Supplementary Materials

Study 1

Memory performance was assessed as a continuous variable with a logistic mixed effect model using the glmer function in R's lme4 package (Bates et al., 2015), the numDeriv package (Gilbert & Varadhan, 2019) to more accurately assess convergence using the Richardson extrapolation, and the sjPlot package (Lüdecke, 2021). Model fit was determined with comparisons using a likelihood ratio test. A model containing random effects of both cue (n = 22 unique cues obtained from within and between event cues) and participant (n = 49) was compared to a model containing only the random effect of participant. The inclusion of cue as a random effect significantly improved model fit ($\chi2(1) = 273.42$, $p < .001$; ICC = .443) and was used as a random intercept in subsequent models. The model was then built to test for the three-way interaction between cue type, age group, and overall memory performance (as a continuous predictor), excluding the random effect of participant.

A logistic mixed effects model (estimated using ML and BOBYQA optimizer with 20000 iterations) was run to predict performance (proportion of correct responses to the cued recall task) from cue type, age group, and free recall performance (proportion of details that were coded as internal), with the effect of participant and cue entered as random effects. We tested whether model fit including all main effects and 2-way interactions would be improved by the inclusion of the 3-way interaction between cue type, age group, and performance. The addition of the 3-way interaction term significantly improved model fit ($\chi2(1) = 3.92$, $p = .048$). As shown in Figure S1, memory for within-event and between-event cues do not differ in low-performing older adults and start to differentiate at higher levels of overall memory performance. Thus, the interpretation of the median-split analysis does not change when treating memory performance as a continuous variable, though again, these results should be interpreted with caution given the small number of scenes and individuals tested.

