Supplemental Material

1. List of Activities in the Everyday Changes Task

Table S1. *List of activities, questions, and correct answers from the memory test*

Nb	Activity	Question	Response 1	Response 2
1.	Alarm	Which of the devices on the nightstand did the actor reach for just after waking up?	Smart Phone	Alarm Clock
2.	Exercise	What did the actor do on the exercise mat?	Ab Crunches	Stretching
3.	Self-Cleaning	How did the actor prepare to wash her body?	Bath	Shower
4.	Hair	What did the actor use to style her hair?	Brush	Comb
5.	Clean Counter	What kind of implement did the actor use to clean the kitchen counter?	Wash Cloth	Paper Towel
6.	Dog Food	Which side was the dog dish the actor took from the floor located on?	Left	Right
7.	Dog Leash	How did the actor attach the leash to her dog?	Harness on its body	leash on its necks
8.	Dog Treat	What kind of container in the cabinet held the treat the actor gave to the dog?	Bag	Jar
9.	Check Device	What electronic device did the actor use to check the news in her home office?	iPad	Laptop
10.	Pills	What kind of pain medication did the actor take in the kitchen?	Aleve	Ibuprofen
11.	Breakfast	What did the actor take from the kitchen counter to eat at breakfast?	Banana	Breakfast Bar
12.	Toothcare	How did the actor care for her teeth immediately after brushing?	Mouthwash	Floss
13.	Book For Bag	What type of book did the actor place in her bag?	Textbook	Notebook
14.	Bag In Car	Where in the car did the actor place her bag?	Passenger Seat	Trunk
15.	Car Music	What type of device did the actor use just after starting her car?	iPod	CD player
16.	Breath Freshener	What type of breath freshener did the actor use while in her car?	Mints	Gum
17.	Enter Home	Which lock on the front door did the actor unlock to enter her home?	Door Handle	Deadbolt

18.	Background Electro	What type of electronics did the actor turn on just after entering her home?	Stereo (Ipod)	Television
19.	Stock Refrigerator	What type of beverage did the actor take from the dining room closet to the fridge?	Sports Drink (two Gatorade bottles)	Carbonated Water (two cans)
20.	Stock Bathroom	Where in the bathroom did the actor hang the towel she took from the closet?	Hook on the back of the door	Rack on the right of the sink
21.	Dog Play	What toy did the actor use when playing with the dog?	Stuffed Bone	Red Ball
22.	Cardio	What type of cardio exercise activity did the actor go outside to perform after stretching downstairs?	Biking	Running
23.	Chest Exercise	How did the actor move her arms when facing toward the camera on the exercise machine?	Circle segment (fly)	Parallel (press) / Rowing
24.	Leg Exercise	What leg exercise did the actor do next to the exercise machine?	Standing Lunges (one leg forward)	Squats
25.	Calf Raise	Where in the basement was the actor standing when she performed calf raises (standing up on her tiptoes)?	First step of the stairs	Ground
26.	Post Workout	What did the actor consume just after her workout?	Protein Drink	Protein Bar
27.	Washer	What form of laundry detergent did the actor use in the washing machine?	Liquid	Powder
28.	Dryer	What form of fabric softener did the actor use in the dryer?	Magnetic bar	Sheets
29.	Folding Laundry	What did the actor first fold after emptying the dryer?	Shirt (Cardinal red)	Towel (grey)
30.	Ironing	What clothing did the actor iron?	Pants (Khaki)	Shirt (dark blue)
31.	Waste To Curb	Which of the three waste containers did the actor take from the driveway to the street?	Middle / smaller one	Right /bigger one
32.	Kitchen Floor	What floor cleaning implement did the actor take from the closet beneath the stairs?	Swiffer	Steam Mop
33.	Greens Container	What type of package held the salad greens?	Clamshell	Bag
34.	Pizza Seasoning	What type of seasoning did the actor use on her pizza?	Red Pepper	Parmesan
35.	Drink	What drink did the actor have with her dinner?	Water	Milk

36.	Dessert	What type of dessert did the actor get after diner from the Tupperware container on the counter?	Cookie	Brownie
37.	Dishes	How did the actor clean the dishes?	Dish Washer	Hand Wash
38.	Candy Bowl	What candy did the actor place in the bowl?	Skittles	Swedish (Red) Fish
39.	Leisure	What did the actor do while sitting on the couch in the basement?	Magazine	Television
40.	Tea	What method did the actor use to heat water for her drink?	Tea Kettle	Keurig (Electronic)
41.	Paint Toenails	When the actor painted her toenails in the bathroom, what color did she use?	Green	Red/Pink
42.	Face Cleaning	What did the actor use to clean her face?	Liquid Cleaner and Water	Wipe
43.	Lotion	On what part of her body did the actor apply lotion?	Arms	Legs
44.	Bed Entertainment	What activity did the actor perform in bed before going to sleep?	Crossword Puzzle	Reading
45.	Pillow	Which pillow did the actor take from the other side of the bed when she went to sleep?	Head Pillow	Knee Pillow

