Event Boundary Identification Procedure. Fifty participants watched the *Sherlock* episode and pressed a key whenever they perceived a meaningful event transition. Timestamps were converted to frame numbers and aggregated using a Gaussian kernel density estimate to create a continuous boundary-likelihood function across the video. Local maxima that met temporal separation criteria and exceeded a minimum response threshold were identified as candidate boundaries. This procedure produced 36 peaks; one occurred too close to the end of the stimulus and was excluded, yielding a final set of 35 event boundaries used throughout the current study.

Cued recall task. Participants viewed 20 randomized 3-second clips from the Sherlock video and typed what happened next in the narrative. See Supplementary Fig. S1 for schematics of this task. Trials included within-event (n = 13) and across-event (n = 7) clips, allowing us to assess memory for content within versus across event boundaries. Within-event cues were taken from the middle of an ongoing event—moments when the situation, location, and characters remained continuous—so that the cue and the target response came from the same event. Across-event cues were taken from the final seconds of one event and thus immediately preceded a boundary leading into a new situation or goal context. In these trials, the cue and target responses came from different events. Each response was scored for accuracy (0 = incorrect, 1 = correct) using a detailed rubric describing the correct sequence of actions within the 5 seconds following the cue clip. Two trained raters independently scored all responses, and disagreements were resolved through discussion (Cohen's $\kappa > .85$). This scoring procedure ensured the robustness and reliability of the task outcomes.

Recognition memory task. Recognition memory was assessed using a 20-item, two-alternative forced-choice (2AFC) task. In each trial, participants were presented with two still images displayed on the right and left sides of the screen. The images, visually similar to each other, featured the same actor, location, and objects. One image was taken from the first 35 minutes of the *Sherlock* episode participants had watched, while the other was from an unseen part of the episode or a different episode. Participants were instructed to identify which image came from the segment they had viewed by using the arrow keys on the keyboard to select the correct image. Each trial was randomized and counterbalanced across participants to control for order effects. Performance was scored as correct or incorrect on each trial. See Supplementary Fig. S1 for a schematic of this task.

Temporal order memory task. The temporal order memory task was similar to the recognition task in that participants completed a 20-item, two-alternative forced-choice (2AFC) task. However, in this task, both images came from the video they watched, and they were instructed to select the image that occurred first in the video. The order of the trials was randomized and counterbalanced across participants. Performance was scored as correct or incorrect on each trial. See Supplementary Fig. S1 for a schematic of this task.

The cued recall, recognition and temporal order memory tasks included only a subset of the 35 events. Because event-level analysis could not be conducted for these measures, they were not included as primary EEG outcomes in the present manuscript. Descriptive statistics for these measures are reported in Supplementary Table S3 and correlations among them are reported in Supplementary Table S4. Consequently, EEG–memory analyses focus exclusively on free recall performance, for which participants' responses were scored at the level of individual events.

To assess the degree of overlap among the memory tasks, we examined correlations between free recall, recognition, and temporal-order performance at the participant level. Free recall showed a weak association with recognition ($r = .15$) and no association with temporal-order memory ($r = -.02$), whereas recognition and temporal-order performance were moderately correlated ($r = .40$). These patterns suggest that free recall captures memory processes that are at least partially distinct from those assessed by probe-based tasks.

Although all memory tasks sampled from the same 35 events, the exact stimuli (video clips or still images) were non-overlapping. The cued recall task used brief dynamic clips, whereas the recognition and temporal-order tasks used static images from different moments within those same events. Thus, no identical image or clip appeared in more than one task.

Psychometric Battery. Participants also completed cognitive assessments using the NIH Toolbox on an iPad with the researcher. Before starting the session, the researcher registered each participant on the iPad and entered their demographic information, including date of birth, sex, race, and education level. Using the NIH Toolbox, we assessed various cognitive domains: learning and memory with the Rey Auditory Verbal Learning Test (RAVLT), processing speed with the Pattern Comparison Processing Speed Test, executive function and attention with the Flanker Inhibitory Control and Attention Test, and language and vocabulary knowledge with the Picture Vocabulary Test. Norms for the NIH Toolbox were created based on a sample of 4,859 participants, aged 3 to 85 years, reflecting the diversity of the U.S. population in terms of age, gender, race/ethnicity, and education. Standardized scores, such as Fully Corrected T-Scores, are calculated using this normative data and adjusted for demographic factors, enabling meaningful comparisons within similar groups (National Institutes of Health (NIH) Toolbox, 2021).