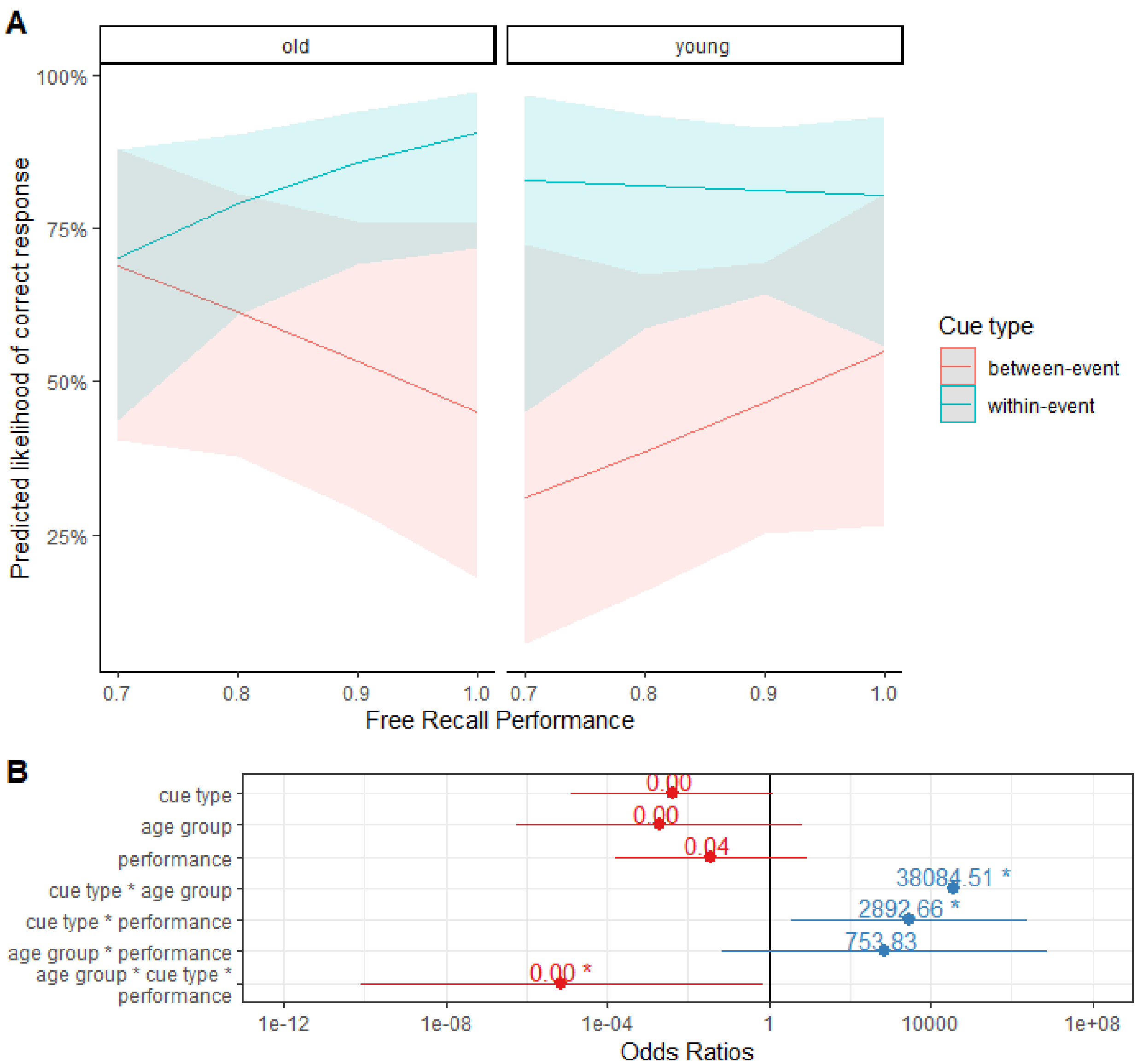

Figure S1. A. Cued recall accuracy as predicted by the logistic mixed effects model including free recall performance (proportion internal details) as a continuous predictor. Shaded area represents 95% CI. B. Logistic regression model coefficients for model predicting cued recall accuracy including performance as a continuous predictor, * $p < .05$, *** $p < .001$.