2. Content of Responses That Were Not Day 1 Consistent Predictions

To further examine the responses that were not Day 1 consistent predictions, we further classified them in different categories based on their content. These categories were (1) "*No Response*" when the participant was unable to make any prediction; (2) "*Activity Only*" when the participant predicted the right ending for the activity but omitted the critical feature that would change across movies; (3) "*Intra-list Intrusions*" when the participant predicted the ending of another activity that they watched in the first movie; (4) "*Alternate-list Intrusions*" when the participant predicted the ending of the alternate version of an activity that they watched in the first movie; (5) "*Unrelated Predictions*" when the participant made a prediction that did not happen in any of the movies. For instance, if a Day 1 consistent prediction was be that the actress took a Swiffer mop from a closet and started cleaning the floor (using a steam mop being the alternate ending), then *"activity only"* would be predicting that she would start cleaning the floor without mentioning what implement she used; "*Intra-list Intrusions*" would be that the actress took from the closet a towel to hang in the bathroom if the participant saw that ending in another activity of the first movie; "*Alternate-list Intrusions*" would be that she started cleaning the floor with the steam mop that she did not took from the closet in the version of the movie that the participant first watched; "*Unrelated Predictions*" would be that she took a boardgame from the closet when no activity or their alternate version involved boardgames. The interrater agreement for these five categories was very high with a Cohen's κ = .82 (Landis & Koch, 1977).

To examine group differences in the frequency of each of these kinds of responses that were not Day 1 consistent predictions, we created a dummy variable for each of them and computed a series of logistic linear mixed models with the dummy variables as dependent variables and a fixed effect of Age. Models 1, 3, and 4 below had random intercept of Participant and Activity only (using more complex random effect structure either resulted in convergence and singularity issues or did not the improve goodness-of-fit; Matuschek et al., 2017). Models 2 and 5 had a random intercept of Participant and a random slope of Age by Activity. The results for the main effect of Age as well as the model estimated probabilities are reported in Table S2.

Table S2. *Model results and model-estimated probabilities conditioned on age group for each subcategory of not Day 1 Consistent responses*

Prediction Type	Model Estimated probability		χ^2	*df*	*p*
	Younger	Older			
1. No Response	.02 [0.01, 0.03]	.04 [0.02, 0.06]	8.97	1	.003
2. Activity Only	.06 [0.04, 0.11]	.07 [0.04, 0.12]	.36	1	.55
3. Intra-List intrusions	.04 [0.03, 0.06]	.05 [0.03, 0.08]	3.06	1	.08
4. Alt.-list intrusions	.02 [0.01, 0.03]	.02 [0.01, 0.03]	0.56	1	.45
5. Unrelated	.07 [0.04, 0.10]	.15 [0.11, 0.20]	17.94	1	<.001

Note: 95% confidence intervals are displayed in brackets.

3. Cued Recall Task

3.1. Older Adults Produced More No Responses

To examine the proportion of No Responses for Day 2 recall, we computed a linear mixed model with no response as a dummy variable and 2 (Age: younger vs. older) × 2 (Activity Type:

repeated vs. changed) fixed effects. The random effect structure only included random intercepts of Participant and Activity as more complex random effect structure either resulted in convergence and singularity issues or did not the improve goodness-of-fit (Matuschek et al., 2017). The model indicated a main effect of Age, $\chi^2(1) = 11.08$, $p < .001$ with more no response in older than younger adults, no significant effect of Activity Type, $\chi^2(1) = 1.53$, $p = .21$, and no significant interaction $\chi^2(1) = 0.17$, $p = .68$. The model estimate probabilities are reported in Table S3.

To examine no response rate for Day 1 recall of changed activity correctly remembered as such, we computed a linear mixed model with no response as a dummy variable and Age (younger vs. older) as fixed effect. The random effect structure only included random intercepts of Participant and Activity as adding a random slope of Age by Activity did not improve goodness-of-fit. The model indicated that the effect of Age was significant $\chi^2(1) = 9.11$, $p = .003$ with older adults making more no responses than younger adults. The model estimate probabilities are reported in Table S3.

Table S3. *Model-Estimated probabilities for No Responses during cued recall test Conditioned on Age and Activity type*

		Activity Type	
Measure	Age	Repeated	Changed
Day 2 No Response	Younger	0.02 [0.01, 0.03]	0.01 [0.01, 0.03]
	Older	0.05 [0.03, 0.09]	0.05 [0.03, 0.08]
		Changed activity correctly remembered as such	
Day 1 No Response	Younger	0.007 [0.003, 0.02]	
	Older	0.03 [0.01, 0.06]	

Note: 95% confidence intervals are displayed in brackets.

