Title: Multi-step planning across the human lifespan can be improved with individualized memory interventions

Sharon M Noh[1†], Keiland W Cooper[2†], Shuheng Guo[1], Dale Zhou[1,2], Craig EL Stark[2,3], Aaron M Bornstein[1,3]*

1. Department of Cognitive Sciences, The University of California, Irvine, Irvine, CA, USA
2. Department of Neurobiology and Behavior, The University of California, Irvine, Irvine, CA, USA
3. Center for the Neurobiology of Learning and Memory, The University of California, Irvine, Irvine, CA, USA

[†] Co-first authors

* Corresponding Author: aaron.bornstein@uci.edu

Word Count: 5345 (1046 Introduction, 1796 Method, 1453 Results, 1050 Discussion)

Number of References: 54

Number of Tables/Figures: 7 + 1 Supplement

Funding:
This work was supported by the National Institute of Aging at the National Institutes of Health (R21AG072673 and R01AG088306 to AMB, R01AG066683 and P30AG066519 to CELS, and F32AG072836 to SMN); and The National Science Foundation (NSF DGE-1839285 to KWC).

Conflict of Interest Statement:
The authors declare no competing financial interests.

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, will be made freely available at the UCI CCNL GitHub repository (https://github.com/uciccnl) upon publication.

Author Contributions
SMN: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing – original draft, writing – review & editing, visualization, supervision, project administration, funding acquisition
KWC: conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing – review & editing, visualization, supervision
SG: methodology, software, validation, formal analysis, data curation, writing – review & editing, visualization
DZ: software, supervision
CELS: conceptualization, methodology, resources, writing – review & editing, supervision, funding acquisition
AMB: conceptualization, methodology, validation, resources, writing – review & editing, supervision, project administration, funding acquisition

Abstract

Objectives. Effective goal-directed decision-making relies on both memory and planning–processes that are each 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 multi-step associations among interconnected stimuli, assessing performance across the adult lifespan. In Experiment 1, we examined relationships between judgment performance and mnemonic discrimination ability. In Experiment 2, we tested whether a learning schedule designed to reduce memory interference (by temporally separating overlapping object pairs) could improve performance, particularly for individuals with weaker memory abilities. We also implemented an artificial neural network simulation varying training sequence and the network's representational capacity to model performance.

Results. Across the lifespan, both young 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. The neural network simulation reproduced these patterns.

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

Word Count: 244

Keywords: individual differences, latent learning, memory precision, inference, training intervention, goal-directed planning

Introduction

Humans across the lifespan 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 multi-step associations and judgments, such as forward planning (Drummond & Niv, 2020; Eppinger et al., 2013; Eppinger & Bruckner, 2015). However, the mechanism by which multi-step inference and planning declines 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 (Konovalov & Krajbich, 2020; Silva et al., 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 4PM Friday flight and plans to call a taxi at noon to allow ample time for travel. Even if this is her first conference, she can engage in multi-step 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." (Rmus et al., 2022). 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 multi-step 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* structure learning occurs, and what type of knowledge representations support multi-step 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., 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 multi-step 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 multi-step inference judgments needed for complex decision-making tasks. In support of this idea, work in 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, 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"). While 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 causes inference failures, then a learning method that reduces memory interference should improve latent structure learning and judgments.

In the present study, we tested whether age-related declines in memory precision affect performance on a multi-step associative inference task (the "graph" task; (Rmus et al., 2022). In Experiment 1, participants studied randomly presented object pairs drawn from an underlying graph of 12 nodes (objects) and 16 edges (object pairs; Fig. 1). 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 (Fig. 2A). We expected performance deficits with age, specifically driven by poor memory precision, so participants also completed the Mnemonic Similarity Task (MST; Fig. 3) 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, Fig. 1C) would reduce working memory load and improve structural inference and multi-step planning for older adults or those with low precision (Schlichting et al., 2015). To affirm the relationship between training condition and memory precision, we simulated task performance using variants of artificial neural network models that differed only in their internal representational capacity and compared the model outputs to our behavioral findings.

Method

Transparency and Openness. All de-identified data, figure scripts, and PyTorch code will be shared at the UCI CCNL GitHub upon publication. This study was not preregistered.

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). We report them separately to reflect the study's logical progression from replication to extension. Experiment 1's intermixed task closely followed Rmus et al. (2022) but included a lifespan sample with older adults, was administered online, and incorporated an additional measure of memory specificity. 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). 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.). We aimed for ~100 participants per experiment post-exclusion. A total of 219 participants (112 female, 108 male; ages 19-84, mean(sd) age = 55.7(14.2)) were recruited via Amazon Mechanical Turk: 113 in Experiment 1 (59 female / 54 male, ages 22-84, mean(sd) age = 56.7(13.8)), and 106 in Experiment 2 (53 female / 53 male, ages

19-79, mean(sd) age = 54.8(14.7)). Upon consent, participants reviewed an online information sheet outlining procedures, rights, risks, and compensation. 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 "Exempt Online Survey" by UCI's IRB. Participants logged in with their Amazon Worker ID, which was anonymized using a one-way hash before storage.

Procedure

Structural Inference ("Graph") Task

Study Phase. Participants viewed and were told to memorize a series of object pairs (Fig. 1A), each presented for 1s 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 (Fig. 1B). In Experiment 1, all 16 unique edge pairs were repeated 44 times (704 trials total) in random sequence (Fig. 1C, "intermixed"). Experiment 2 grouped object pairs into 4 mini-blocks, each with 4 unique, non-overlapping pairs (Fig. 1C, "blocked"). This separation of potentially confusing (overlapping) edges across time aimed to reduce memory interference and improve learning of the graph structure.