Study 2

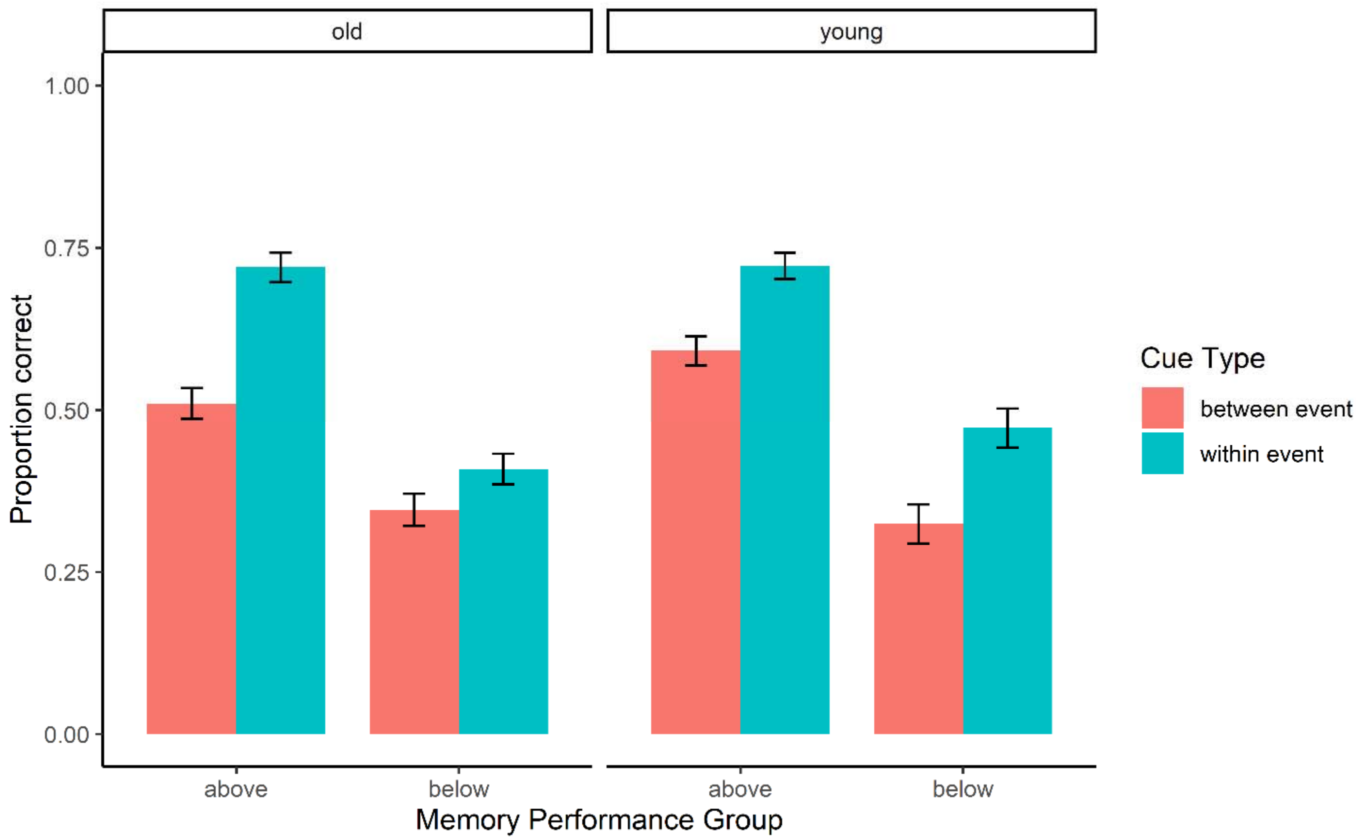

Figure S2. Mean cued recall accuracy in Study 2 divided according to cue type, age group, and performance group (median split of overall performance within each group).

Error analysis

As an exploratory analysis, we also examined the types of errors people made; that is, whether they wrongly reported an event from earlier in the movie, the next scene (i.e., the one immediately following the correct response), or long after. To assess whether there were any differences in the type of errors made by older and younger adults, separate 2 Age (between-subjects; old, young) by 3 Error Type (within-subjects; earlier, next, long after) by 2 Performance Group (between-subjects; high performers, low performers) mixed ANOVAs were run for each cue type (between-event, within-event)[1]. For both cue type analyses, sphericity was violated (within; $\varepsilon = .894$, between; $\varepsilon = .931$), so Greenhouse-Geisser corrected results are reported. For within-event cues, a main effect of performance group ($p < .001$) was observed such that high performers made fewer temporal errors. This finding was expected because the grouping was determined by overall memory performance. There was no main effect of age group ($p = .670$). However, the main effect of error type was significant, $F(1.79, 173.4) = 82.56$, $p < .001$, $\eta^2_p = .460$, such that errors where participants described earlier scenes were less frequent than errors describing either the next, or a long after scene (ps > .001), which did not differ from each other ($p = .758$; see Figure 5). There were no significant interactions with age (ps > .302). However, there was a significant 2-way interaction between performance group and error type $F(1.79, 173.4) = 7.72$, $p < .001$, $\eta^2_p = .074$. Pairwise comparisons of performance groups for each error type revealed that this interaction was driven by low performers more often recalling both the next scene ($p = .010$) and scenes that were long after the target ($p < .001$), they did not differ in the frequency of earlier scene errors (ps = .996). Thus, for within-event cues, regardless of age group, high and low performers made a similar number of earlier scene errors (which were relatively infrequent), but low performers more often made errors describing both the next, and long after scenes.

[1] Note that this analysis does not include trials on which participants typed “I don’t know” or provided no response and thus, is not simply the inverse of the accuracy analysis. It also does not include incorrect responses that did not correspond to any time point in the movie (there were very few of these; see “Incorrect Information” in Table 1).

Table S1. Mean accuracy and errors on the cued recall task separated by age and performance groups.