3.2. Model selection

We used the backward-selection method proposed by Matuschek and colleagues (2017) to determine the random effect structure of our models. We first attempted to fit the maximal model with random intercepts for Participant and Activity and random slopes for each relevant fixed effect and their interactions and gradually decreased the complexity of the random effect structure in case of convergence or singularity issues until a fit occurred. We then continued to use this backward-selection method to reduce model complexity until a further reduction would lead to a significant decrease in the goodness-of-fit. The model trimming followed the following sequence: Maximal model → remove interaction random slopes for Activity → remove interaction random slopes for Participant → remove random slopes for Activity → remove random slopes for Participant (i.e., model with random intercepts only). Adding a step to the sequence where we uncorrelated the random slopes and intercept did not influence the selected models. In case of convergence issue, the BOBYQA optimizer (Powell, 2009) was used in attempt to fit the models. The goodness-of-fit was tested with a likelihood ratio test with the default settings of the anova function of the R stats package and a liberal threshold of alpha = .20 (Matuschek et al., 2017). Table S4 below present the final structure of the models from the main manuscript, in Wilkinson notation.

Table S4. *Model Structure used for the analyses of the cued recall task in the main manuscript*

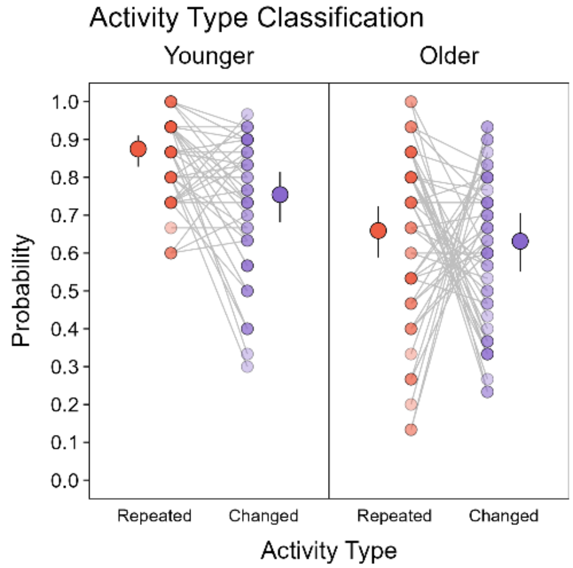
Nb.	Analysis	Model Specification
1.	Effects of Age (older vs. younger) and Activity Type (repeated vs. changed) on Prediction Type	MemConsPred01 ~ Age * Activity_Type + (1| ParticipantNb) + (1 | Activity)
2.	Effects of Age (older vs. younger) and Activity Type (repeated vs. changed) on Day 2 Recall	Recall_D2_ACC ~ Age * Activity_Type + (Activity_Type | ParticipantNb) + (Age + Activity_Type | Activity)
3.	Effects of Age (older vs. younger) on Day 1 Intrusions for changed activities	Recall_D2_INT ~ Age + (1 | ParticipantNb) + (1 | Activity)
4.	Effects of Age (older vs. younger) and Prediction (Day 1 consistent vs. not Day 1 consistent) on Day 2 Recall for changed activities	Recall_D2_ACC ~ Age * MemConsPred + (MemConsPred | ParticipantNb) + (1 | Activity)
5.	Effects of Age (older vs. younger) and Prediction (Day 1 consistent vs. not Day 1 consistent) on Day 1 Intrusions for changed activities	Recall_D2_INT ~ Age * MemConsPred + (1 | ParticipantNb) + (1 | Activity)
6.	Effect of Age (older vs. younger) and Prediction (Day 1 consistent vs. not Day 1	Recall_D1_ACC ~ Age * MemConsPred + (1 | ParticipantNb) + (1 | Activity)

	consistent) on Change Recollection for changed activities	
7.	Effect Age (younger vs. older), Prediction (Day 1 consistent vs. not Day 1 consistent) and Classification (recollected vs. not recollected) on Day 2 Recall for changed activities	Recall_D2_ACC ~ Age * MemConsPred * Recall_D1_ACC + (1 | ParticipantNb) + (1 | Activity)
8.	Effect Age (younger vs. older), Prediction (Day 1 consistent vs. not Day 1 consistent) and Classification (recollected vs. not recollected) on Day 1 Intrusions for changed activities	Recall_D2_INT ~ Age * MemConsPred * Recall_D1_ACC + (MemConsPred + Recall_D1_ACC | ParticipantNb) + (1 | Activity)

3.3. Older Adults Showed Poorer Memory for Whether an Activity Had Changed

EMRC predicts that if older adults' poorer event memory updating reflected impaired access to configural representations, then older adults should also show poorer overall memory for whether activities had earlier changed. We tested this additional hypothesis by comparing the accuracy of judgments during cued recall about whether activities had repeated or changed between days (Figure S1) using a 2 (Age: younger vs. older) × 2 (Activity Type: repeated vs. changed) model. The random effect included a random slope of Activity Type by Participant in addition to the random intercepts of Participant and Activity. There was a significant effect of Age, $\chi^2(1) = 39.50$, p <.001, showing that younger adults had better overall memory for activity types than older adults. There was also a significant effect of Activity Type, $\chi^2(1) = 4.67$, p = .003, and a tendency for an interaction effect, $\chi^2(1) = 3.62$, $p < .06$. Post-hoc tests showed that younger adults better remembered repeated than changed activities as such, z ratio = 2.88, p = .004, whereas older adults showed no significant difference between activity types, z ratio = 0.46, p = .65. Taken Together with the age-related differences in cued recall and Day 1 consistent predictions in the main manuscript, these findings suggest that older adults encoded and later retrieved fewer configural representations following Day 1 consistent predictions for changed activities.

