Aging shifts recall of naturalistic events from temporal to topic organization

Angelique I. Delarazan[1], Katherine March[1],

Elena Markantonakis[2], June Dy[3], Zachariah M. Reagh[1]

[1]Washington University in St. Louis,

[2]Radboud University, [3]University of California, San Francisco

*Corresponding Author: Angelique I. Delarazan, a.delarazan@wustl.edu

ABSTRACT

Aging impacts episodic memories, including the way they are organized. Younger adults often organize items based on temporal context, whereas older adults show greater reliance on semantic relationships. Given that these findings are largely derived from studies using word-lists, an important question remains: Do these phenomena extend to complex, real-world memories that consists of meaningfully connected events? Here, we conducted an experiment using a naturalistic stimulus that allows one to organize recall along different principles: Temporal associations, semantic relationships, topics, and causally connected storylines. Older (N=36) and younger (N=46) participants viewed and recalled an episode of a "slice-of-life" television sitcom. We applied list-learning analysis techniques (e.g., lag-conditional response probabilities, temporal clustering, and topic clustering scores) to examine age differences in recall organization. Our results revealed that younger adults showed greater reliance on temporal structure, whereas older adults exhibited increased topic-based clustering. Older adults also showed diminished forward asymmetry even within shared topics, suggesting a broader shift away from fine-grained temporal structure. Despite these organizational differences, overall recall performance did not differ between groups; however, temporal organization was associated with better recall performance across age groups. Additionally, we found that events with more causal connections were recalled more than events with fewer causal connections across age groups, demonstrating the critical role of narrative structure in supporting memory retrieval. Together, our results demonstrate that aging shifts the principles underlying memory organization, and highlight the utility of naturalistic paradigms in revealing subtle changes in memory processes across the lifespan.

INTRODUCTION

Our ability to retrieve detailed information about the past declines as we get older (Park & Reuter-Lorenz, 2009; Naveh-Benjamin et al. 2003). Importantly, aging not only affects the likelihood of recalling specific details, but also influences the way those details are organized. Memory retrieval is a dynamic, interconnected process (Schacter et al., 2007), in which recalling one event can cue or reshape the recall of others (Radvansky & Dijkstra, 2007). For example, remembering a recent vacation might not only bring to mind the day-by-day sequence of activities during the trip itself, but also related memories of similar places visited or emotions associated with travel. These processes rely on retrieving isolated details as well as maintaining the relationships among them, including their temporal, thematic, and causal connections.

Most prior studies of recall organization in aging derive from list-learning tasks, where participants are presented with a sequence of items, often individual words, and later asked to recall them. When recalling lists, younger adults tend to cluster their recall based on the temporal proximity of items during encoding (Estes, 1955; Kahana et al., 1996), and typically demonstrate a strong forward recall bias (Howard & Kahana, 2002). These suggest that temporal context plays a major organizing role in how younger adults structure their recall. In contrast, older adults' recall of lists is less temporally organized (Healey & Kahana, 2016), which may contribute to their increased susceptibility to memory errors and misattributions, including familiarity-based errors (Jacoby, 1999) and difficulty inhibiting irrelevant information (Hasher et al., 1999). Rather than organizing memories based on time, older adults may rely more heavily on abstract associations, such as semantic relationships. For example, older adults are more likely to consecutively recall words based on shared meaning, such as belonging to the same taxonomic category (e.g., fruits; Golomb et al., 2008), suggesting a shift toward more meaningful-based retrieval strategies.

However, these findings largely stem from experiments using simple, isolated stimuli. Memories of real-world experience differ fundamentally from word lists. Everyday experiences are rich, complex, and embedded within meaningfully-connected structures that unfold over time. Reconstructing memories of real-world events requires not only retrieving specific details but also integrating higher-order information into a coherent narrative (Cohn-Sheehy et al., 2022l; Delarazan et al., 2025). This distinction raises important questions about whether the patterns of memory organization observed in simple laboratory tasks generalize to the retrieval of complex, naturalistic memories. One influential view suggests that memory difference between younger and older adults may reflect not just deficits in memory fidelity, but shifts in the way memories are represented. According to this view, older adults are more likely to form gist-based memory representations, capturing the essential meaning of experiences while specific details are lost (Grilli & Sheldon, 2022; Greene & Naveh-Benjamin, 2023). Importantly, these kinds of shifts in memory representations may not be well-captured by word lists.

Recent research has begun incorporating more naturalistic stimuli to bridge the gap between laboratory-based tasks and real-world memory. However, studies using naturalistic approaches have produced mixed findings. One study suggest that older adults' reliance on temporal organization is reduced when recalling events embedded in richer contexts, such as a museum tour (Diamond & Levine, 2020). Others have found that when events follow strong causal or narrative structures, older adults perform similarly to younger adults in maintaining the temporal structure of everyday event scripts (e.g., familiar routines; Schank & Abelson, 1977; Rosen et al., 2003) or linear short films (Fenerci et al., 2024). These mixed findings suggest that the extent of age-related changes in recall organization may depend on the complexity of stimulus and the memory strategies available. Critically, these findings also highlight the importance to study memory organization in contexts where multiple structures including temporal, semantic, and causal connections are simultaneously available and may exert important influences.

Thus, a key question remains: Do older and younger adults differ in how they structure memories for complex, naturalistic events that are separated in time but connected by meaning? To address this question, we conducted an experiment using a movie stimulus that allows for multiple recall organization strategies, including temporal associations, semantic relationships, topic-based clustering, and causally connected storylines. Older and younger adults watched a 26-minute episode of a television sitcom featuring four central storylines interleaved with one another throughout the episode. The episode's nonlinear structure provided a unique opportunity to examine how recall is organized based on different principles, including meaning and causality.

After viewing, participants completed a free recall task, allowing us to assess age-related differences in recall organization for naturalistic event memories. Specifically, we examined whether older adults show similar reduced reliance on temporal recall clustering and increased reliance on meaningful organization typically observed in list-learning paradigms. We also assessed the role of causal event structure in event recall performance and explored whether neuropsychological measures predict individual differences in recall strategies.

Our findings reveal a clear age-related shift in recall organization, with older adults prioritizing broad, meaningful connections over strict temporal order to a greater extent than younger adults. Additionally, our results highlight the critical role of causal centrality in recall, with high-centrality events significantly more likely to be remembered. Finally, we found that standard neuropsychological test scores were not significantly associated with recall performance, temporal clustering, or topic clustering scores, highlighting that these neuropsychological assessments may not be suited to capture age-related changes that influence specific organizational strategies used during recall, particularly in more naturalistic memory tasks.

METHODS

Participants

Eighty-two participants were recruited to participated in the study: 36 older adults (M = 72.22, SD = 6.93; range= 60-93; 20 female) and 46 younger adults (M = 19.96, SD = 1.58; range=18-25; 34 female). Younger adults were recruited from a pool of undergraduate students enrolled in psychology courses at the University of California, Davis and Washington University in St. Louis. Older adults were recruited from the Davis, CA community through online advertisement, flyers and word of mouth. Older adults were recruited from the St. Louis, MO community through Volunteer for Health Participant Registry. Data collection was initially conducted in the Davis community but was disrupted due to the COVID-19 pandemic. Recruitment later resumed in the St. Louis community, allowing us to reach our target sample size. Inclusion criteria included: Normal or corrected-to-normal hearing, normal or correct-to-normal vision, no history of major neurological or psychiatric illness, and English as a native language.

Procedure

Encoding. Older and younger participants watched a twenty-six-minute episode of a television "slice-of-life" sitcom (HBO's Curb Your Enthusiasm, S01E07 AAMCO), involving storylines of different topics that were interleaved with one another throughout the episode (Fig. 1a). Four central storylines were identified: Caterer, Chevy, Dinner Party, and Other (see *Recall Scoring Approach* below). After encoding, participants completed a free recall task.

Free Recall. In the free recall task, participants were presented with a fixation point in the center of a computer screen, and were instructed to verbally recall everything that they can remember about the episode (Fig. 1b). While participants were told that they could recall the events in any order, it was advised that it may be easier to recall it in the order of how it was encoded. Participants were encouraged to continue recalling as long as they remember additional details.