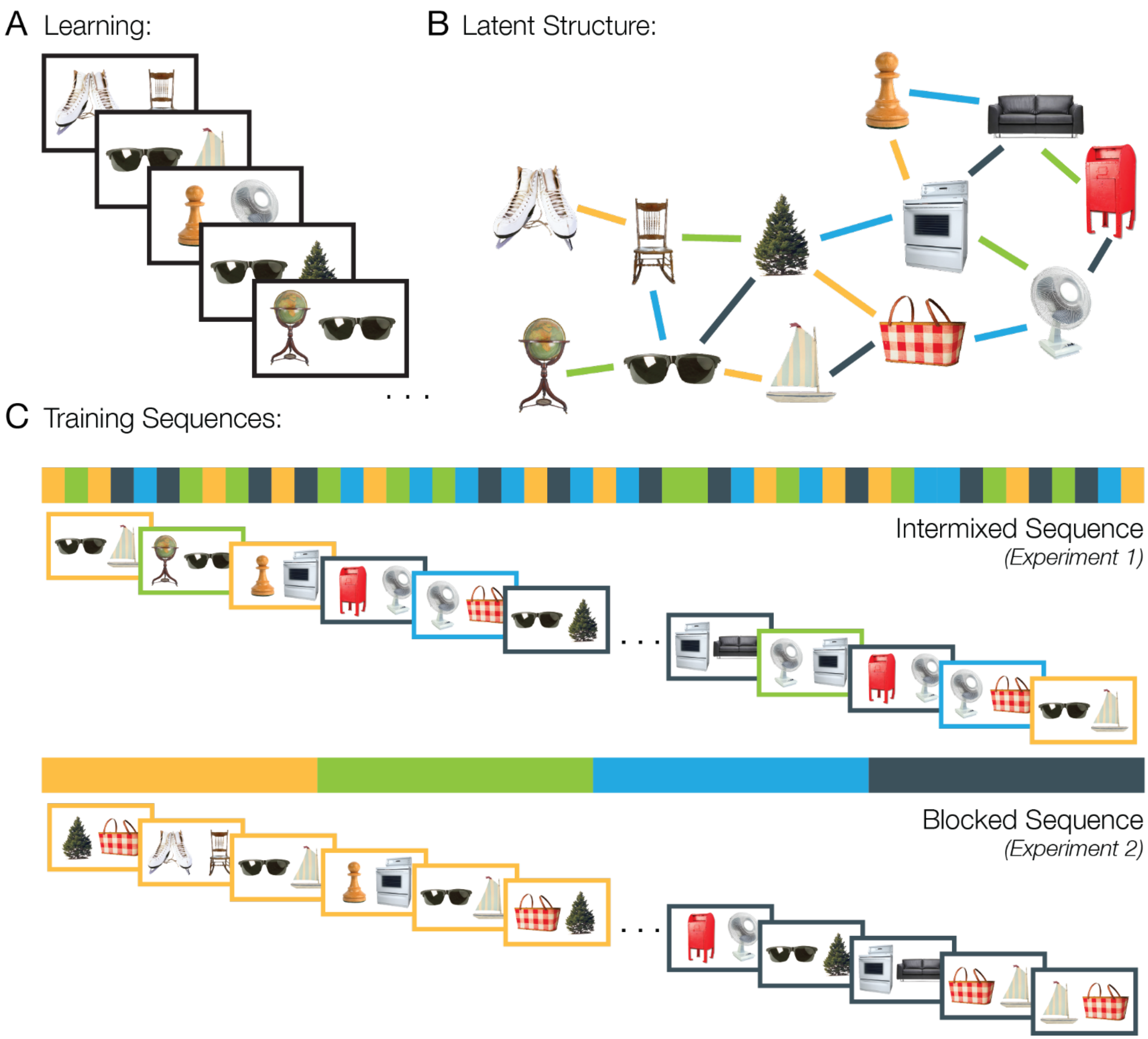


Figure 1. Structural Inference Task learning phase. Participants learn individual edges (A) drawn from a latent structure made up of several overlapping edges (B). In Experiment 1, edges are drawn randomly and presented in an intermixed fashion (C, intermixed), whereas in Experiment 2, overlapping edges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

Following the study phase, participants completed two different tests to measure how well participants can *use* their knowledge of the graph structure (Fig. 2A, *Judgment Test*), as well as how well they *learned* the graph structure (Fig. 2B, *Graph Reconstruction Test)*.

Judgment Test. After completing the learning phase, participants completed a relative distance judgment task (Fig. 2A). This task measured participants' ability to use learned information for structural inference. Participants were asked to judge which of two objects (left vs. 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 seconds 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 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. Data were first combined across all conditions (age, training sequence). We excluded trials with RTs <400 ms, which were determined to be unrealistic (the lowest participant-level average RT was 463.2 ms), and >3000 ms, which may indicate connectivity issues, lag, or lapses in attention. This initial filter yielded a general acceptable range of [400, 3000 ms]. From this set, we further retained trials within twice the interquartile range ([Q3 – 1.5×IQR, Q1 + 1.5×IQR]) of the remaining data ([494, 2226 ms]). This range still encompassed the IQR of the original data (median = 1368 ms, IQR = [951, 2062]) and preserved sufficient variability to examine potential strategy differences, while excluding trials where participants were likely disengaged. These criteria align with prior approaches for filtering noise in online RT data (Ratcliff & Hendrickson, 2021).

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. interleaved) were 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 (Fig. 2B). 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 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 scored reconstructions as follows: "hits" for correctly linked study pairs (Fig. 5A), "false alarms" for incorrectly linked pairs (Fig. 5B), and total number of edges drawn (Fig. 5C). Accuracy was computed in two ways: 1) sensitivity (d') defined as hit rate - false alarm rate (Fig. 5D; $hits/16 - false\ alarms/50$), and 2) precision (Fig. 5E; $hits\ /\ total\ edges\ drawn$). The precision metric adjusts performance by 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 ~ 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).

