17014. https://doi.org/10.1038/s41598-023-44107-5

Noh, S. M., Yan, V. X., Bjork, R. A., & Maddox, W. T. (2016). Optimal sequencing during category learning: Testing a dual-learning systems perspective. *Cognition*, *155*, 23–29. https://doi.org/10.1016/j.cognition.2016.06.007

Noh, S. M., Zhou, D., Cooper, K. W., Guo, S., Dinh, E. T., & Bornstein, A. M. (2026). Sparsity and memory constraints interact with training sequence to bias learning of associative maps. *Neuropsychologia*, *225*, 109396. https://doi.org/10.1016/j.neuropsychologia.2026.109396

Pettigrew, C., & Martin, R. C. (2014). Cognitive declines in healthy aging: Evidence from multiple aspects of interference resolution. *Psychology and Aging*, *29*, 187–204. https://doi.org/10.1037/a0036085

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., Van Willigen, B., & Maintainer, R. (2017). Package 'nlme. *Linear and nonlinear mixed effects models (Version, 3)*. http://cran.rapporter.net/web/packages/nlme/nlme.pdf

Poppenk, J., Evensmoen, H. R., Moscovitch, M., & Nadel, L. (2013). Long-axis specialization of the human hippocampus. *Trends in Cognitive Sciences*, *17*, 230–240. https://doi.org/10.1016/j.tics.2013.03.005

Pudhiyidath, A., Morton, N. W., Duran, R. V., Schapiro, A. C., Momennejad, I., Hinojosa-Rowland, D. M., Molitor, R. J., & Preston, A. R. (2022). Representations of temporal community structure in hippocampus and precuneus predict inductive reasoning decisions. *Journal of Cognitive Neuroscience*, *34*, 1736–1760. https://doi.org/10.1162/jocn_a_01864

Raz, N. (2000). Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings (2nd ed.). In F. I. M. Craik, & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 1–90). Lawrence Erlbaum Associates Publishers.

Rmus, M., Ritz, H., Hunter, L. E., Bornstein, A. M., & Shenhav, A. (2022). Humans can navigate complex graph structures acquired during latent learning. *Cognition*, *225*, 105103. https://doi.org/10.1016/j.cognition.2022.105103

Samanez-Larkin, G. R. (2013). Financial decision making and the aging brain. *APS Observer*, *26*, 30–33. https://www.ncbi.nlm.nih.gov/pubmed/23946614

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, *16*, 486–492. https://doi.org/10.1038/nn.3331

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications*, *6*, 8151. https://doi.org/10.1038/ncomms9151

Schlichting, M. L., & Preston, A. R. (2015). Memory integration: Neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences*, *1*, 1–8. https://doi.org/10.1016/j.cobeha.2014.07.005

Schnoor, A. (2020, February 6). *The problem of the aging CEO*. The Intelligence of Everything. https://medium.com/the-intelligence-of-everything/the-problem-of-the-aging-ceo-e09e8d21bb65

Shimamura, A. P., & Jurica, P. J. (1994). Memory interference effects and aging: Findings from a test of frontal lobe function. *Neuropsychology*, *8*, 408–412. https://doi.org/10.1037/0894-4105.8.3.408

Stark, S. M., Kirwan, C. B., & Stark, C. E. L. (2019). Mnemonic similarity task: A tool for assessing hippocampal integrity. *Trends in Cognitive Sciences*, *23*, 938–951. https://doi.org/10.1016/j.tics.2019.08.003

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

Toner, C. K., Pirogovsky, E., Kirwan, C. B., & Gilbert, P. E. (2009). Visual object pattern separation deficits in nondemented older adults. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *16*, 338–342. https://doi.org/10.1101/lm.1315109

Vikbladh, O. M., Meager, M. R., King, J., Blackmon, K., Devinsky, O., Shohamy, D., Burgess, N., & Daw, N. D. (2019). Hippocampal contributions to model-based planning and spatial memory. *Neuron*, *102*, 683–693.e4. https://doi.org/10.1016/j.neuron.2019.02.014

Wilson, I. A., Gallagher, M., Eichenbaum, H., & Tanila, H. (2006). Neurocognitive aging: Prior memories hinder new hippocampal encoding. *Trends in Neurosciences*, *29*, 662–670. https://doi.org/10.1016/j.tins.2006.10.002

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*, 515–525. https://doi.org/10.1016/j.tins.2011.06.006

Yassa, M. A., Stark, S. M., Bakker, A., Albert, M. S., Gallagher, M., & Stark, C. E. L. (2010). High-resolution structural and functional MRI of hippocampal CA3 and dentate gyrus in patients with amnestic Mild Cognitive Impairment. *NeuroImage*, *51*, 1242–1252. https://doi.org/10.1016/j.neuroimage.2010.03.040

Yoo, J., Chrastil, E. R., & Bornstein, A. M. (2024). Cognitive graphs: Representational substrates for planning. *Decision (Washington, D.C.)*, *11*, 537–556. https://doi.org/10.1037/dec0000249

Zeithamova, D., & Preston, A. R. (2017). Temporal proximity promotes integration of overlapping events. *Journal of Cognitive Neuroscience*, *29*, 1311–1323. https://doi.org/10.1162/jocn_a_01116

Zeithamova, D., Schlichting, M. L., & Preston, A. R. (2012). The hippocampus and inferential reasoning: Building memories to navigate future decisions. *Frontiers in Human Neuroscience*, *6*, 70. https://doi.org/10.3389/fnhum.2012.00070

Zhou, Z., Singh, D., Tandoc, M. C., & Schapiro, A. C. (2023). Building integrated representations through interleaved learning. *Journal of Experimental Psychology: General*, *152*, 2666–2684. https://doi.org/10.1037/xge0001415

The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 2026, 81, 1–13
https://doi.org/10.1093/geronb/gbag038
Research Article