Neuropsychological Tests. Prior to encoding, older participants also completed a battery of neuropsychological tests to test for cognitive impairments that might be relevant for assessing task performance (Table 1). These tests included: Craft21 Recall Immediate, Craft21 Recall Delayed, Montreal Cognitive Assessment (MoCA), and Multilingual Naming Test (MINT; see Table 1). Briefly, Craft21 assesses recall for narratives, MoCA coarsely assesses cognitive ability, and MINT assesses the ability to name objects in English.

Analyses

Data and Code Availability. Materials, primary data, and analysis scripts are publicly available (https://osf.io/9pfmk/; https://github.com/aidelarazan/curbage_recall). No aspects of the experiment was preregistered.

Recall Scoring Approach. To investigate age-related differences in the way participants organize recall for complex events, similar analyses as free recall list-learning paradigms were applied. Annotations about the television episode were created by three independent annotators (A.I.D., E.M., and J.D.). Annotators identified events based on meaningful transitions (Zacks et al., 2007) between topics: Caterer, Chevy, Dinner Party, and Other, resulting in 37 total events (10 Caterer, 10 Chevy, 12 Dinner Party, and 5 Other). Annotators then provided written descriptions about the events and agreed on the final descriptions that were used as the “ground truth” to compare participants’ recall to.

Two raters (A.I.D and K.M.) independently segmented participants’ recall and labeled each segment according to the event and topic it pertained to using the annotations (Fig. 2c). Following prior approaches used to apply list-learning analyses to naturalistic stimuli (Diamond & Levine, 2020; Heusser et al., 2021; Delarazan et al. 2025), each segment was assigned its ordinal position from the original presentation. From this, we derived a vector of order tags, representing the events recalled by the participant and the sequence in which they were recalled.

Analyses applying free recall list-learning methods (i.e., lag-CRPs, temporal clustering scores, serial position curves, and topic clustering scores) were conducted using the psifr package in Python (Morton 2020). Additionally, given that our sample involved two different cohorts (i.e., samples collected from both the Davis and St. Louis communities), the following mixed-effect linear model was consistently used across our measures: *Dependent Variable (e.g., Recall Performance) ~ Age Group + (1| Cohort)*. The models included age group as a fixed effect and cohort as a random effect, allowing us to estimate the overall effect of age group on recall performance while controlling for differences attributable to cohort effects. Our mixed-effects linear model was analyzed using statsmodel package in Python.

Recall Performance. Recall performance was operationalized as the total number of words recalled, as has been done in prior studies (Flores et al., 2017; Delarazan et al. 2025). Additionally, a separate analysis on a subset of the recall data (N=41) revealed that word count highly correlated with manual scoring of recall (Pearson $r(41)=0.969$, CI=[0.92, 0.98], $p<.001$; Table 1). The manual scoring method was adapted from the Autobiographical Memory Interview (Levine et al., 2002): Each recall transcript was segmented into meaningful detail units and then assigned to labels that describe the content. Two raters scored the number of verifiable details that each participant recalled for each event. Intraclass correlation coefficient (ICC) was calculated to determine consistency among the raters and revealed high inter-rater reliability (Single raters absolute: ICC = 0.83, $F(40, 41) = 11.18$, $p < .001$, 95% CI = [0.71, 0.91]; Single random raters: ICC 4 = 0.84, $F(40, 40) = 28.09$, $p < .001$, 95% CI = [0.13, 0.95]; Average raters absolute: ICC = 0.91, $F(40, 41)$

= 11.18, $p < .001$, 95% CI = [0.83, 0.95]; and Average random raters: ICC = 0.91, $F(40, 40)$ = 28.09, $p < .001$, 95% CI = [0.24, 0.98]; Table 2). Recall performance was then analyzed using a mixed effects linear model: *Recall performance ~ Age Group + (1|Cohort)*.

Lag-CRP. To quantify temporal organization, we analyzed lag-conditional response probability (lag-CRP; Kahana, 1996), a standard behavioral recall measure of temporal context reinstatement. The lag-CRP curve quantifies the probability of recalling an event based on its relative encoding position (lag) to the previously recalled event (Kahana, 1996). A lag of +1 indicates that the recalled event was encoded immediately after the prior recalled event, whereas a lag of -2 means it was encoded two events earlier. To compute the lag-CRP, we determined the lag for each event transition in recall, normalizing by the total number of possible transitions. This produced a participant-by-lag matrix, where lags ranged from -37 to +37 (excluding 0), resulting in 74 possible lags. Finally, we averaged across participants and age groups to generate two separate lag-CRP curves for each age group. The lag-CRP curves were then used to derive overall measures of temporal clustering and forward asymmetry, which were then compared across age groups.

Forward Asymmetry. Forward asymmetry indicates the likelihood that participants follow the original timeline of the encoded episode during recall. It focuses on forward recall transitions (+1 lag), with higher proportions demonstrating a greater tendency to recall the next encoded event to the just recalled event. Using a similar mixed-effects linear model as recall performance: *Forward asymmetry ~ Age Group + (1|Cohort)*, we compared the proportion of all transitions that moved forward in time with respect to the encoded order among older and younger adults.

Temporal Clustering Scores. Forward asymmetry looks primarily at the next event recalled, but we can also evaluate one’s tendency to successively recall nearby events (regardless of forward or backward direction transitions). Temporal clustering scores quantifies, for each transition between recalled events, the extent to which a person's event transitions are close versus far in terms of the original order. We sorted all not-yet-recalled event transition according to absolute lag. We then computed the percentile rank of the net event the participant recalled and then averaged the percentile ranks across participants within the same age group. Recalling event closest to the just recalled event during encoding would yield a score of 1 and recalling the event farthest to just recalled event during encoding would yield a score of 0. Recalling the events in random order would yield a score of 0.5, corresponding to chance clustering. Clustering above chance is evidence of a temporal contiguity effect. We compared the extent to which age groups tended to recall events based on temporal context using a similar mixed-effects linear model as before: *Temporal Clustering Scores ~ Age Group + (1|Cohort).*

Topic Clustering Scores. While analyses involving lag-CRP, forward asymmetry, and temporal clustering scores focus on recall based on temporal context, recall organization can also be driven by how related information are. In the context of list-learning tasks, these could be recalling items of the same taxonomic category (e.g., animals, occupation, vegetables) together even when presented randomly. Applying similar logic as list-learning tasks to topic categories, we calculated an adjusted ratio of clustering (ARC) score, which estimates the probability of successively recalling items or events of the same category topic (Roenker et al., 1971). ARC scores have a maximum of 1, with higher values indicating increasingly above-chance clustering (score of 0). We compared the extent to which age groups tended to recall events based on topics using a linear mixed-effects model: *Topic Clustering Scores ~ Age Group + (1|Cohort).*

Temporal Organization within Topics. Prior research suggests that older adults may rely more on semantic or higher-order structures when recalling information, potentially at the expense of precise temporal order. However, it remains unclear whether older adults maintain temporal organization within meaningful clusters, such as shared topics, or whether their recall reflects a broader shift away from temporal sequencing.

To determine whether event recall remains temporally structured within topics, we calculated forward asymmetry and temporal clustering scores separately for each topic using the same approach as before. These scores were then averaged across topics, resulting in a single temporal clustering and forward asymmetry value per participant. We then compared these values across age groups. To assess differences in within-topic temporal organization between age groups, we used a linear mixed-effects model: *Average Forward Asymmetry within Topics* or *Average Temporal Clustering Score within Topics ~ Age Group + (1|Cohort).* Due to the relatively limited number of recallable events and low frequency of observations for the *Other* topic, we only conducted these analyses for the topics involving the *Caterer*, *Chevy, and Dinner Party*.

Centrality Networks. Recall organization can also be influenced by how causally connected the events are within the episode (Mandler, 2014; Lee & Chen, 2022). We assessed the degree to which semantic relationships and causal connections between events contribute to how well they are remembered across age groups using narrative networks. Narrative networks assess the centrality of each event with higher centrality events forming stronger and greater number of connections with other events. We created separate centrality networks for semantic relationships and causal connections and events were categorized as low and high centrality events. The following linear mixed-effects model was used to test the influence of centrality, separately for causal and semantic centrality: *Recall Proportion ~ Age Group + Centrality + (Age Group * Centrality) + (1|Cohort)*. The model includes age group, event centrality, and their interaction as fixed effects and cohort as a random effect.