Figure S1 *Classification of Activity Types in Cued Recall*



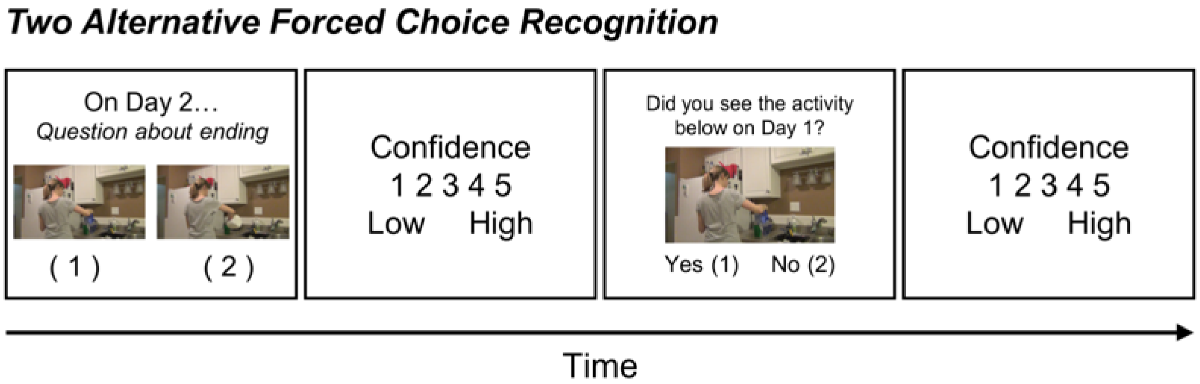
Note. The outer points are estimated probabilities from a mixed effects model. The inner points with connecting lines are the observed probabilities for each participant. Individual points with higher color intensities indicate more participants with the same probability. The error bars are 95% confidence intervals.

4. 2AFC Recognition

4.1.Task Description

After the cued recall test, participants completed a two-alternative forced choice recognition test. Participants first selected the activity ending from Day 2 from two still shots. The shots appeared side-by-side and included features that changed between the two version of each activity (e.g., the water pitcher [left] and milk jug [right]). The assignment of activity versions to screen position was evenly counterbalanced, and the same version appeared consecutively no more than thrice. Participants selected the left shot by pressing the “1” key and the right shot by pressing the “2” key. After responding, the unchosen shot appeared centered on the screen. Participants indicated if it appeared on Day 1, pressing the “1” key for “yes” and the “2” key for “no.” Responding was self-paced, and participants rated their confidence after each memory judgment from 1 (Low) to 5 (High).

Figure S2 *Schematic of the recognition task*



Note. Participants first selected between two still shots the activity ending from Day 2 then indicated if the unchosen ending had appeared on Day 1. Participants made confidence ratings for all memory responses.

4.2. Model Selection

The model selection was performed using the same procedure from Matuschek and colleague (2017) as for the cued recall task. Table S5 presents the final structure of the models.

Table S5. *Model Structure used for the analyses of the recognition task*

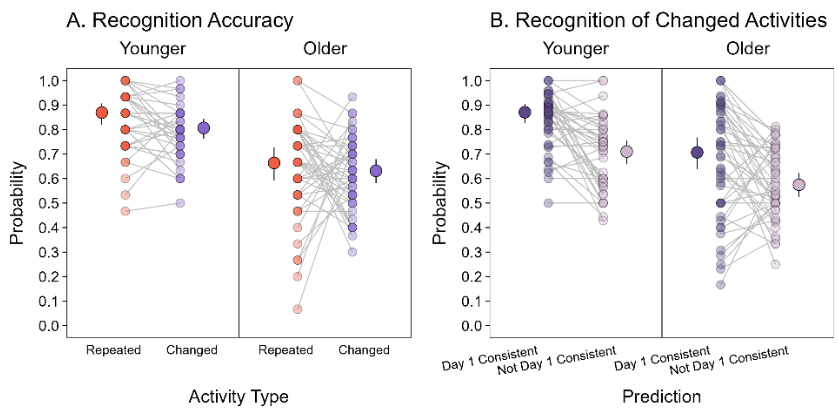
Nb.	Analysis	Model Specification
1.	Effects of Age (older vs. younger) and Activity Type (repeated vs. changed) on Day 2 Recognition	Reco_D2_ACC ~ Age * Activity_Type + (Activity_Type | ParticipantNb) + (Age + Activity_Type | Activity)
2.	Effects of Age (older vs. younger) and Prediction (Day 1 consistent vs. not Day 1 consistent) on Day 2 Recognition for changed activities	Reco_D2_ACC ~ Age * MemConsPred + (MemConsPred | ParticipantNb) + (1 | Activity)
3.	Effects of Age (older vs. younger) and Activity Type (repeated vs. changed) on Memory for Activity Type	Reco_ChangeReco_ACC ~ Age * Activity_Type + (Activity_Type | ParticipantNb) + (1 | Activity)
4.	Effect of Age (older vs. younger) and Prediction (Day 1 consistent vs. not Day 1 consistent) on Change Recollection for changed activities	Reco_ChangeReco_ACC ~ Age * MemConsPred + (1 | ParticipantNb) + (1 | Activity)
5.	Effect Age (younger vs. older), Prediction (Day 1 consistent vs. not Day 1 consistent) and	Reco_D2_ACC ~ Age * MemConsPred * Reco_ChangeClass + (MemConsPred +