	Young Adults			Older Adults		
	Low Performers	*High Performers*	*Overall*	*Low Performers*	*High Performers*	*Overall*
Accuracy						
Within-event	0.48 (.15)	0.72 (.11)	0.61 (.18)	0.41 (.11)	0.72 (.12)	0.58 (.18)
Between-event	0.33 (.15)	0.59 (.12)	0.47 (.19)	0.35 (.12)	0.51 (.13)	0.44 (.15)
Temporal Errors						
Earlier scene						
Within-event	0.03 (.05)	0.02 (.04)	0.02 (.05)	0.03 (.04)	0.02 (.04)	0.03 (.04)
Between-event	0.16 (.13)	0.10 (.08)	0.13 (.11)	0.14 (.12)	0.14 (.11)	0.14 (.11)
Next scene						
Within-event	0.20 (.13)	0.14 (.08)	0.17 (.11)	0.22 (.13)	0.16 (.08)	0.18 (.11)
Between-event	0.17 (.10)	0.12 (.07)	0.15 (.09)	0.16 (.08)	0.15 (.08)	0.15 (.08)
Long after						
Within-event	0.22 (.16)	0.14 (.12)	0.18 (.14)	0.25 (.12)	0.10 (.09)	0.16 (.12)
Between-event	0.24 (.14)	0.15 (.12)	0.19 (.13)	0.24 (.11)	0.17 (.09)	0.20 (.11)
Any temporal error						
Within-event	0.43 (.21)	0.27 (.11)	0.34 (.18)	0.49 (.14)	0.25 (.12)	0.36 (.18)
Between-event	0.54 (.14)	0.35 (.13)	0.44 (.17)	0.55 (.14)	0.45 (.14)	0.49 (.15)
Incorrect Information						
Within-event	0.04 (.07)	0.01 (.03)	0.03 (.06)	0.02 (.04)	0.02 (.03)	0.02 (.04)
Between-event	0.03 (.07)	0.02 (.03)	0.02 (.05)	0.02 (.04)	0.01 (.03)	0.02 (.03)

Note. Accuracy is represented as proportion correct. All other categories are proportion of incorrect trials containing the relevant coded category. SDs are in brackets.

For between event cues, there was a main effect of performance group ($p < .001$). There was no significant main effect of age ($p = .234$). The main effect of error type was significant, $F(1.86, 180.5) = 9.03$, $p < .001$, $\eta^2_p = .085$, such that long after errors were more frequent than those describing either earlier scenes ($p = .002$), or the next scene ($p = .003$), earlier and next scene errors did not differ ($p = .290$). In contrast to the analysis of within-event cues, the interaction between error type and performance group was not significant ($p = .263$), and no other interactions were significant (ps > .077). Therefore, for between event cues, regardless of age group or performance group, errors describing temporally distant scenes were more frequent than other temporal errors (see Table 1 for means and Figure 5).