Semantic Centrality Networks. Within a narrative network, events are represented by nodes and connections between events are represented with edges. We created a semantic narrative network based on similarities between events using Google's Universal Sentence Encoder (USE; Cer et al., 2018), a natural language processing model that converts inputted texts into 512-dimensional embedding vectors. USE uses pre-weighted layers, previously trained on an expansive textual database and calculates a cosine similarity between pairs of inputted text. Annotations were used as inputs, computing the semantic similarity between all event pairs. Node size is derived from the number of connections an event has with other events (thresholded at cosine similarity = 0.6) and edge thickness are proportional to edge weights. Individual events were then categorized as low and high semantic centrality based on the number and strength of connections the event had with other events (Lee & Chen, 2022).

Causal Narrative Networks. We also created a causal narrative network by gathering causality ratings from a separate group of raters (N = 25, M = 20.40, SD = 2.03; range= 18-26; 12 female). Raters were recruited from a pool of undergraduate students enrolled in psychology courses at Washington University in St. Louis and followed the same inclusion criteria as our original sample. The raters watched the same episode and identified causally related event pairs (e.g., cause event and effect event) based on the annotations. The ratings were then transformed into a causal narrative network with the size of each node derived from the total number of connections to other events. That is, the total number of times raters identified an event as a cause or an effect event. The thickness of the lines represents the causal strength between paired events and is operationalized by the total number of raters identified cause and effect relationships between paired events. Individual events were then ranked ordered, and a median split half was applied to categorize events as low or high causal centrality.

RESULTS

As previously stated, our analyses applied free recall list-learning methods to assess age-related differences in recall organization using a naturalistic stimulus. Since our sample included two distinct cohorts (participants from the Davis and St. Louis communities), we accounted for potential cohort differences using a linear mixed-effects model: *Dependent Variable (e.g., Recall Performance) ~ Age Group + (1| Cohort)*. In this model, age group was included as a fixed effect, while cohort was included as a random effect to control for variability attributable to sample differences. This approach allowed us to estimate the overall effect of age group on recall performance while accounting for potential cohort-related differences. With the exception of overall recall performance, the variance attributed to cohort in our models was negligible (cohort variance < 0), suggesting minimal cohort-related variability in our measures (e.g., forward asymmetry, temporal clustering scores, topic clustering scores, etc.) beyond the fixed effect of age group.

Younger and older adults do not differ in overall recall performance

Analyses examining the effect of age group on recall performance, operationalized by the number of words recalled, revealed no meaningful differences in recall performance between younger and older adults ($\beta = 1.15$, SE = 2.27, $z = 0.505$, 95% CI = [-3.31, 5.60], $p = 0.613$; Fig. 2). The random effect of cohort suggests that differences in recall contributed to variability in performance (cohort variance = 151.15, SE = 3.94). However, the inclusion of this factor in the model accounts for potential influences on our comparisons of interest. Together, results indicate that while overall recall performance differed by cohort, there were no difference across age groups, even when accounting for this difference.

Age-related shifts from temporal to topic organization

Effective memory retrieval relies on organizational principles that structure recall transitions. We examined lag-conditional response probability, forward asymmetry, and clustering scores, allowing us to quantify the extent to which memory retrieval was driven by temporal context versus meaningful connections across age groups.

Lag-CRP and Forward Asymmetry. Temporal structure is crucial for recall, with individuals typically retrieving events in the order they were experienced. To assess age-related differences in temporal recall, we calculated lag conditional response probability (lag-CRP), which measures the likelihood of recalling events based on their temporal proximity. Additionally, we examined forward asymmetry, which builds on lag-CRP by focusing specifically on +1 lag to determine the tendency to recall events in their exact sequential order.

Recall transitions between events revealed patterns reflecting reliance on temporal structure, with older and younger generally showing higher lag-CRP at +1 compared to other positions. However, we observed a significant age difference in the degree of forward asymmetry ($\beta = 0.131$, $SE = 0.031$, $z = 4.203$, 95% CI = [0.070, 0.192], $p < 0.001$; Fig. 3a), indicating that younger adults were more likely to recall events in their originally-encoded order compared to older adults.

Temporal Clustering Scores. Building on these patterns, we calculated temporal clustering scores to get an assessment of overall likelihood to recall nearby events, regardless of forward or backward direction transitions. We found weaker temporal clustering among older adults relative to younger adults ($\beta = 0.073$, $SE = 0.015$, $z = 4.900$, 95% CI = [0.044, 0.102], $p < 0.001$; Fig. 3b), which coincides with our findings involving age-related differences and forward asymmetry above.

Topic Clustering Scores. Given that older adults exhibited less reliance on temporal structure during recall, we next investigated whether older adults relied on alternative organizing principles. Specifically, we examined whether older adults demonstrated a greater tendency to structure recall based on shared topics rather than temporal proximity. That is, given the recall of an event related to a specific topic (e.g., an event involving a caterer), we estimated the likelihood that the next recalled event also pertained to the same topic. Older adults were more likely than younger adults to structure recall based on topic, often transitioning between non-adjacent events within the story to maintain topic continuity ($\beta = -0.085$, SE = 0.037, $z = -2.289$, 95% CI = [-0.157, -0.012], $p = 0.022$; Fig. 3c).

These findings suggest that while younger adults rely more on temporal structure to guide recall, older adults shift toward organizing memories based on topic-level connections. This shift in recall strategy highlights broader age-related changes in memory retrieval, potentially reflecting a transition from precise structures to those based on meaningful connections.

Temporal Organization within Topics. Our findings demonstrate heightened reliance to topic-level organization among older adults, potentially reflecting deficits in veridical memory. That is, older adults may have greater difficulty precisely reconstructing the sequence of events for the entire television episode and thus show a general deficit in temporal memory organization. Alternatively, older adults may still show strong temporal organization within shared topics, potentially reflecting strategic shifts in recall organization instead. To further investigate these age-related changes, we investigated temporal organization within each topic through forward asymmetry and temporal clustering scores.

After assuming topic-level organization, results show that younger adults have a greater tendency to recall the exact temporal sequence of events compared to older adults ($\beta = 0.085$, SE

= 0.034, z = 2.539, 95% CI = [0.019, 0.214], p = 0.011; Fig. 4a). Interestingly, we found no age-related differences in temporal clustering scores after accounting for within-topic organization (β = 0.029, SE = 0.017, z = 1.738, 95% CI = [-0.004, 0.061], p = 0.083; Fig. 4e). It is important to note that age-related differences varied across topics for both forward asymmetry (Fig. 4b-4d; Caterer: β = -0.026, SE = 0.031, z = -0.578, 95% CI = [-0.115, 0.063], p = 0.563; Chevy: β = 0.115, SE = 0.050, z = 2.305, 95% CI = [0.017, 0.214], p = 0.021; Dinner Party: β = 0.170, SE = 0.058, z = 2.917, 95% CI = [0.056, 0.284], p = 0.004) and temporal clustering scores (Fig. 4f-4h; Caterer: β = 0.007, SE = 0.028, z = 0.237, 95% CI = [-0.049, 0.062], p = 0.813; Chevy: β = 0.022, SE = 0.022, z = 1.019, 95% CI = [-0.020, 0.064, p = 0.308]; Dinner Party: β = 0.058, SE = 0.018, z = 3.211, 95% CI = [0.022, 0.093], p = 0.001), indicating that recall organization was partly influenced by the stimulus. Since we had no prior hypotheses about topic-specific differences, we averaged scores to assess overall temporal organization within topics. Therefore, our findings involving the averaged scores suggest that narrative event structures may be informative for older adults to use, but do not have an exact representation of the order in which it has happened. That is, older adults have a coarse representation of the way in which the events unfold but these representations are more gist-based accounts of the events.

Altogether, these findings indicate that while both age groups relied on temporal structure in recall transitions for the entire episode, younger adults showed significantly greater forward asymmetry and temporal clustering compared to older adults. On the other hand, older adults showed higher topic clustering scores, suggesting a tendency to deviate from the original timeline to connect events that shared topics when recalling the narrative. Additional analyses of temporal organization within topics revealed age-related differences in forward asymmetry, but not temporal clustering scores suggesting an age-related shift in recall strategies. Older adults rely more on coarse, meaningful connections and less on the exact temporal sequence of events when retrieving memories of dynamic events compared to younger adults.