Classification (recollected vs. not recollected) on Day 2 Recognition for changed activities	Reco_ChangeClass | ParticipantNb) + (1 | Activity)

4.2. Older Adults Showed Poorer Overall 2AFC Recognition Accuracy

We tested for age-related differences in 2AFC recognition accuracy for Day 2 activities (Figure S3A, Model 1 in Table S5) using a 2 (Age: younger vs. older) × 2 (Activity Type: repeated vs. changed) model. 2AFC recognition accuracy was measured as correct identification of the Day 2 activity when the two shots appeared at test. The model indicated a significant effect of Age, $\chi^2(1) = 56.27$, $p < .001$, showing higher accuracy for younger than older adults. There was a tendency for an effect of Activity Type, $\chi^2(1) = 3.66$, $p = .06$ with higher accuracy for repeated than changed activities. There was not a significant Age × Activity Type interaction, $\chi^2(1) = 1.74$, $p = .19$. These results support Hypothesis 1b and are consistent with cued recall in showing memory deficits associated with age, but are less reliable regarding the effect of Activity Type.

Figure S3. *2AFC Recognition Accuracy for Both Activity Types and for Changed Activities Conditioned on Prediction*



Note. The outer points are estimated probabilities from a mixed effects model. The inner points with connecting lines are the observed probabilities for each participant. Individual points with higher color intensities indicate more participants with the same probability. The error bars are 95% confidence intervals.

4.3. Older Adults Showed Updating Benefits Following Day 1 consistent predictions

We examined the association between Day 1 consistent predictions and 2AFC recognition of changed activity features only (Figure S3B) using a 2 (Age: younger vs. older) × 2 (Prediction: Day 1 consistent vs. not Day 1 consistent) model. The model indicated significant effects of Age,

$\chi^2(1) = 30.20$, $p < .001$, and Prediction, $\chi^2(1) = 40.38$, $p < .001$, showing that recognition accuracy was higher following Day 1 consistent predictions (supporting Hypothesis 2a) and a tendency for a significant interaction effect, $\chi^2(1) = 3.46$, $p = .06$. Although the effect tended to be stronger in younger adults, post-hoc tests showed that both younger adults, *z* ratio = 5.67, p < .001, and older adults, *z* ratio = 3.71, p < .001 benefitted from Day 1 consistent predictions. These results do not support Hypothesis 2b and is inconsistent with the findings of the cued recall task.

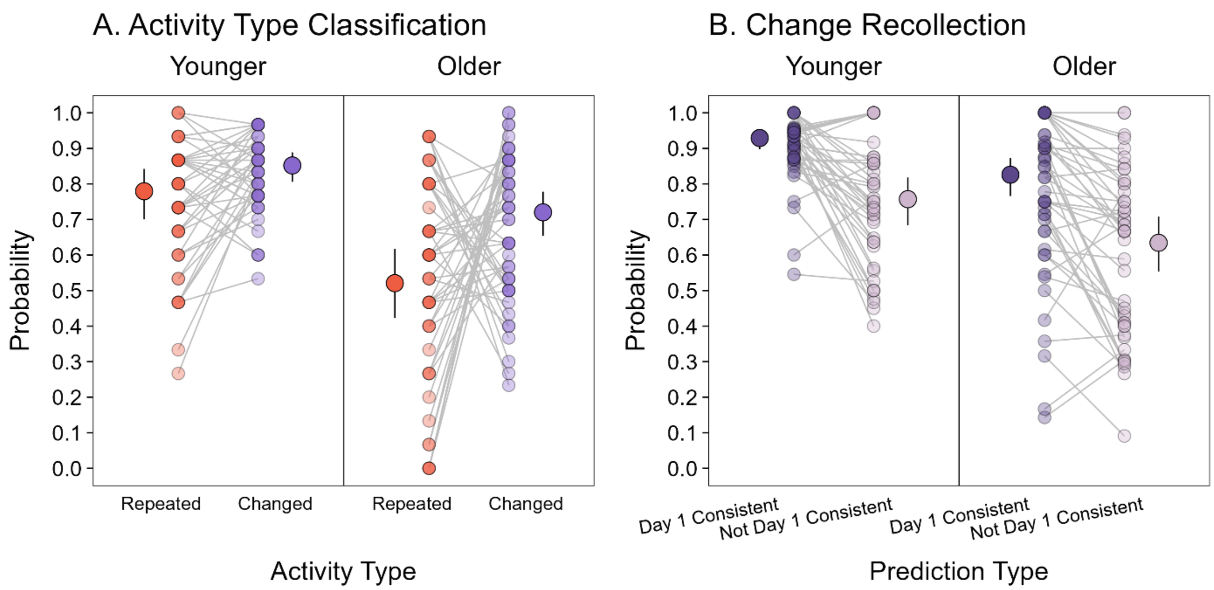
4.4. Older Adults Showed Poorer Memory for Whether an Activity Had Changed