Rey Auditory Verbal Learning Test (RAVLT). Episodic long-term memory was assessed using the RAVLT. Participants were asked to learn a list of 15 words presented verbally. Care was taken to ensure the volume was turned up loud enough for participants to hear. Participants first listened to the word list and then recalled as many words as possible. The correct responses were listed on the iPad screen that the researcher held, ensuring that the participants could not see. As the participant recalled the words on the list, the researcher touched the corresponding word on the screen to log the response. Only correct responses were recorded. If the participant recalled an incorrect word or repeated a response, this was not recorded. This encode-recall process was repeated three times to produce an immediate memory score. After a delay of approximately 10 minutes, participants were tested again, this time without hearing the list again. The total time for this task was 4 minutes. This task yields an immediate and a delayed test score.

Pattern Comparison Processing Speed Test. Participants completed the Pattern Comparison Processing Speed Test on their own using the iPad. This task assessed processing speed by asking participants to quickly determine whether two presented stimuli were the same or different. Participants were shown pairs of images and instructed to decide whether the images were the same or different while moving through the task as quickly as possible without making mistakes. Participants completed as many trials as possible within 4 minutes.

Flanker Inhibitory Control and Attention Test. The Flanker Inhibitory Control and Attention Test assessed participants' ability to focus on a central target while ignoring distractions. Participants were shown five fish with arrows on them. Their task was to determine the direction of the arrow on the center fish while ignoring the “flanker” fish on either side. The arrows on the flanker fish sometimes pointed in the same direction as the center fish (congruent

trials) and sometimes pointed in the opposite direction (incongruent trials). This task measured participants' ability to maintain focus on a specific target while filtering out irrelevant distractions. The total time for this task was 4 minutes.

Picture Vocabulary Test. The Picture Vocabulary Test assessed participants' vocabulary knowledge. For each trial, participants listened to an audio recording of a word and were shown four pictures. Their task was to select the picture that most closely matched the meaning of the word. The total time for this task was 3 minutes.

Conscientiousness Survey. Participants completed a personality survey administered via Qualtrics. The survey focused on conscientiousness, one of the major domains of personality, based on participants' self-reported answers. Participants were instructed to describe themselves as they generally are now, rather than how they wish to be in the future. They were asked to evaluate themselves honestly in comparison to others of the same sex and roughly the same age. To encourage truthful responses, participants were assured that their answers would remain confidential. Conscientiousness is composed of six facets, or subgroups: self-efficacy, orderliness, dutifulness, achievement-striving, self-discipline, and cautiousness. Participants rated each statement on a 5-point scale: 1 = Very Inaccurate, 2 = Moderately Inaccurate, 3 = Neither Accurate nor Inaccurate, 4 = Moderately Accurate, and 5 = Very Accurate.

Scores for each of the NIH Toolbox measures are reported separately for young and older adults in Supplementary Table S3. The values reported in this table are the standardized scores that are adjusted for age and education levels. Because of these adjustments, we would not expect to see large age-related differences in performance on the individual tasks.

Supplementary Table S1. Linear mixed-effects regression results for predicting RSA

Parameter	*B*	*SE*	*z*	*p*
Intercept	0.0353	0.0180	1.95	.057
Trial Type (Within vs. Across)	0.0131	0.0060	2.16	.031
Age Group (Young vs. Old)	0.0084	0.0107	0.79	.434
Perceptual Change (Z-score)	–0.0455	0.0031	–14.90	< .001
Trial Type × Age Group	0.0066	0.0084	0.79	.432

Note. Results from a linear mixed-effects model predicting RSA. *B* = fixed-effect regression coefficient; *SE* = standard error; *z* = Wald test statistic; *p* = two-tailed *p*-value. Trial Type contrast is coded as Across vs. Within (0 = Across, 1 = Within). Age Group contrast is coded as Old vs. Young (0 = Old, 1 = Young). Random intercepts for subjects and events were included.

Supplementary Table S2. Correlations between within-event RSA and temporal gap for young and older adults.

Age Group	n	r	95% CI Lower	95% CI Upper	t	*p*
Young Adults	1470	–0.049	0.100	0.002	-1.897	0.058
Older Adults	1470	–0.090	0.140	–0.039	-3.444	0.001

Note. *n* = number of observations; *r* = Pearson correlation coefficient; CI = 95% confidence interval; *t* = test statistic; *p* = two-tailed *p*-value.

The correlation was small and not significant for young adults but significant and negative for older adults. I also ran a mixed-effects model (RSA ~ scaled Gap × Agegroup + (1|SubNum)), confirmed a significant main effect of temporal gap ($\beta = –0.018$, $p < .001$) but no significant interaction with age group.