Temporal, but not topic, organization aids overall recall performance

Temporal Clustering Scores and Recall Performance. While younger adults relied more on temporal structure, and older adults showed a greater tendency to organize recall by topic, it remains unclear whether these strategies enhance or hinder overall recall ability. To address this, we next examined the relationship between clustering scores and recall performance for the entire episode. We also compared the strength of correlations between temporal clustering and recall performance as well as topic clustering and recall performance across younger and older adults to examine whether the relationship between clustering scores and recall performance differed by age group.

We correlated temporal clustering scores with recall performance, and found that both younger ($r(46) = 0.36$, CI = [0.08, 0.59], $p = .013$) and older ($r(38) = 0.33$, CI = [-0.06, 0.54], $p = .045$) adults benefited from following the temporal structure of the original episode (Fig. 3d). However, a direct comparison of these correlation coefficients revealed that the strength of this relationship did not significantly differ between younger and older adults ($z = 0.15$, $p = .44$), suggesting that both groups benefited similarly from organizing recall based on temporal structure.

Topic Clustering Scores and Recall Performance. Conversely, we found a significant negative correlation between topic clustering and recall performance in younger adults ($r(46) = -0.34$, 95% CI = [-0.57, -0.05], $p = .042$), indicating that a stronger reliance on topic-based recall transitions was associated with poorer recall performance. Older adults showed a non-significant negative relationship in the same direction ($r(38) = -0.23$, 95% CI = [-0.51, 0.1], $p = .174$). A direct statistical comparison revealed no significant difference in the strength of these correlations between younger and older adults ($z = -0.53$, $p = .299$), suggesting that the observed negative relationship between topic clustering and recall performance was comparable across age groups, and was numerically stronger in younger adults (Fig. 3e).

Together, results indicate that temporal clustering was beneficial for recall performance in both younger and older adults, with no significant differences in effect size between groups. In contrast, while younger adults showed a significant negative relationship between topic clustering and recall performance, this relationship was not significantly different from that of older adults. This suggests that relying on topic-based recall organization was similarly not beneficial for memory performance in both age groups and could perhaps be detrimental to overall recall success.

Contributions of narrative centrality in event recall performance

Beyond temporal and topic-based organization, recall structure may also be influenced by the centrality of events within a narrative. Specifically, events that serve as key semantic and causal anchors within an episode may be more likely to be remembered. To examine this, we created narrative networks to assess the degree to which semantic relationships and causal connections between events contribute to recall performance across age groups.

Semantic Narrative Centrality and Event Recall Performance. Using the Universal Sentence Encoder (Cer et al. 2018), we computed cosine similarity between paired events with higher cosine similarity indicating greater semantic relatedness. Events (thresholded at cosine similarity > 0.6) were then rank-ordered and categorized as low and high semantic centrality events (Fig. 5a). A linear mixed-effects model was used to assess the age group, semantic centrality, and their interaction on recall event probability, while accounting for cohort effects as a random effect.

Results revealed a significant main effect of age group, with younger adults recalling significantly more events overall than older adults ($\beta = 0.140$, $SE = 0.035$, $z = 3.970$, 95% CI =

[0.071, 0.209], $p < .001$). However, semantic centrality did not significantly predict recall performance, as the main effect of low-centrality events was not statistically significant ($\beta = 0.017$, SE = 0.037, z = 0.460, p = 0.646, 95% CI = [-0.056, 0.089]). The interaction between age group and semantic centrality was also not significant ($\beta = -0.053$, SE = 0.052, z = -1.060, p = 0.289, 95% CI = [-0.155, 0.045]), indicating that the relationship between semantic centrality and recall did not differ significantly between younger and older adults (Fig. 5b-5c).

These findings suggest that age-related differences in recall performance are not driven by semantic centrality, as events with stronger semantic associations were not more likely to be recalled than those with weaker associations. Instead, younger adults demonstrated better recall overall for events that share semantic associations greater than 0.6 cosine similarity, independent of event centrality, while older adults did not show a differential reliance on high-centrality events.

Causal Narrative Centrality and Event Recall Performance. While semantic centrality did not significantly predict recall performance, event causality may provide an alternative organizational structure that supports memory retrieval. As previously stated, we gathered causality ratings from a separate group of raters, where the raters watched the same episode and identified causally related event pairs (e.g., cause event and effect event). The ratings were then assessed in terms of causality centrality by the number of connections to other events. Events were then rank-ordered and categorized as low and high causal centrality events (Fig. 6a). A similar mixed-effects linear model as semantic narrative centrality, but with causal narrative centrality.

Results revealed a significant main effect of causal centrality, with low-centrality events being recalled significantly less often than high-centrality events ($\beta = -0.319$, SE = 0.034, z = -9.285, 95% CI = [-0.386, -0.253], $p < .001$). Additionally, age group significantly predicted recall performance, with younger adults recalling more events than older adults ($\beta = 0.074$, SE = 0.024,

$z = 3.045$, 95% CI = [0.026, 0.121], $p = .002$). However, the interaction between age group and causal centrality was not significant ($\beta = 0.075$, SE = 0.046, $z = 1.614$, 95% CI = [-0.016, 0.166], $p = .106$), suggesting that the relationship between causal centrality and recall performance did not differ significantly between younger and older adults (Fig. 6b-6c).

These findings indicate that causal centrality plays a critical role in the probability of recalling specific events, with high-centrality events being significantly more likely to be remembered. In contrast to semantic centrality, which did not predict recall, these results suggest that memory retrieval is more strongly influenced by causal event structure, rather than purely semantic relationships between events. Furthermore, the benefit of causality is shared across age groups.

Neuropsychological tests did not correlate with recall performance or clustering scores

As a final step, we examined whether broader cognitive abilities, as measured by neuropsychological tests: Craft21 Recall Immediate, Craft21 Recall Delayed, Montreal Cognitive Assessment (MoCA), and Multilingual Naming Test (MINT) were related to recall performance or clustering scores. We correlated neuropsychological test scores with overall recall performance, temporal clustering scores, and topic clustering scores.

None of the neuropsychological test scores were significantly associated with recall performance, temporal clustering, or topic clustering scores (see Table 1). While there were trend-level associations between Craft 21 (Immediate and Delayed) recall and overall recall performance, these relationships did not reach statistical significance. These findings suggest that traditional neuropsychological measures may not strongly predict the specific memory retrieval patterns observed in this study.

DISCUSSION

Real-world experiences are inherently complex, leading to variability in how associated memories are represented and recalled (Palombo et al., 2018; Sheldon et al., 2019). This complexity raises a key question: Does aging alter the way past experiences are organized and remembered? To investigate this, we examined age-related differences in recall organization using a naturalistic stimulus. Older and younger participants viewed an episode of a television sitcom with interwoven topics, and then freely recalled the episode. We found no overall difference in the amount of recalled information across age groups; however, age groups differed in the way recall was organized from event-to-event. Younger adults showed strong temporal organization, while older adults clustered events by topic at the expense of temporal order. Although recall differences in the exact temporal order of events were still apparent within topics, temporal clustering effects were no longer statistically significant across age-groups after accounting for topic-level organization. Temporal organization was related to better recall across age groups, whereas topic-level organization was related to worse recall. Finally, we found that both age groups showed a strong reliance on causality in the narrative, with causally central events being more recalled than events with relatively few causal connections.

A major motivation of the present study was to assess phenomena from traditional list-learning tasks in a naturalistic stimulus, with extended lifelike events. While prior studies have shown that temporal organization of list recall is diminished in older adults (Estes, 1955; Kahana et al., 1996; Howard & Kahana, 2002), and that older adults may rely more on semantic connections (Healey & Kahana, 2016), it is possible that prior age-related differences merely reflect the arbitrary nature of word lists. On the other hand, older adults may show no apparent deficit when encoding and recalling a meaningful narrative with connected topics, much like one encounters in everyday life. However, we found evidence consistent with prior findings: When recalling events, older adults show evidence for diminished temporal organization. Furthermore, older adults showed a shift toward higher-order topic organization.