We tested for age-related differences in memory for whether an activity was changed or repeated by comparing the accuracy of the second judgments made during 2AFC recognition trials. When participants were asked whether the still shot that they did not choose on the first judgment had appeared in the Day 1 movie, "yes" responses were counted as correct for changed activities and "no" responses were counted as correct for repeated activities. Note that this measure of activity type classification assesses if participants thought that there was more than one activity ending regardless of whether they correctly identified the still shot depicting the Day 2 activity ending on the first judgment. A 2 (Age: younger vs. older) × 2 (Activity Type: repeated vs. changed) model, see Figure S4A, indicated significant effects of Age, $\chi^2(1) = 55.50$, $p < .001$, and Activity Type, 10.36, $p = .001$, showing higher accuracy for younger than older adults and for changed than repeated activities. The interaction was not significant, $\chi^2(1) = 0.78$, $p = .38$. These are consistent with the pattern in cued recall showing an age-related deficit in memory for activity types (see section 3.3. and Figure S1).

4.5. Older Adults Showed Weaker Associations Among Day 1 Consistent Predictions, Recollecting that Activities had Changed, and Recognition of Changed Features

EMRC predicts that younger adults should recollect more changes than older adults, especially following Day 1 consistent predictions. Change recollection was defined as instances when participants indicated on the second judgment that the featured activity ending appeared in the Day 1 movie. Note that these instances are the same as the correct activity type classifications for changed activities in the previous analysis. The association between Day 1 consistent predictions and change recollection (Figure S4B) were examined with a 2 (Age: younger vs. older) × 2 (Prediction: Day 1 Consistent vs. not Day 1 consistent) model. The model indicated significant effects of Age, $\chi^2(1) = 12.35$, $p < .001$, and Prediction, $\chi^2(1) = 103.93$, $p < .001$, showing that change recollection was higher for younger than older adults and for Day 1 consistent predictions, supporting Hypothesis 3a. The interaction was also significant, $\chi^2(1) = 3.86$, $p = .049$ indicating that the beneficial effect of Day 1 consistent predictions on change recollection was stronger in the younger, *z* ratio = 8.21, $p < .001$, than older adults, *z* ratio = 6.94, $p < .001$, supporting Hypothesis 3b. This pattern was consistent with cued recall in the main manuscript.

Figure S4. *Classification of Activity Types and Change Recollection Conditioned on Prediction Type*



Note. The outer points are estimated probabilities from a mixed effects model. The inner points with connecting lines are the observed probabilities for each participant. Individual points with higher color intensities indicate more participants with the same probability. The error bars are 95% confidence intervals.

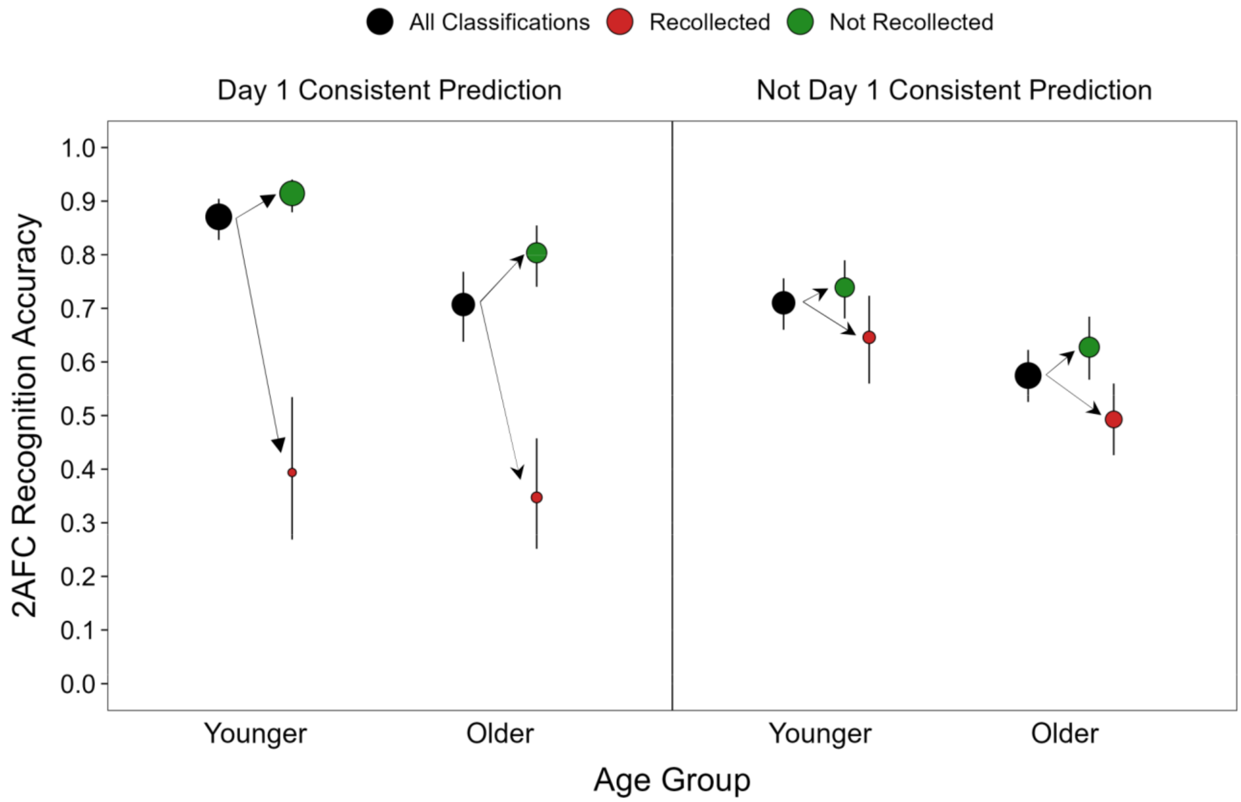
EMRC also predicts that, when mnemonic prediction errors promote the integration of features into configural representations, the association in memory for both features of changed activity should be stronger than when mnemonic prediction errors did not occur. For age differences, the account also proposes that age-related impairment in the contribution of mnemonic prediction errors to configural encoding should lead to a weaker association between recognition of changed Day 2 features on the first judgment and memory for the activity having changed on the second judgment (i.e., "yes" responses indicating that participants believe that they recollected changes). We tested this prediction by comparing 2AFC recognition accuracy for changed activity features conditioned on the change recollection measure from that task (Figure S5). Table S6 clarifies how observation differences across conditional cells contributed to overall recall differences by displaying the classification proportions for each age group and prediction type.