Supplementary Table S3. Comparison of young and older adults across NIH Toolbox, event memory tests & conscientiousness measures

Variable	Young adults		Older adults		*t*	*f*	*p*
	M	*SD*	*M*	*SD*			
NIH toolbox							
Pattern Comparison Processing Speed	48.7	10.87	49.9	10.07	0.51	0.275	.60
Picture Vocabulary	37.3	10.14	50.5	7.08	6.52	44.31	.001***
Flanker Inhibitory Control & Attention	51.4	10.09	50.0	6.90	-0.66	.45	.50
RAVLT (Immediate)	51.3	9.35	54.8	11.88	1.45	2.403	.13
RAVLT (delayed)	56.6	8.77	60.8	8.77	2.28	5.712	.02*
Event Memory tests							
Free recall	35.20	0.47	34.0	0.47	-0.73	0.544	0.46
Cued Recall	49.50	0.50	36.80	0.48	–5.13	26.32	.001
Recognition	.86	0.118	.86	0.09	-0.34	0.115	.73
Recognition Reaction Time	5.56	2.25	6.14	3.07	1.27	1.587	.22

Temporal Order	.69	0.166	.66	0.154	-0.99	0.912	.34
Temporal Order Reaction Time	5.24	1.84	7.42	2.74	4.71	23.194	.001***
Conscientiousness Survey							
Conscientiousness	227.24	29.59	237.56	18.80		4.387	.04*
Self-efficacy	38.51	5.26	39.91	4.01		1.898	.17
Orderliness	38.61	7.11	38.38	5.50		0.033	.86
Dutifulness	42.37	4.48	45.13	3.96		10.906	.001***
Achievement striving	39.51	6.88	41.53	4.38		3.048	.08
Self-discipline	33.32	7.43	35.78	5.26		3.391	.07
Cautiousness	34.93	6.34	36.82	4.04		3.427	.07

Note. M = Mean; SD = Standard Deviation. *$p < .05$, **$p < .01$, ***$p < .001$. All NIH Toolbox scores are age- and education-adjusted standardized scores.

Supplementary Table S4 presents the correlations among neural pattern similarity measures (Within RSA, Across RSA), event memory performance (free recall, within- and across-event cued recall, recognition, and temporal order), and cognitive measures from the NIH Toolbox (Pattern Comparison, Picture Vocabulary, Flanker, and RAVLT immediate and delayed recall) separately for young and older adults. As expected, Within and Across RSA values were strongly positively correlated in both age groups. Several cognitive measures were moderately correlated with memory performance measures, particularly in young adults, where free recall

was positively associated with Flanker and RAVLT performance. Notably, within-event cued recall was significantly associated with multiple memory measures, including recognition and temporal order, in both age groups, whereas across-event cued recall showed no significant associations with other memory measures in older adults and only a single significant association in young adults. Correlations between RSA and behavioral measures were generally small and non-significant. These results provide an overview of the interrelationships between neural similarity, memory outcomes, and standardized cognitive abilities across the two age groups.

Supplementary Table S4. Correlations between variables

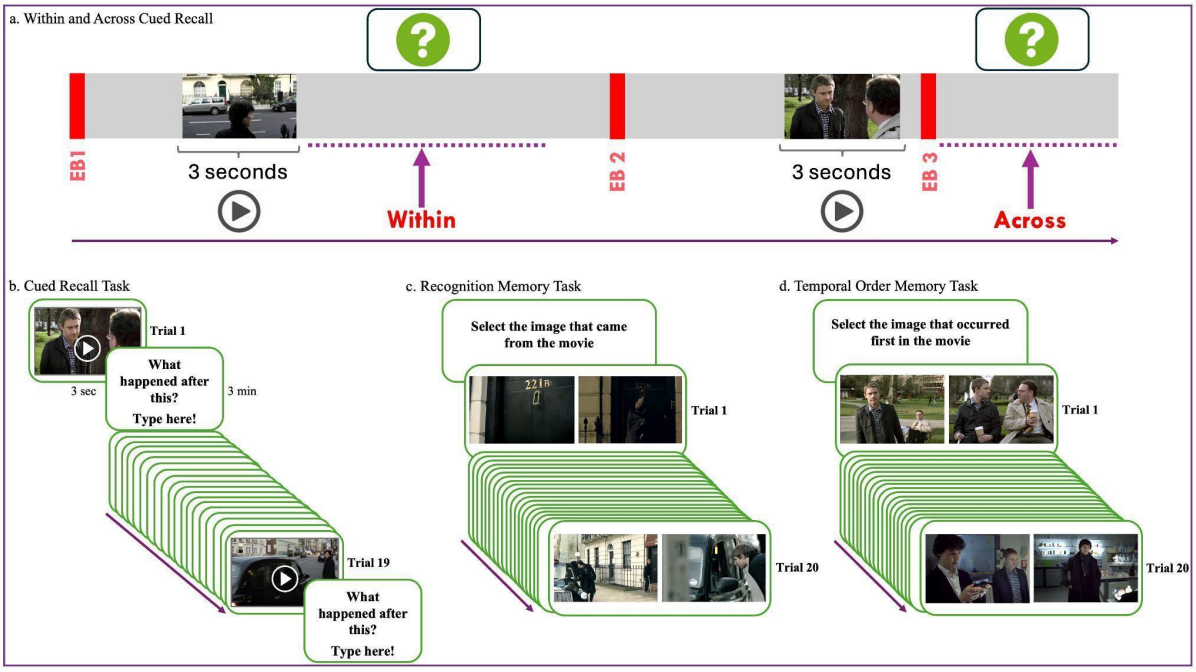
Group		Variable	1	2	3	4	5	6	7	8	9	10	11
YA	Event memory												
YA	1	Within RSA											
YA	2	Across RSA	0.80***										
YA	3	Free recall	0.04	-0.05									
YA	4	Recognition	-0.33	-0.33	0.18								
YA	5	Temporal order	-0.10	0.03	-0.07	0.28							