Use of a naturalistic stimulus offered important insights beyond this confirmatory evidence. We found that temporal organization of events during recall was positively correlated with overall recall success, whereas the opposite relationship was found for topic-level organization. This implies that topic-level organization in older adults is likely not compensatory, as it relates to poorer recall overall; rather, this age-related shift may be strategic rather than an attempt to mitigate memory deficits. That is, while we suggested to all participants that recalling the episode in the order it was experienced is an effective strategy, older adults recalled the episode out of order to link events that shared the same topic. Older adults may have difficulty suppressing associations that share topics, leading them to recall related events together even when doing so disrupts the original narrative sequence. These results run somewhat counter to previous findings demonstrating that older adults tend to have difficulty inhibiting irrelevant information that occurred close in temporal proximity (Hasher et al., 2001; Hasher et al., 1991). However, prior studies often use simpler stimuli (e.g., visual or word distractors) rather than events that are meaningfully connected across time (Cohn-Sheehy et al., 2022; Delarazan et al., 2025). This suggests that the presence of meaningful event structures may play a crucial role in how older adults organize and retrieve memories.

To better understand these age-related differences in recall organization, we also examined how temporal structure was maintained within individual topics. It is possible that older adults tend to organize recall by topic more broadly, but show no evidence for a temporal organization deficit within topics. In contrast, our findings revealed that, even within a given topic, older adults tended to reconstruct the sequence of events less precisely compared to younger adults. This may reflect a broader shift toward gist-based memory representations, where general event structures are retained while fine-grained details, such as sequential order, become less emphasized (Gilboa & Moscovitch, 2021; Grilli & Sheldon, 2022). That is, while narrative structures to may be helpful in guiding recall, older adults do not necessarily preserve the fine-grained order of events within those topics. Our findings align with previous research on verbatim versus gist-based memory (Brainerd & Reyna, 2004; Reyna & Brainerd, 2015) and extend this work by examining recall using a naturalistic stimulus with interwoven topics over an extended timescale.

Recent studies have increasingly turned to naturalistic approaches to studying memory in older adults. As previously noted, these studies have sometimes revealed conflicting results. Whereas one recent study involving a movie stimulus found no major differences between older and younger adults (Fenerci et al., 2024), a study involving recall of a real-world museum tour showed diminished temporal organization in older adults' recall (Diamond & Levine, 2020). A primary difference between these studies may lie in the complexity of the stimuli. The movie used in the former study was intentionally pared-down, and involved events that were causality connected that unfolded in a linear timeline. In contrast, the museum tour involved item-based recall of artwork that lacked causal event structure. Given the recall boost afforded by causally connected events, these differences suggest that a stimulus that is optimal for testing age-related differences in memory may benefit from a mixture of high causality and relatively arbitrary low causality events with topics interwoven with each other, much like real life. Our study could be considered to be an intermediate step between these experiments, highlighting contributions of varying degrees of causality and multiple narrative structures on memory for everyday events.

Similar to prior studies, our findings indicate that causality is a key component of event recall (Lee & Chen, 2022), and possibly an important target in future studies of age-related memory decline.

Although this study did not examine age-related pathology, our results may have implications on current neuropsychological assessments. Previous research has shown that older adults with early-stage Alzheimer's disease exhibit reduced temporal organization effects in list-learning tasks (Quenon et al., 2015). In our study, we found no significant associations between clustering scores or overall recall performance and traditional neuropsychological test measures, suggesting that our naturalistic stimuli may engage cognitive mechanisms that differ from those assessed by standard clinical tests. That is, current neuropsychological assessments may lack sensitivity to detect subtle shifts in memory organization that emerge with age-related pathology. Examining recall transitions for events that reflect naturalistic events, such as those used in our study, may provide deeper insights into episodic memory impairments associated with age-related pathology. Given that healthy older adults show increased reliance on topic-level organization compared to younger adults, it is possible that the inability to organize based on topic-level organization could be an indicator to age-related pathology decline. Further studies should explore how recall transitions relate to early markers of Alzheimer's disease and whether narrative recall patterns could serve as a more sensitive diagnostic tool.

In summary, our findings highlight important age-related differences in recall organization when remembering naturalistic events. While both younger and older adults benefited from causal event structures, older adults showed a shift away from strict temporal order toward topic-based organization. Importantly, our results suggest that naturalistic stimuli offer valuable insights beyond traditional list-learning tasks, revealing nuanced patterns of memory organization that could inform future research on aging and cognitive decline.

TABLES AND FIGURES

Table 1. Neuropsychological test scores from older adults and correlations with recall performance, temporal clustering scores, and topic clustering scores

		Recall Performance		Temporal Clustering Score		Topic Clustering Score	
Neuropsychological Test	**Score M (SD)**	**r**	**p**	**r**	**p**	**r**	**p**
Montreal Cognitive Assessment (MoCA)	26.324 (3.11)	0.148	.384	-0.082	.630	-0.066	.698
Craft 21 Immediate	24.243 (9.488)	0.322	.052	0.152	.370	-0.053	.755
Craft 21 Delayed	22.216 (10.127)	0.315	.058	0.085	.619	-0.140	.410
Multilingual Naming Test (MINT)	30.459 (1.710)	0.013	.941	0.138	.415	-0.103	.544

Table 2. Intraclass correlation coefficient (ICC) of verifiable details.

	ICC	95% Confidence Interval	F	df1	df2	p
Single Rater	0.835	(0.71, 0.91)	11.18	40	41	.000
Average Rater	0.915	(0.83, 0.95)	11.18	40	41	.000

Table 3. Correlations between average verifiable details and recall performance based on age group.

Total (N=41)		Younger Adults (N=20)		Older Adults (N=21)	
r	**p**	**r**	**p**	**r**	**p**
0.969	< .000	0.983	< .000	0.897	< .000

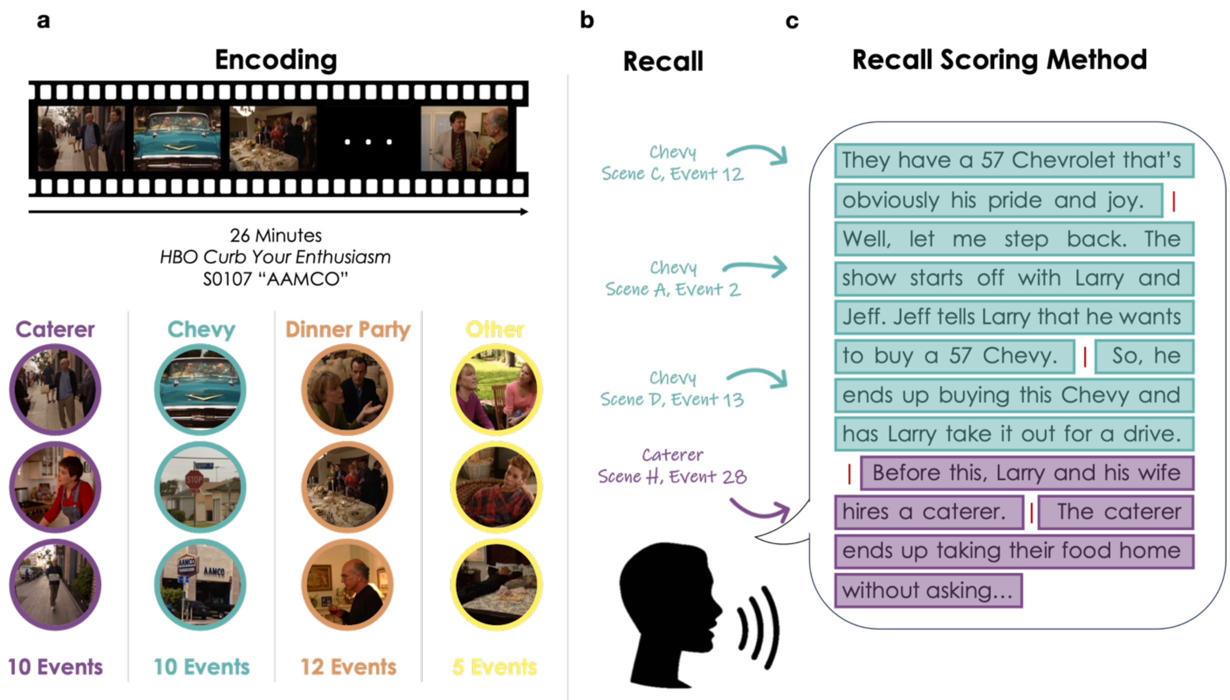

Figure 1. Methods and Materials. (a) Encoding. *Participants were presented with an episode of a television sitcom (HBO's Curb Your Enthusiasm, S01E07: AAMCO). The episode consisted of 37 events and four different topics: Chevy, Caterer, Dinner Party, and Other.* ***(b) Recall.*** *Participants were then instructed to recall the episode in as much detail as possible in any order.* ***(c) Recall Scoring Methods.*** *For each participant's recall, raters segmented and labeled each segment by the event it pertained to (e.g. Topic: Chevy, Event Number: 12, Recall Order: 1; Topic: Dinner Party; Event Number 2, Recall Order: 2).*