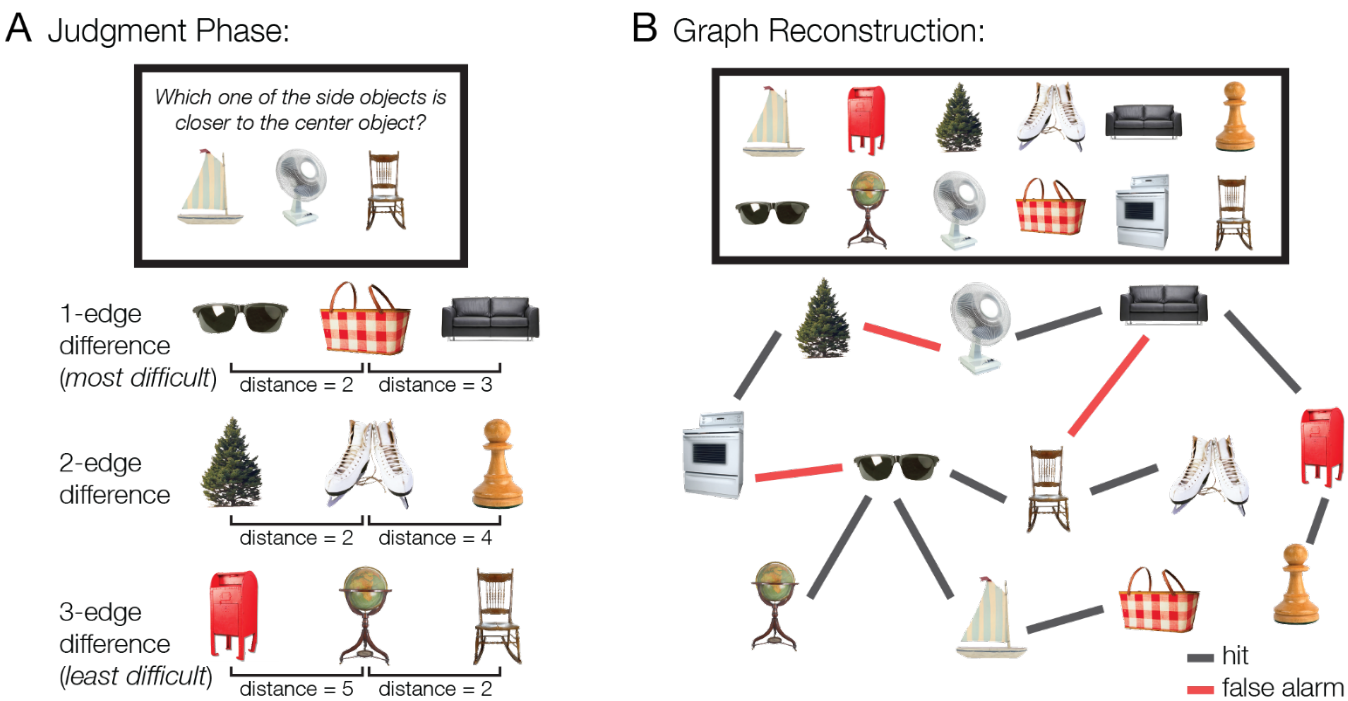


Figure 2. Structural Inference Task test phases. (A) After the learning phase, structural inference-based judgments were assessed for each participant. Participants were presented with 3 objects and asked whether the object on the left or right was closer to the center object, based on the associations 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 5 (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 object on the canvas, and each object had to be connected to at least one other object to complete this phase.

<u>Mnemonic Similarity Task (MST)</u>

Participants completed the MST (Fig. 3) 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).

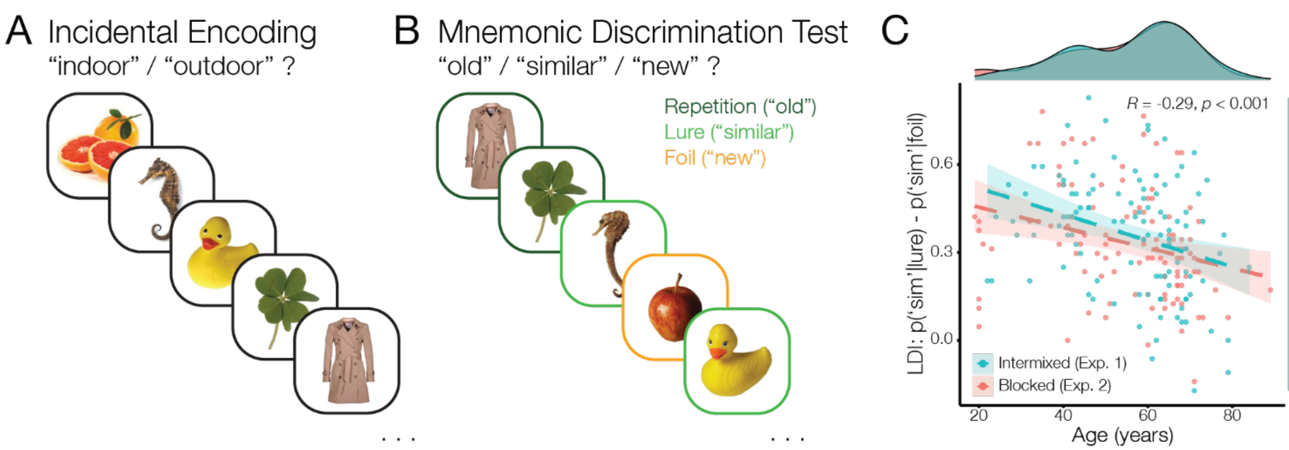


Figure 3. The Mnemonic Discrimination Task. (A) Participants view a sequence of objects during an incidental encoding phase in which participants are asked to classify each object as an indoor or outdoor object. (B) In a surprise discrimination test, participants view a series of objects and are asked 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).