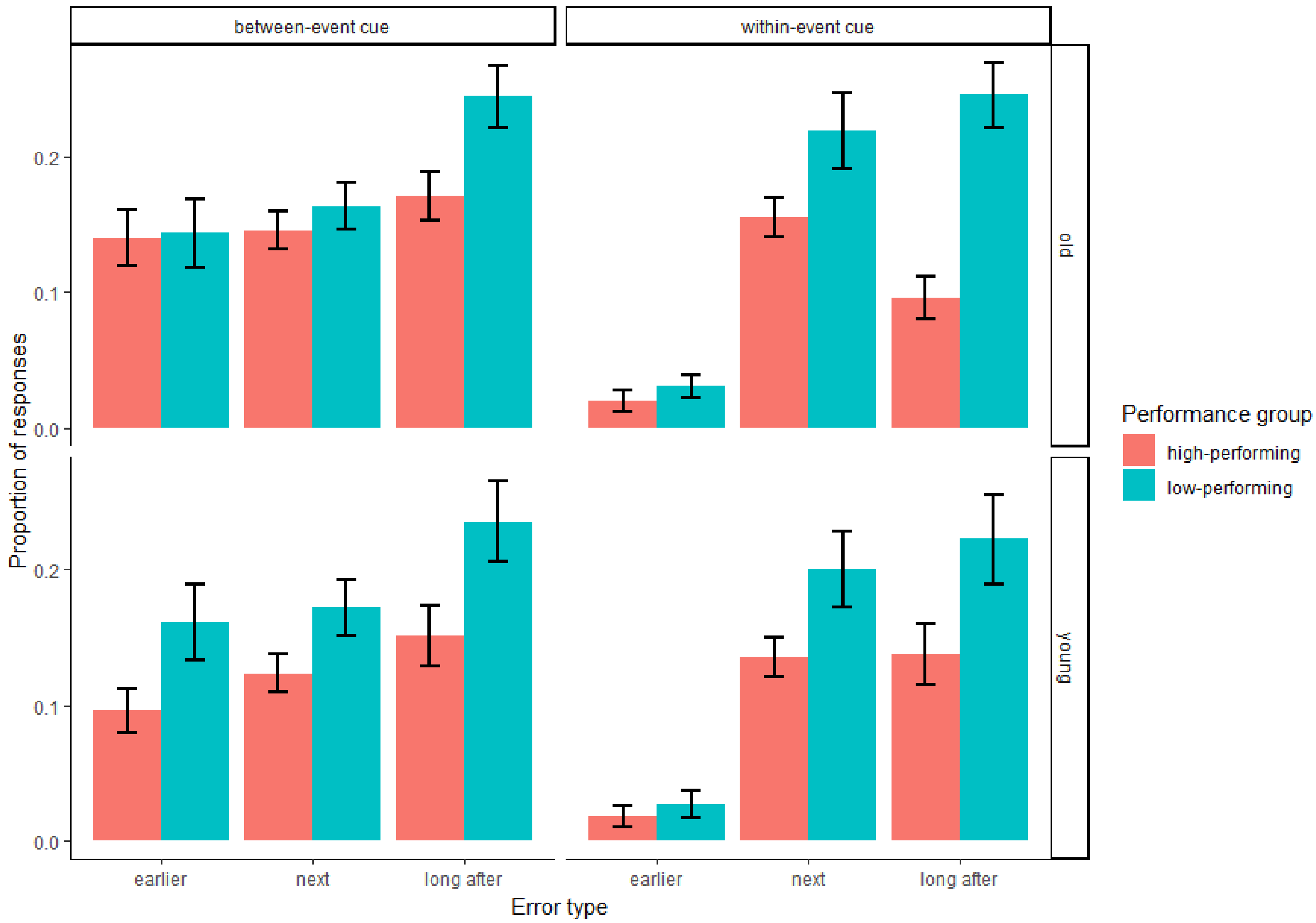

Figure S3. Proportion of responses on the cued recall task containing each type of temporal error separated by cue-type, performance group, and age. Error bars represent standard errors.

Our error analysis revealed that the types of temporal errors made by older and younger adults were similar, but for within-event cues, low performers made more errors describing scenes that came after the target (both near and distant). For between-event cues, regardless of age or performance group, errors describing scenes that occurred in the distant future were more common than other temporal errors.

Discussion

The distribution of temporal errors in the cued recall task was similar across older and younger adults. This is in line with recent work by Diamond et al. (2020), who found no age difference in error rates (which were remarkably low) during free recall of an in-person art tour. Our analysis focused on the temporal properties of errors, showing that participants made few errors describing earlier scenes, but relatively more describing the next or long-after scenes. The preference for recalling in a forward direction is consistent with the robust *asymmetry effect* observed in episodic recall wherein participants are more likely to move forward than backwards in time during recall (see Kahana et al., 2008, for review). The relatively large number of next scene errors is also consistent with the *temporal contiguity effect* (see Healey et al., 2019, for review), which describes the tendency for items temporally adjacent to the last recalled item to have a higher likelihood of retrieval (Polyn & Kahana, 2008). This effect was likely enhanced in the present study by the ordered presentation of cues, as previous work has shown that the temporal contiguity effect is compounded when previous transitions were from adjacent items (Lohnas & Kahana, 2014). This preference for next scene errors was seemingly diminished for between-event cues, which tended to elicit more 'long-after' errors. This is likely a product of the temporal distortion that occurs at event boundaries (Bangert et al., 2020; Lositsky et al., 2016) or simply the nature of the cues, since making a 'next' scene error for between-event cues requires participants to cross two event boundaries and may therefore already be perceived to be quite temporally distant. This may have weakened the temporal contiguity effect resulting in more errors due to non-serial jumps between events (DuBrow & Davachi, 2016). Both older and younger adults have been shown to display these temporal organization effects in free recall, though, there is evidence that contiguity has a weaker influence on older adults (Kahana et al., 2002). The lack of age differences in the present study may be due to our use of cued recall and the ordered presentation of cues. Taken together, our error analysis fits reasonably well with previous work on the temporal organization of episodic recall.