YA	6	Within cued recall	-0.09	0.04	0.11	0.32	0.38*						
YA	7	Across cued recall	-0.10	-0.07	0.26	0.45*	0.46*	0.42**					
YA	NIH Toolbox												
YA	8	Pattern comparison	-0.11	-0.01	0.26	0.15	-0.14	0.32	0.29				
YA	9	Picture Vocabulary	0.16	0.26	0.08	0.21	0.07	0.35	0.39*	0.70***			
YA	10	Flanker	-0.25	-0.26	0.41*	0.38	-0.22	0.02	0.20	0.74***	0.39*		
YA	11	RAVLT_immediate	-0.01	-0.06	0.54**	0.02	-0.36	0.17	0.04	0.49**	0.24	0.36	
YA	12	RAVLT_delayed	0.18	0.19	0.42*	0.13	-0.05	0.32	0.17	0.15	0.18	0.03	0.74***
OA	Event memory												
OA	1	Within RSA											

OA	2	Across RSA	0.75***								
OA	3	Free recall	-0.07	-0.16							
OA	4	Recognition	-0.13	-0.24	0.13						
OA	5	Temporal order	-0.20	-0.05	0.00	0.47**					
OA	6	Within cued recall	0.08	-0.06	0.15	0.14	0.00				
OA	7	Across cued recall	0.09	0.12	0.20	0.45**	0.34*	0.42**			
OA	NIH Toolbox										
OA	8	Pattern comparison	-0.07	-0.21	0.12	0.17	0.03	0.08	0.23		
OA	9	Picture Vocabulary	0.18	0.13	-0.14	0.17	0.04	0.14	0.25	0.19	
OA	10	Flanker	-0.05	-0.21	-0.19	0.00	0.19	-0.01	-0.16	0.29	-0.05

OA	11	RAVLT_immediate	0.19	0.29	-0.02	0.04	0.11	0.13	0.24	0.13	0.57***	0.27	
OA	12	RAVLT_delayed	0.03	0.15	0.11	0.13	0.25	0.01	0.23	-0.08	0.41**	0.03	0.75***

Note. YA = young adults; OA = older adults; Within RSA = mean neural pattern similarity for within-event segment pairs; Across RSA = mean neural pattern similarity for across-event segment pairs; Free Recall, Within-Event Cued Recall, Across-Event Cued Recall, Recognition, and Temporal Order tasks were scored as proportion correct; Pattern Comparison, Picture Vocabulary, Flanker, and RAVLT Immediate and Delayed Recall are subtests from the NIH Toolbox, and all NIH Toolbox values are standardized scores adjusted for age and education. Correlations were computed across participants within each age group using participant-level mean values, reflecting between-subject relationships.



Supplementary Fig. S1. Overview of memory tasks used to assess event memory during naturalistic movie viewing.

(a) Cued Recall Trial Types: Each 3-second video clip was drawn from either within a single event (within-event trial) or spanned an event boundary (across-event trial).

(b) *Cued Recall Task:* Participants completed 20 randomized trials and typed what they thought would happen next after each clip.

(c) Recognition Memory Task: In 20 two-alternative forced-choice (2AFC) trials, participants selected the image that had appeared in the movie segment.

(d) Temporal Order Memory Task: In 20 2AFC trials, participants chose which of two images occurred earlier in the movie.