Artificial Neural Network Model (ANN)

To further examine the behavior of participants under different simulated conditions, we implemented a minimal-assumption feedforward network in PyTorch (Version 1.13.1, (Paszke et al., 2019)) under varying representational capacities. To avoid terminology confusion, *node/edge* refer to the graph task and *unit/weight* to the neural network. Given one node of a studied pair, the network learned to predict (pattern-complete) its complementary vertex. To simulate variance in memory encoding ability, the number of units in the 2nd and 4th layers of the network was varied. We used a fully connected five-layer network not including the 12 units input and output layers (one per item) with an 12-unit hidden layer. To model different degrees of pattern separation, the widths of the second and fourth layers were varied at increments of one-quarter the input size: 6, 12, 18, 24, 36 units. Smaller widths simulate low memory capacity by forcing many-to-one mappings (greater potential for interference). The model's activation function at unit *i* of layer *l* was defined as:

$$h_l(i) = f(w^T{}_{il} h_{l-1})$$

where *f* is a pointwise function, $\mathbf{w}_{il}$ is a vector of learnable parameters, $\mathbf{h}_{l-1}$ represents the output of the previous layer, and *T* is the matrix transposition operation. The input layer is referred to as h_0. For each condition (blocked vs. intermixed), the model saw the same image-pair stimuli as participants, <x,y>. Each of the 12 images was one-hot encoded (the input unit for the presented node set to 1, others 0). The target was a one-hot vector for the paired node (1 for the complement; 0 otherwise). Training used both directions of each undirected pair with equalized exposures per direction. We estimated the parameters $\mathbf{w}_{il}$, for all *i* and *l*, using stochastic gradient descent with the Adam optimizer (Kingma & Ba, 2014) to minimize the cosine similarity between the input and output. We used a rectified nonlinear unit for the pointwise function, f, for all layers except the output layer, where we used a sigmoid function. To account for variability due to stochastic weight initialization and to evaluate model generalizability, we trained 1000 models per condition (blocked and interleaved) across 5 different latent layer widths (6, 12, 18, 24, 36). All models used He initialization, which samples weights from a uniform distribution scaled for ReLU activations to maintain stable gradients. We optimized cross-entropy over the 12-way next-item distribution using Adam without label smoothing, applying ReLU after the first two layers (L1 and L2), a linear bottleneck layer (B), and symmetric decoding layers back to the output logits. Learning rate and weight decay were set to 0.38 and 0.13, respectively, based on prior runs and sweep tests. Models were trained for 5 epochs, with each epoch containing 704 trials. In the blocked condition, these were divided into 4 blocks of 176 trials each, presented in a strict block-wise schedule; in the interleaved condition, all 704 trials were shuffled.

To assess the ability of the model to infer the latently learned graph, we performed a relative distance judgment task. We evaluated all non-adjacent triples of nodes (reference, option 1, option 2) with a shortest-path length difference |d12−d13| of 1 to 3. The model's internal representations were taken from the bottleneck layer. For each trial, we estimated the number of hops from the reference to each option using a beam search (keeping only the top-k candidates by similarity at each step), and selected the closer option via a softmax over the two estimated hop counts.

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 (Fig. 4A): judgment accuracy ~ age + distDiff + age*distDiff, random = ~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 = 0.011), with older adults performing worse than younger adults. There was also a main effect of difficulty ($\beta_{distDiff}$ = 0.032, SE = 0.009, t(216) = 3.50, p < 0.001), with performance improving as relative distance increased. The age × difficulty interaction was significant ($\beta_{age*distDiff}$ = -0.022, SE = 0.009, t(216) = -2.41, p = 0.017) such that younger adults showed greater performance gains with increasing relative distance.

Adding mnemonic discrimination ability (LDI) to the model revealed it was a stronger predictor of judgment accuracy than age: judgment accuracy ~ age + distDiff + LDI + age*distDiff + age*LDI + distDiff*LDI + age*distDiff*LDI, random = ~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 = 0.088) and its interaction with difficulty ($\beta_{age*distDiff}$ = -0.012, SE = 0.010, t(214) = -1.24, p = 0.218) were no longer significant. Instead, higher LDI predicted better performance (β_{LDI} = 0.031, SE = 0.013, t(108) = 2.42, p = 0.017), with a significant LDI × difficulty interaction ($\beta_{LDI*distDiff}$ = 0.033, SE = 0.010, t(214) = 3.34, p = 0.001) indicating that memory precision benefits emerged primarily in easier trials. Age × LDI and the three-way interaction were non-significant. Given LDI's predictive strength, we re-ran the model excluding age (Fig. 4B, adjusted $R^2 = 0.26$, AIC = -164.19). This model confirmed significant main effects of LDI ($\beta_{LDI} = 0.035$, SE = 0.012, $t(110) = 2.99$, $p = 0.003$) and difficulty ($\beta_{distDiff} = 0.032$, SE = 0.009, $t(216) = 3.50$, $p < 0.001$) as well as a robust LDI × difficulty interaction ($\beta_{LDI*distDiff} = 0.034$, SE = 0.009, $t(216) = 3.84$, $p < 0.001$), with higher LDI predicting greater gains in easier trials.