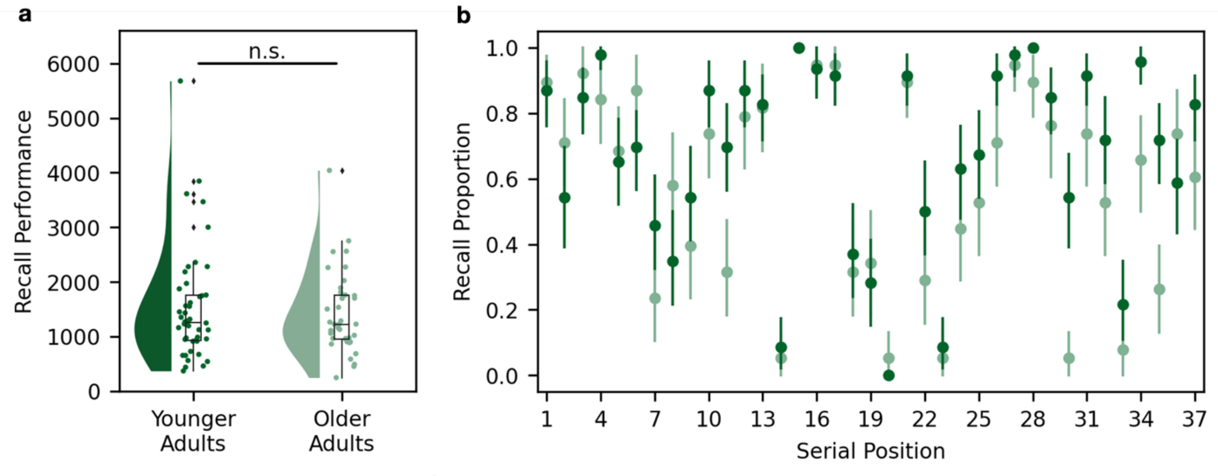

Figure 2. (a) Recall Performance. *Reflects the average amount of words recalled for each age group.* ***(b) Serial Position Curves.*** *Reflects the proportion of participants in each age group who remembered each event as a function of the event's position during encoding.* ***Key:*** *Points represent individual participants' average performance. Boxplot and violin plot display distribution of data, including the range, IQR and median.* ***Significant tests:*** *n.s. > .05, * p < .05, ** p < .01, *** p < .001*

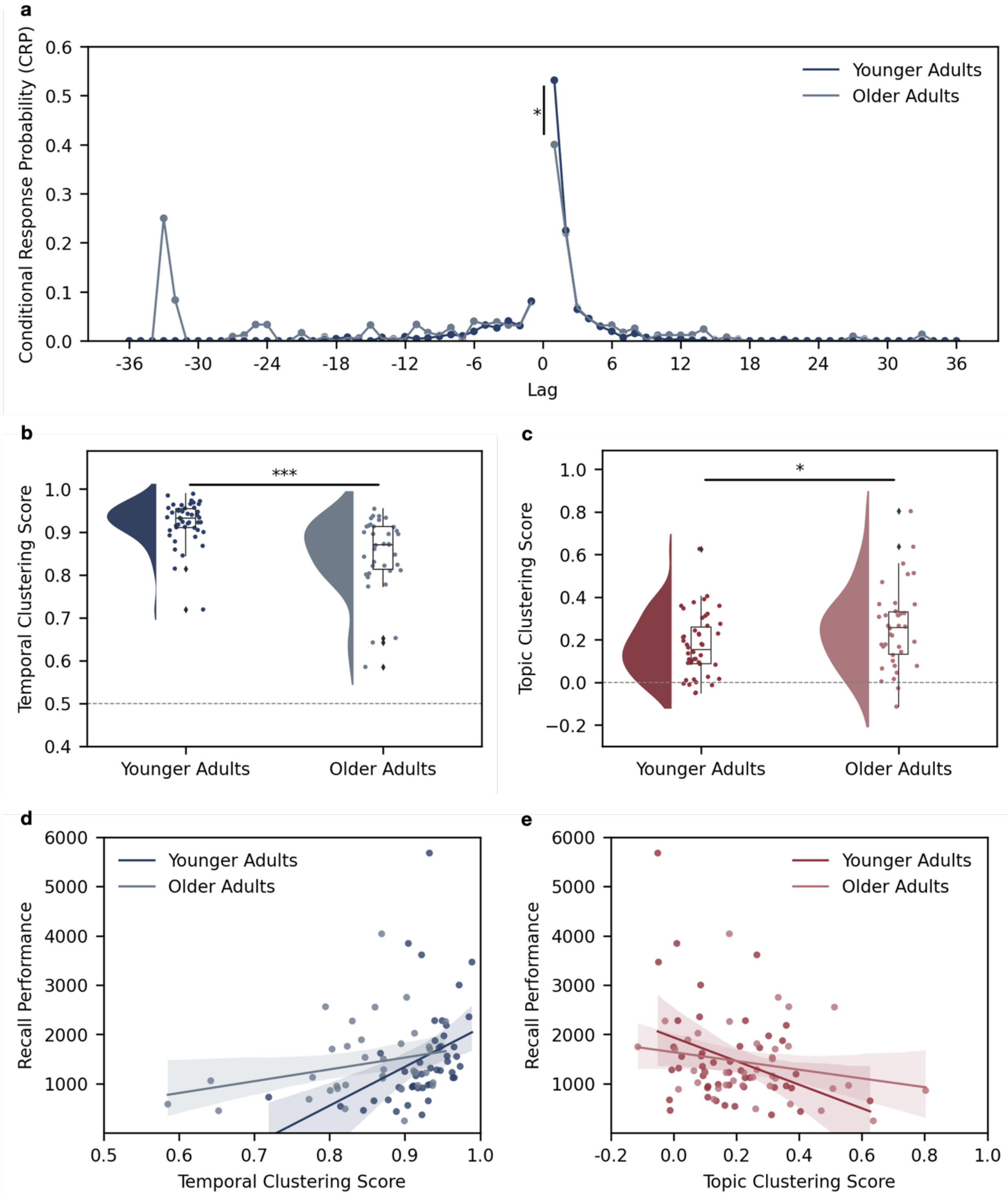

Figure 3. (a) Lag-CRP. *Reflects the probability of recalling a given item after the just-recalled item, as function of their relative encoding positions (lag).* ***(b) Temporal Clustering Scores.*** *Reflects a tendency to organize recall in relation to the original story's sequence.* ***(c) Topic Clustering Scores.*** *Reflects a tendency to organize recall based on shared topic.* ***(d)*** *Correlations between recall performance and temporal clustering scores.* ***(e)*** *Correlations between recall performance and topic clustering scores.* ***Key:*** *Points represent individual participants' average*

performance. Boxplot and violin plot display distribution of data, including the range, IQR and median. Solid lines represent line of best fit and confidence interval. ***Significant tests:*** *n.s. > .05, * p < .05, ** p < .01, *** p < .001*