Table S6. *Cell Proportions for Changed Activities by Prediction and Change Classification*

		Change Classification		
Age	Prediction Type	Recollected	Not Recollected	Total
Younger (N = 44)	Day 1 consistent	.52	.05	.57
	Not Day 1 consistent	.31	.12	.43
	Total	.83	.17	1.00
Older (N = 47)	Day 1 consistent	.34	.09	.43
	Not Day 1 consistent	.34	.23	.57
	Total	.68	.32	1.00

Table S7 displays the complete results from a 2 (Age: younger vs. older) × 2 (Prediction: Day 1 consistent vs. not Day 1 consistent) × 2 (Classification: recollected vs. not recollected). Here, we focus on the significant effects that are not redundant with those described above. There was a significant effect of Classification, $\chi^2(1) = 82.01$, $p < .001$, qualified by a significant Prediction × Classification interaction, $\chi^2(1) = 60.67$, $p < .001$, showing stronger associations between change recollection and recognition of changed activity features following Day 1 consistent predictions. The key age-related comparison of recognition accuracy following Day 1 consistent predictions when change was recollected (left panel, green points) showed a stronger association in recognition of both features from changed activities for younger than older adults, *z* ratio = 3.76, $p < .001$. Collectively, these results converge with the findings from cued recall in the main manuscript (i.e., hypotheses 4a and 4b) by showing that Day 1 consistent predictions led to associations between activity features that better supported memory updating for younger than older adults.

Figure S5 *2AFC Recognition Accuracy for Changed Activities Only Conditioned on Prediction and Change Classifications:*



Note. The black points are model-estimated probabilities conditioned on the type of prediction made during Day 2 viewing. These estimates are the same as those displayed in Figure 5B. The colored points are probabilities conditioned on both prediction type and the type of activity type classification made after the initial recall responses during the cued recall test. The green points are changed activities that were identified as such followed by correct recall of Day 1 features (change recollected). The blue points are changed activities that were identified as such followed by incorrect recall of Day 1 features (change remembered). The red points are changed activities that were not identified as such (change not remembered). The error bars are 95% confidence intervals and are obscured when the intervals are smaller than the point diameters

Table S7. *Model Results for 2AFC Recognition Accuracy for Changed Activities Only Conditioned on Prediction and Change Classification:*

Effect	χ^2	*df*	*p*
Age	20.67	1	< .001
Prediction Type	19.70	1	< .001
Change Classification	82.01	1	< .001
Age × Prediction	0.80	1	= .37
Age × Change Classification	0.43	1	= .51
Prediction × Change Classification	60.67	1	< .001
Age × Prediction × Change Classification	3.48	1	= .06

5. Response Times

5.1. Correct Memory Responses were Faster After Day 1 Consistent Predictions

For the Sake of the preregistration, we report here analyses examining if there was an association between Day 1 consistent predictions and standardized responses times (zRTs) for correct memory responses for changed activities on the following measures: 1) change classification during cued recall, 2) identification of Day 2 features during recognition, and 3) identification of Day 1 features during recognition. Note that these were measures for which participants did not type recalled responses. zRTs were assessed in standard deviation units by transforming times across all responses into participant-level *z*-scores. Then, *z*-scores for correct answers were modeled separately for each memory measure using 2 Age (Age: younger vs. older) × 2 (Prediction: Day 1 consistent vs. not Day 1 consistent) linear mixed effect models with participant and activity as random intercepts. We do not report the main effects of Age as the z-scoring made them difficult to interpret. Following the application of the above-mentioned model selection procedure (Matuschek et al., 2017), the three models included a random intercept of activity only.