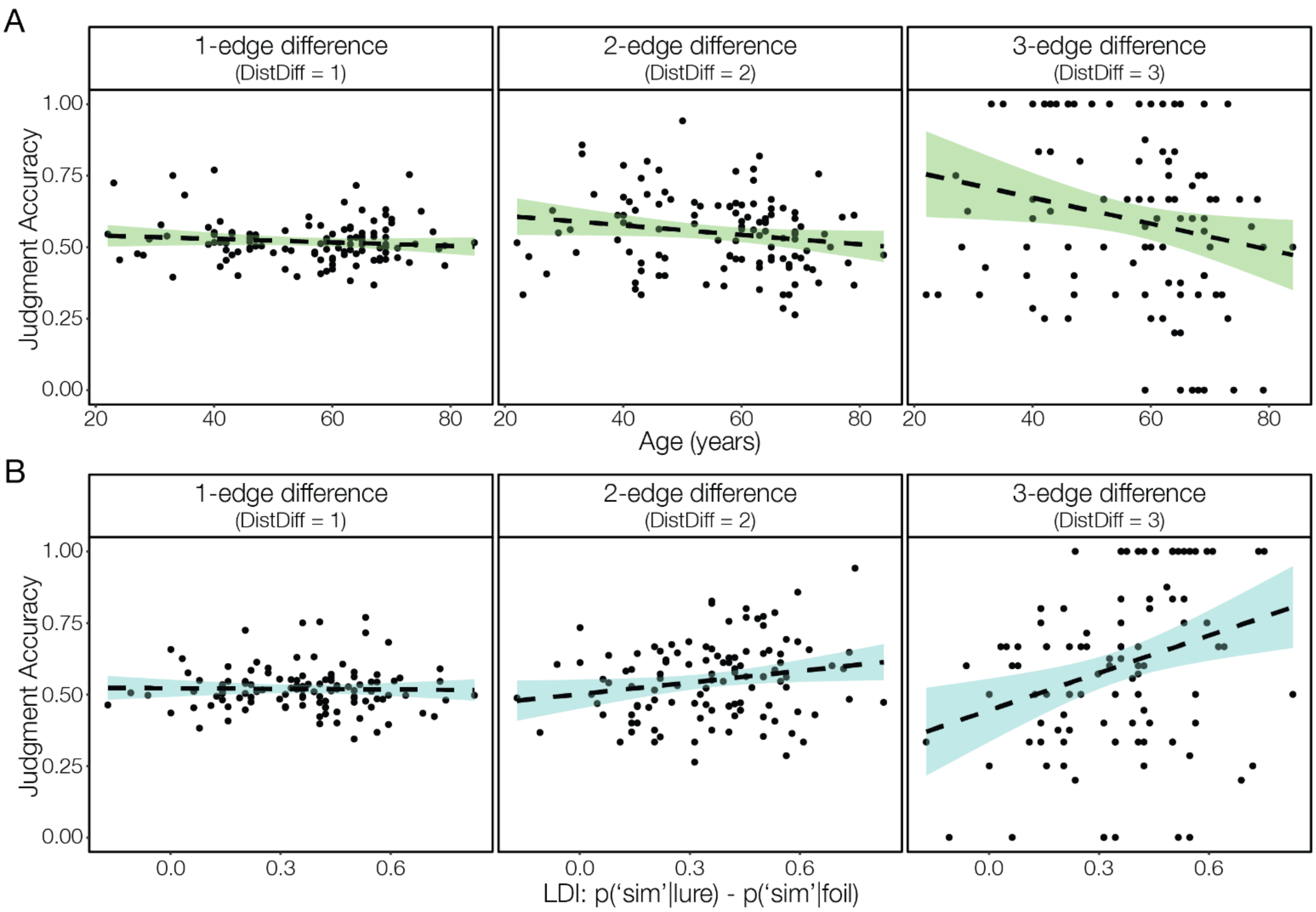


Figure 4. Behavioral results from Experiment 1. (A) Judgment accuracy as a function of chronological age and trial difficulty (distance difference between choice options). The green band indicates 95% confidence intervals around the best fit regression line. (B) Judgment accuracy as a function of mnemonic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

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. 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 study 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 (Fig. 5): judgment accuracy ~ LDI + distDiff + Sequence + LDI*distDiff + Sequence*LDI + Sequence*distDiff + LDI*Sequence*distDiff, random = ~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 ($\beta_{LDI*distDiff*sequence}$ = 0.054, SE = 0.013, t(422) = 4.10, p < 0.001) 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 ($\beta_{LDI*sequence}$ = 0.038, SE = 0.016, t(214) = 2.26, p = 0.025): 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.

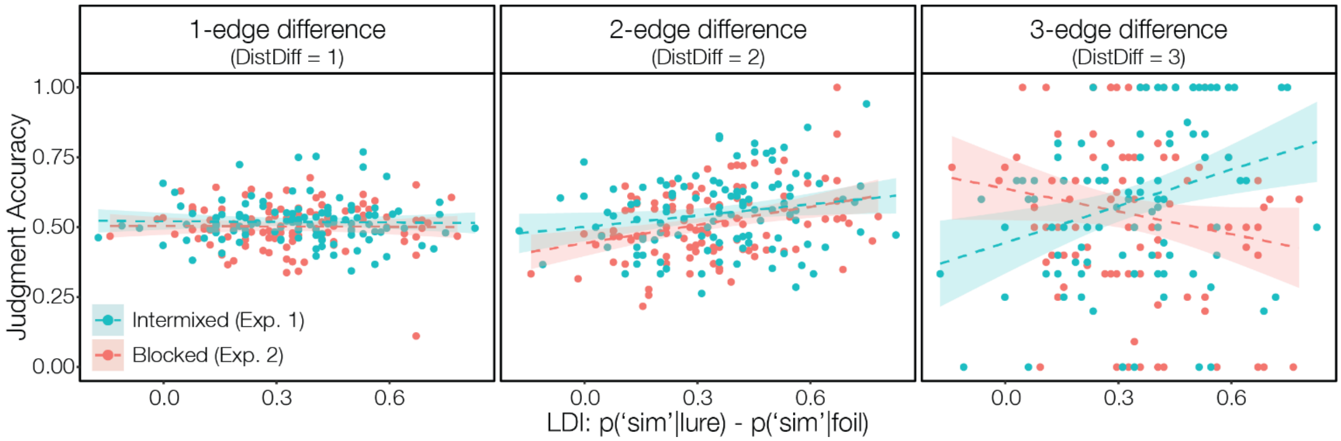


Figure 5. Behavioral results combining data from Experiment 1 (blue) and Experiment 2 (red) showing judgment accuracy as a function of trial difficulty (distance difference between choice options), mnemonic discrimination ability (LDI), and training condition (blocked vs. intermixed learning sequence).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

Artificial neural network 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 1–2, we tested whether such effects could be reproduced by manipulating a neural network's representational capacity. 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 from the intermixed version of the task (704 randomized presentations; see *Methods*: *Artificial Neural Network Model*). The trained network performed the judgment task by estimating hop counts using its bottleneck representations and choosing the closer option using a softmax over the two estimates. Consistent with the hypothesis that the network provides a minimal mechanism for multi-step inference, its distance-dependent performance matched that of Rmus et al. (2022; aged 18–27, $M = 22$; Fig. 4a in that paper) and the young/high-LDI participants in Experiment 1 (Fig. 4).