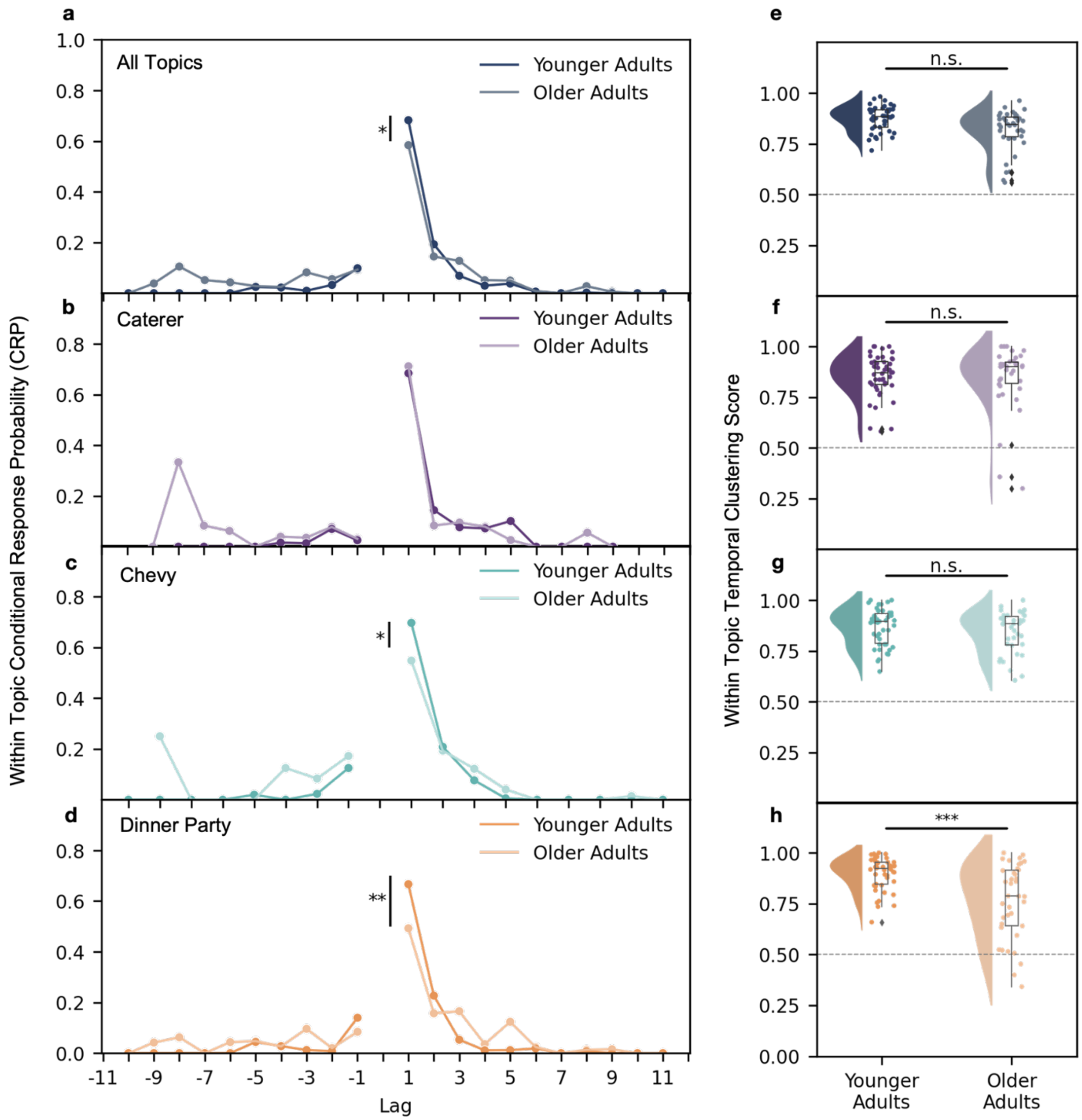

Figure 4. Within Topic Lag-CRP. *Reflects the probability of recalling a given item after the just-recalled item, as a function of their relative encoding positions (lag) for each topic:* ***(a)*** *average across topics,* ***(b)*** *Caterer,* ***(c)*** *Chevy, and* ***(d)*** *Dinner Party.* ***Within Topic Temporal Clustering Scores.*** *Reflects a tendency to organize recall in relation to the original story's sequence for each topic:* ***(e)*** *average across topics,* ***(f)*** *Caterer,* ***(g)*** *Chevy, and* ***(h)*** *Dinner Party.* ***Key:*** *Points represent individual participants' average performance. Boxplot and violin plot display*

distribution of data, including the range, IQR and median. Solid lines represent line of best fit and confidence interval. ***Significant tests:*** *n.s. > .05, * p < .05, ** p < .01, *** p < .001*

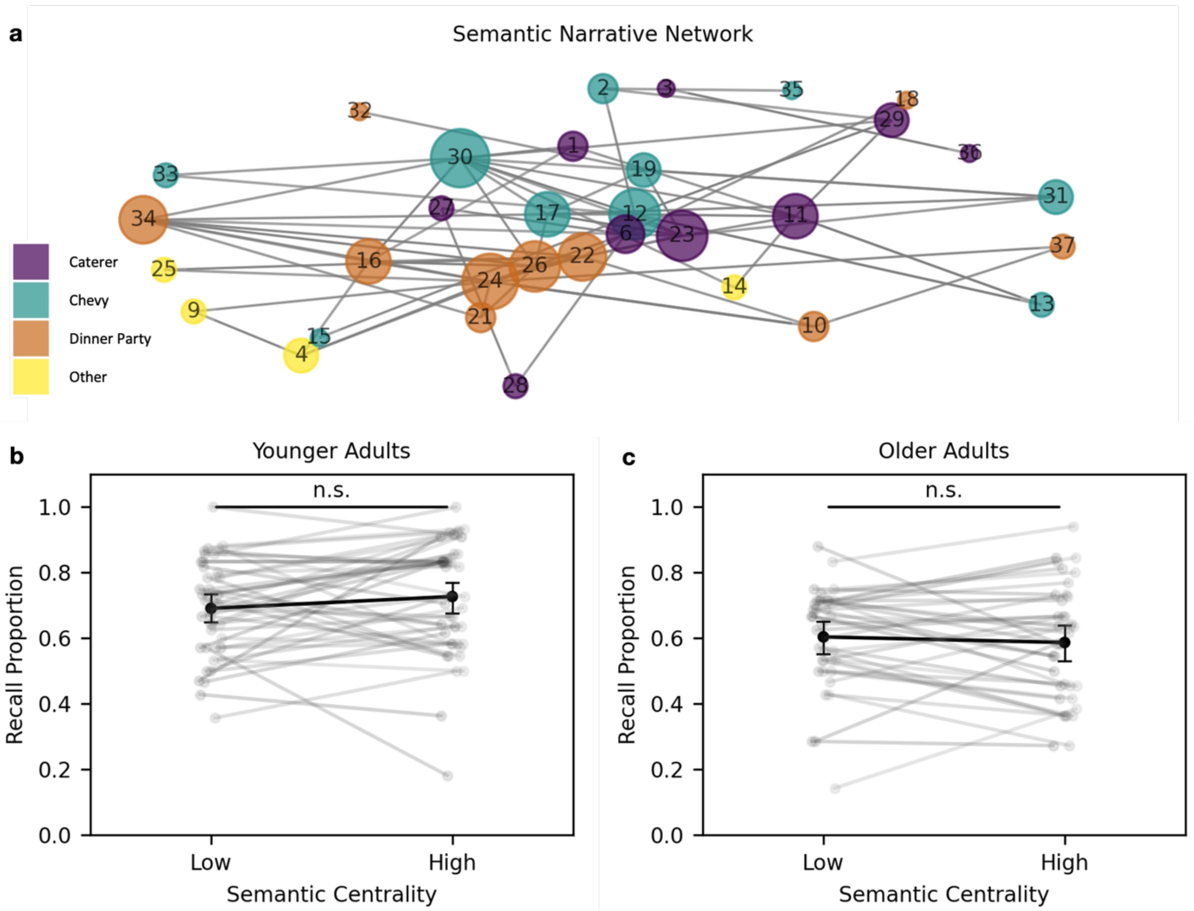

Figure 5. (a) Semantic Narrative Network. *To create semantic narrative networks, each movie was split into events, and independent annotators provided text descriptions of the events. The text descriptions were transformed into sentence embedding vectors using Google's Universal Sentence Encoder (USE). Semantic similarity between events was computed as the cosine similarity between the USE vectors. A semantic narrative network was defined as a network whose nodes are events and the edge weights are the semantic similarity between the events. Edge weights were thresholded at cosine similarity = 0.6 for visualization purposes. Node size is proportional to centrality. Nodes are color-coded based on the topic. Edge thickness is proportional to edge weights. Recall probability for Low and High semantic centrality events for* ***(b)*** *younger and* ***(c)*** *older adults.* ***Key:*** *Points represent individual participants' average performance. Solid lines represent line of best fit.* ***Significant tests:*** *n.s. > .05, * p < .05, ** p < .01, *** p < .001*