To model variability in memory precision, we manipulated the number of units in the second and fourth hidden layers (6, 12, 18, 24, or 36; Fig. 6a). Larger widths corresponded to more distinct, nonoverlapping representations of each node (high LDI), whereas smaller widths simulated representational merging before latent structure extraction (low LDI). This manipulation was loosely inspired by pattern separation in the dentate gyrus, where higher neuron density supports more precise memory representations (Yassa & Stark, 2011). Consistent with this hypothesis, we found that judgment task performance scaled with hidden-layer width, with the largest effects on distDiff = 3 trials that required precise, orthogonalized representations to compare long vs. short paths (Fig. 6b). We fit a linear mixed-effects regression predicting judgment accuracy from difficulty (distDiff: 1, 2, 3), training sequence (blocked/intermixed), and layer width (6–36), with model ID as a random effect. For distDiff = 3 trials, the analysis was: *judgment accuracy* ~ *Width* + *Sequence* + *Width* × *Sequence*, random = ~1|model_ID (AIC = −27,558.8; ML fit). At the baseline width, blocked models achieved a mean accuracy of 84.3%, whereas intermixed models were lower by 4.8% (SE = 0.0020, $z = -23.83$, $p < .001$). This substantial and highly significant difference indicates that, in low-difficulty conditions, blocked training improved structural inference accuracy, paralleling the behavioral advantage observed for low-LDI participants.

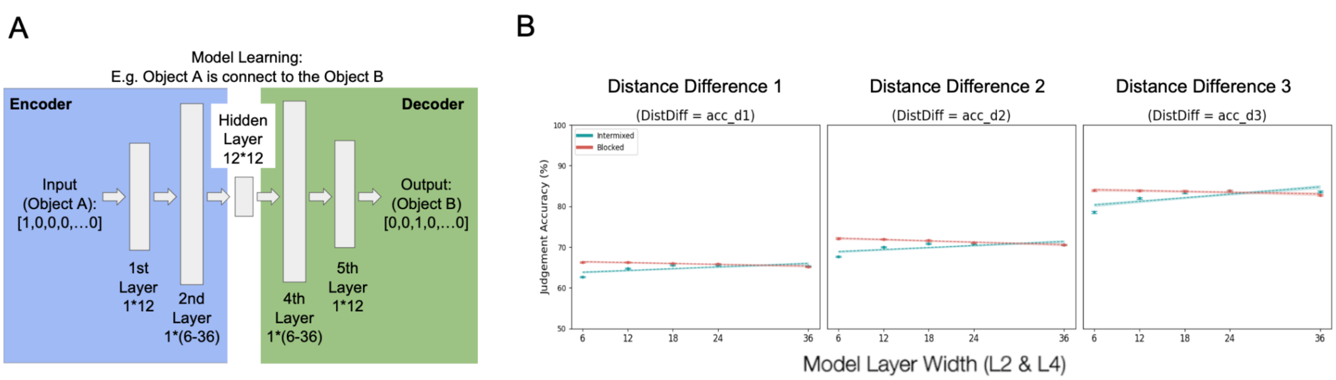


Figure 6. (A) Schematic of the artificial neural network. To model memory capacities' effect on judgment accuracy, the width of the first and third hidden layer were varied from 6 units (low-LDI; reflecting the hypothesized diminished capacity for separating inputs) to 324 units (high-LDI; allowing for 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. For Distance Difference 3, The effect of layer width differed significantly between training conditions($F(1,11{,}996)= 181.29$, $p<.001$, $\eta p^2=.015$), indicating that network capacity influenced structural inference accuracy differently for blocked and intermixed training. Accuracy decreased with width for blocked models (slope

= −0.00035) but increased for intermixed models (slope = +0.00148), with blocked outperforming intermixed at low widths and intermixed closing the gap at higher widths.

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 reconstruction phase as a function of mnemonic discrimination ability (LDI) and training sequence (blocked vs. intermixed). 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 $R^2 = .03$, $F(2, 212) = 4.09$, $p = .018$, with a main effect of LDI, $\beta = 0.65$, 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 (Fig. 7A). For false alarms, the model (*false alarms ~ LDI + Sequence*) was significant, adjusted $R^2 = .03$, $F(2, 212) = 4.07$, $p = .018$, with a main effect of training sequence, $\beta = 1.53$, SE = 0.55, $t(212) = 2.77$, $p = .006$, such that participants in the intermixed condition produced more false edges; LDI was not significant (Fig. 7B). For total edges, the model (*total edges ~ LDI + Sequence*) was significant, adjusted $R^2 = .05$, $F(2, 212) = 6.38$, $p = .002$, with main effects of LDI, $\beta = 0.79$, SE = 0.33, $t(212) = 2.36$, $p = .019$, and Sequence, $\beta = 1.71$, SE = 0.67, $t(212) = 2.55$, $p = .011$, indicating that higher-LDI participants and those in the intermixed condition drew more edges overall (Fig. 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 ~ 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 (Fig. 7D). For precision, which adjusts for the tendency of low-memory participants to draw fewer edges, the model (*precision ~ 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 (Fig. 7E).

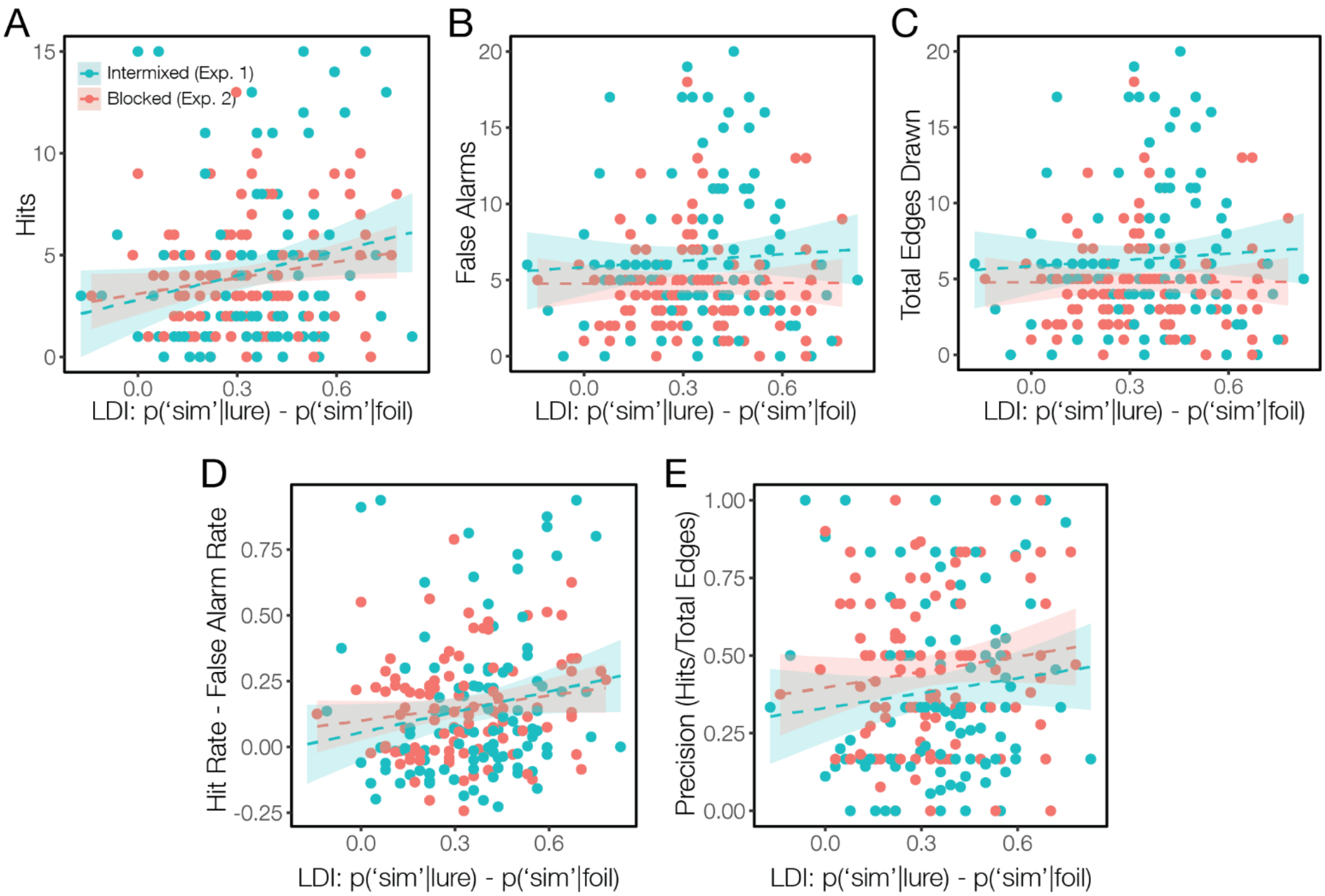


Figure 7. Behavioral results combining data from Experiment 1 (blue) and Experiment 2 (red) for the graph reconstruction phase. Shaded bands indicate 95% confidence intervals around the best fit regression line for each training condition (dashed lines).

Discussion

The present study examined whether multi-step 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. 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.

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. 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 (Supplemental Fig. 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 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—while the high-LDI blocked group exhibited low false alarms (Fig. 7B) and high graph reconstruction precision (Fig. 7E), suggesting more localist coding. However, this gain in precision came at the cost of impaired judgment performance (Fig. 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 (Supp. Fig. 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 in an intermixed schedule. With 16 overlapping pairs, interference may be so severe that learners can only focus on memorizing only 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). 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 (Fig. 5) and fewer false alarms (Fig. 7B), indicating reduced interference. 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. Future work should test how disruptions in latent structure learning impair multi-step 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 lifespan.

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*, *319*(5870), 1640–1642.

Blanco Malerba, S., Micheli, A., Woodford, M., & Azeredo da Silveira, R. (2024). Jointly efficient encoding and decoding in neural populations. *PLoS Computational Biology*, *20*(7), e1012240.

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*(12), e1003387.

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.

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*(5), 559–573.

Campbell, K. L., Hasher, L., & Thomas, R. C. (2010). Hyper-binding: a unique age effect. *Psychological Science*, *21*(3), 399–405.

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.

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*(7), 1078–1085.

Chandak, S., Shah, P., Borkar, V. S., & Dodhia, P. (2024). Reinforcement learning in non-Markovian environments. *Systems & Control Letters*, *185*(105751), 105751.

Craik, F. I. M., & Salthouse, T. A. (2011). *The Handbook of Aging and Cognition: Third Edition*. Psychology Press.

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.

Drummond, N., & Niv, Y. (2020). Model-based decision making and model-free learning. *Current Biology: CB*, *30*(15), R860–R865.

Eichenbaum, H. (2001). The hippocampus and declarative memory: cognitive mechanisms and neural codes. *Behavioural Brain Research*, *127*(1-2), 199–207.

Eichenbaum, H. (2017). Memory: Organization and Control. *Annual Review of Psychology*, *68*, 19–45.

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.

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.

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*. Insider. 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*(13), e2216524120.

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. In *arXiv [cs.LG]*. arXiv. https://proceedings.neurips.cc/paper/2019/hash/3b712de48137572f3849aabd5666a4e3-

Abstract.html

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/

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. In *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/1412.6980

Kirwan, C. B., & Stark, C. E. L. (2007). Overcoming interference: an fMRI investigation of pattern separation in the medial temporal lobe. *Learning & Memory* , *14*(9), 625–633.

Konovalov, A., & Krajbich, I. (2020). Mouse tracking reveals structure knowledge in the absence of model-based choice. *Nature Communications*, *11*(1), 1893.

Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. *Psychological Review*, *119*(3), 573–616.

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* , *18*(1), 15–18.

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.

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.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. In *arXiv [cs.LG]*. arXiv. https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html

Pettigrew, C., & Martin, R. C. (2014). Cognitive declines in healthy aging: evidence from multiple aspects of interference resolution. *Psychology and Aging*, *29*(2), 187–204.

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*(1). 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*(5), 230–240.

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*(10), 1736–1760.

Ratcliff, R., & Hendrickson, A. T. (2021). Do data from mechanical Turk subjects replicate accuracy, response time, and diffusion modeling results? *Behavior Research Methods*, *53*(6), 2302–2325.