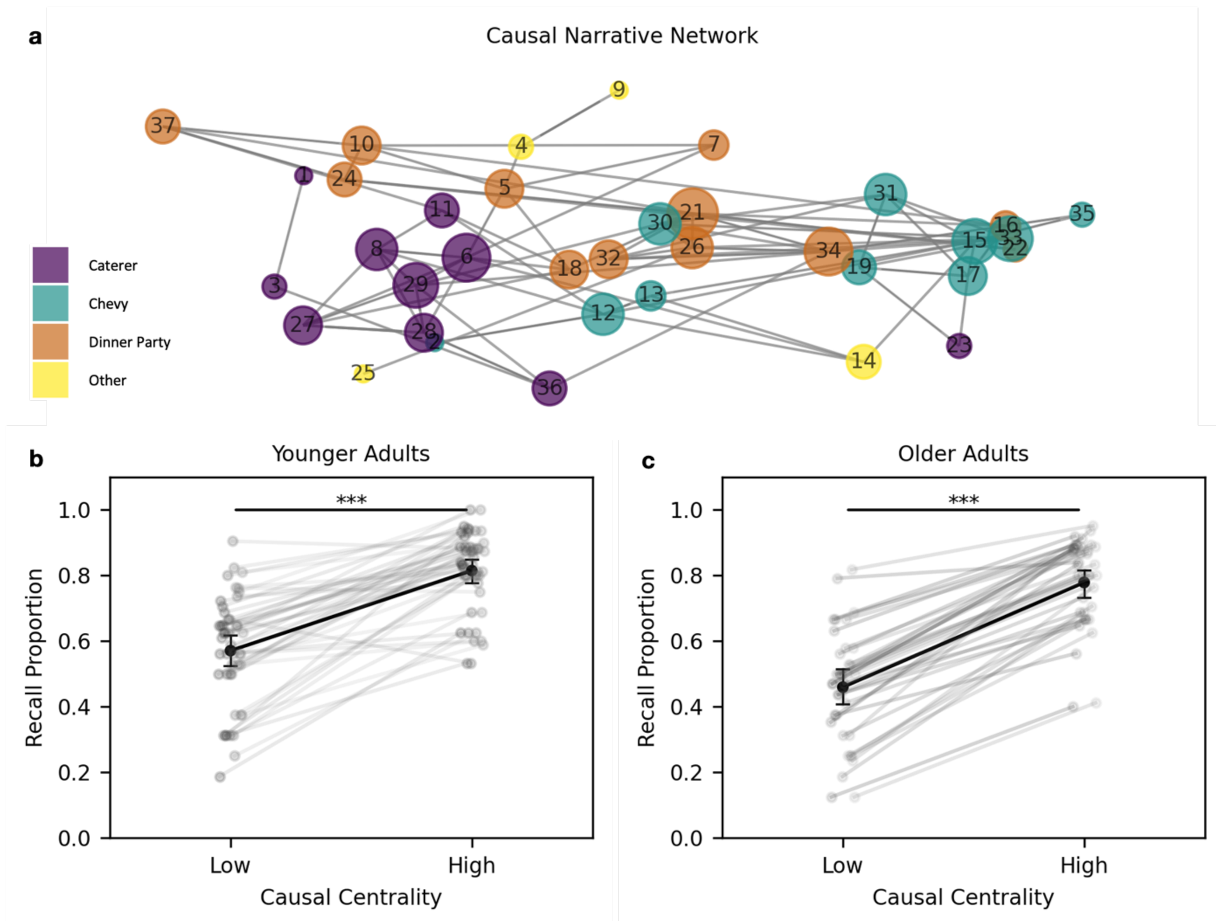

Figure 6. ***(a) Causal Narrative Network.*** *To create causal narrative networks, each movie was split into events, and independent annotators provided text descriptions of the events. Independent raters identified causal relations between paired events. A causal narrative network was defined as a network whose nodes are events and the edge weights are the causal relationships between the events. Edge weights were thresholded at cosine similarity = 0.6 for visualization purposes. Node size is proportional to centrality and calculated based on the number of connections it has with other events. Nodes are color-coded based on the topic. Edge thickness is determined by the number of times the pairs were rated as causally related. Recall probability for Low and High causal centrality events for* ***(b)*** *younger and* ***(c)*** *older adults.* ***Key:*** *Points represent individual participants' average performance. Solid lines represent line of best fit.* ***Significant tests:*** *n.s. > .05, * p < .05, ** p < .01, *** p < .001*

REFERENCES

Cer, D., Yang, Y., Kong, S. Y., Hua, N., Limtiaco, N., John, R. S., ... & Kurzweil, R. (2018, November). Universal sentence encoder for English. In *Proceedings of the 2018 conference on empirical methods in natural language processing: system demonstrations* (pp. 169-174).

Cohn-Sheehy, B. I., Delarazan, A. I., Crivelli-Decker, J. E., Reagh, Z. M., Mundada, N. S., Yonelinas, A. P., ... & Ranganath, C. (2022). Narratives bridge the divide between distant events in episodic memory. Memory & Cognition, 50(3), 478-494.

Lee, H., & Chen, J. (2022). Predicting memory from the network structure of naturalistic events. *Nature Communications*, *13*(1), 4235.

Delarazan, A. I., Bosak, E., Cohn-Sheehy, B. I., Foureaux-Lee, V., Zacks, J. M., & Reagh, Z. (2025, March 17). Narrative coherence warps the timeline of recalled naturalistic events.

Diamond, N. B., & Levine, B. (2020). Linking detail to temporal structure in naturalistic-event recall. *Psychological Science*, *31*(12), 1557-1572.

Flores, S., Bailey, H. R., Eisenberg, M. L., & Zacks, J. M. (2017). Event segmentation improves event memory up to one month later. Journal of Experimental Psychology: Learning, Memory, and Cognition, 43(8), 1183.

Fenerci, C., Davis, E. E., Henderson, S. E., Campbell, K. L., & Sheldon, S. (2024). Shift happens: Aging alters the content but not the organization of memory for complex events. *Aging, Neuropsychology, and Cognition*, 1-24.

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

Grilli, M. D., & Sheldon, S. (2022). Autobiographical event memory and aging: Older adults get the gist. Trends in Cognitive Sciences, 26(12), 1079-1089.

Golomb, J. D., Peelle, J. E., Addis, K. M., Kahana, M. J., & Wingfield, A. (2008). Effects of adult aging on utilization of temporal and semantic associations during free and serial recall. *Memory & Cognition*, *36*, 947-956.

Hasher, L., Zacks, R. T., & Rahhal, T. A. (1999). Timing, instructions, and inhibitory control: some missing factors in the age and memory debate. Gerontology, 45(6), 355-357.

Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. Psychological review, 123(1), 23.

Heusser, A. C., Fitzpatrick, P. C., & Manning, J. R. (2021). Geometric models reveal behavioural and neural signatures of transforming experiences into memories. Nature Human Behaviour, 5(7), 905-919.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of mathematical psychology*, *46*(3), 269-299.

Jacoby, L. L. (1999). Ironic effects of repetition: measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 3.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & cognition*, *24*(1), 103-109.

Mandler, J. M. (2014). Stories, scripts, and scenes: Aspects of schema theory. Psychology Press.

Naveh-Benjamin, M., Hussain, Z., Guez, J., & Bar-On, M. (2003). Adult age differences in episodic memory: further support for an associative-deficit hypothesis. Journal of Experimental Psychology: Learning, Memory, and Cognition, 29(5), 826.

Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: aging and neurocognitive scaffolding. Annual review of psychology, 60(1), 173-196.

Quenon, L., de Xivry, J. J. O., Hanseeuw, B., & Ivanoiu, A. (2015). Investigating associative learning effects in patients with prodromal Alzheimer's disease using the temporal context model. Journal of the International Neuropsychological Society, 21(9), 699-708.

Radvansky, G. A., & Dijkstra, K. (2007). Aging and situation model processing. Psychonomic Bulletin & Review, 14, 1027-1042.

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. Psychological bulletin, 76(1), 45.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. Nature reviews neuroscience, 8(9), 657-661.

Schank, R. C., & Abelson, R. P. (2013). Scripts, plans, goals, and understanding: An inquiry into human knowledge structures. Psychology press.

Stine-Morrow, E. A., Miller, L. M. S., & Hertzog, C. (2006). Aging and self-regulated language processing. *Psychological bulletin*, *132*(4), 582.