Raz, N. (2000). Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings. *The Handbook of Aging and Cognition., 2nd Ed.*, *2*, 1–90.

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.

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*(4), 486–492.

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.

Schlichting, M. L., & Preston, A. R. (2015). Memory integration: neural mechanisms and implications for behavior. *Current Opinion in Behavioral Sciences*, *1*, 1–8.

Schnoor, A. (2020, February 6). *The Problem of the Aging CEO*. Medium: 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*(3), 408–412.

Silva, C. F. da, da Silva, C. F., & Hare, T. A. (2020). Humans primarily use model-based inference in the two-stage task. In *Nature Human Behaviour* (Vol. 4, Issue 10, pp. 1053–1066). https://doi.org/10.1038/s41562-020-0905-y

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*(11), 938–951.

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*(12), 2442–2449.

Toner, C. K., Pirogovsky, E., Kirwan, C. B., & Gilbert, P. E. (2009). Visual object pattern separation deficits in nondemented older adults. *Learning & Memory* , *16*(5), 338–342.

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*(3), 683–693.e4.

Wilson, I. A., Gallagher, M., Eichenbaum, H., & Tanila, H. (2006). Neurocognitive aging: prior memories hinder new hippocampal encoding. *Trends in Neurosciences*, *29*(12), 662–670.

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525.

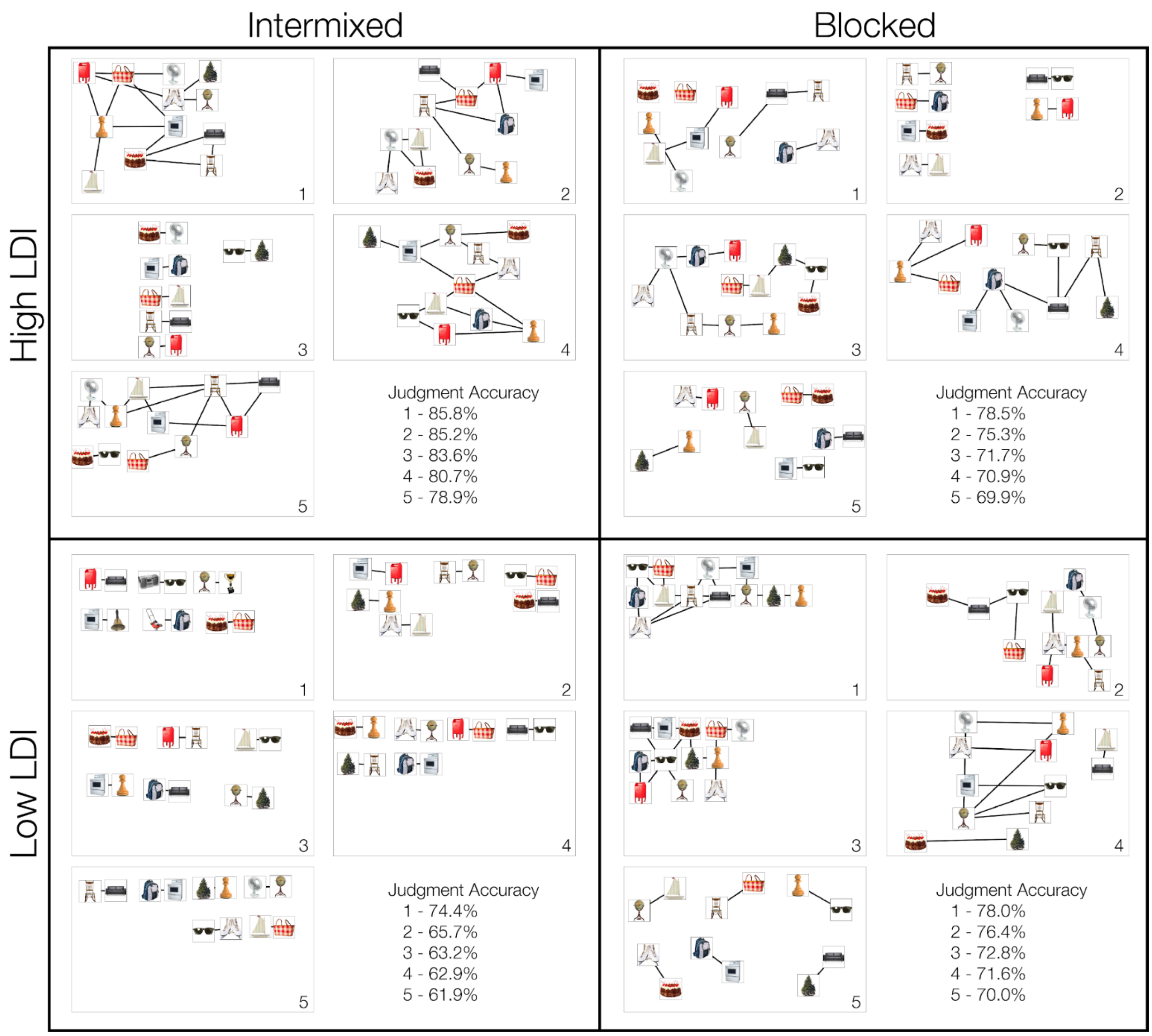
Yoo, J., Chrastil, E. R., & Bornstein, A. M. (2024). Cognitive graphs: Representational substrates for planning. *Decision (Washington, D.C.)*, *11*(4), 537–556.

Zeithamova, D., & Preston, A. R. (2017). Temporal Proximity Promotes Integration of Overlapping Events. *Journal of Cognitive Neuroscience*, *29*(8), 1311–1323.

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.

Zhou, Z., Singh, D., Tandoc, M. C., & Schapiro, A. C. (2023). Building integrated representations through interleaved learning. *Journal of Experimental Psychology. General*, *152*(9), 2666–2684.

Supplementary Material



Supplementary Figure 1. Sample graph reconstructions from participants with the overall best judgment accuracy within each condition. LDI groups were created using a median split.