Table S8 displays the model-estimated values for each measure. The model for correct change classification during cued recall (top rows), indicated a significant effect of Prediction, $\chi^2(1) = 13.17$, $p < .001$, showing that classifications were faster following Day 1 consistent predictions. The model for correct Day 2 recognition (middle rows) also indicated a significant effect of Prediction, $\chi^2(1) = 9.55$, $p = .002$, showing that recognitions were faster following Day 1 consistent predictions. Finally, the model for correct Day 1 recognition (bottom rows) indicated a significant effect Prediction, $\chi^2(1) = 19.39$, $p < .001$, showing that recognitions were faster following Day 1 consistent predictions. There were no significant Age × Prediction interactions, largest, $\chi^2(1) = 0.31$, $p = .58$.

Table S8 *Model-Estimated Standardized Reaction Times for Correct Responses to Memory Measures Conditioned on Prediction in Cued Recall and Recognition Tasks*

		Prediction	
Measure	Age	Day 1 consistent	Not Day 1 consistent
Change Classification	Younger	-0.07 [-0.18, 0.04]	0.12 [-0.01, 0.26]
	Older	-0.17 [-0.30, -0.05]	-0.03 [-0.15, 0.09]
Day 2 Recognition	Younger	-0.14 [-0.28, < 0.01]	< 0.01 [-0.15, 0.15]
	Older	-0.16 [-0.31, -0.01]	-0.04 [-0.19, 0.11]
Day 1 Recognition	Younger	-0.15 [-0.26, -0.03]	0.05 [-0.08, 0.19]
	Older	-0.26 [-0.39, -0.14]	-0.05 [-0.18, 0.08]

Note: 95% confidence intervals are displayed in brackets.

For completeness, we compared zRTs for correct Day 2 recognition for repeated and changed activities. We computed a 2 (Age: younger vs. older) × 2 (Activity Type: repeated vs. changed) linear mixed model with Activity as only random intercept and random slopes Age and Activity Type (Table S9, top rows show model-estimated values). Results indicated no significant effect of Activity Type, $\chi^2(1) = 0.97$, $p = .32$, and no significant Age × Activity Type interaction, $\chi^2(1) = 0.25$, $p = .62$, showing that zRTs did not vary across activity types. Finally, we also compared zRTs for correct Day 2 recognition of changed activities only, conditioned on change recollection. A 2 (Age: younger vs. older) × 2 (Classification: recollected vs. not recollected) linear mixed model with activity as only random intercept and random slopes of Age and Activity Type (Table S9, bottom rows, shows model-estimated values) indicated a significant effect of Classification, $\chi^2(1) = 9.55$, $p = .002$, showing faster correct recognition of changed Day 2 activity features when change was recollected than when it was not recollected. There was no significant Age × Classification interaction, $\chi^2(1) < 0.01$, $p = .95$.

Table S9 *Model-Estimated Standardized Reaction Times for Correct Day 2 Recognitions for Each Activity Type and for Changed Activities Conditioned on Change Recollection*

	Activity Type	
Age	Repeated Activities	Changed Activities
Younger	-0.03 [-0.19, 0.13]	-0.08 [-0.23, 0.06]
Older	-0.07 [-0.23, 0.08]	-0.10 [-0.24, 0.05]
	Classification for Changed Activities Only	
Age	Change Recollected	Change Not Recollected
Younger	-0.11 [-0.25, 0.03]	0.07 [-0.16, 0.29]
Older	-0.14 [-0.28, < 0.01]	0.05 [-0.14, 0.23]

Note: 95% confidence intervals are displayed in brackets.

6. Monitoring Resolution

6.1. Resolution for Day 2 Recall and Recognition of Changed Activities is Not Associated with Day 1 Consistent Predictions

We next examined if Day 1 Consistent predictions were associated with how well confidence judgments for memory measures discriminate between correct and incorrect Day 2 recall and recognition responses (i.e., monitoring resolution) for changed activities. For Day 2 recall, we first excluded all trials where participants did not give an answer. Confidence judgments were assessed in standard deviation units by transforming judgments across all responses into participant-level *z*-scores (zConfidence). This was necessary to control for differences in how participants used the confidence scale. We tested for differences in zConfidence using separate 2 (Age: younger vs. older) × 2 (Memory Accuracy: correct vs. incorrect) × 2 (Prediction: Day 1 consistent vs. not Day 1 consistent) models with activity as random intercept. As for all our other model, the random effect structure of models in this section was determined by applying the procedure of Matuschek and colleagues (2017). We were primarily interested in differences in monitoring resolution potential shown by Age and Prediction moderating confidence differences between correct and incorrect responses. We also report the main effect of Prediction to determine if this variable was associated with confidence judgments independently of other variables. We do not report the main effects of Age as the z-scoring made them difficult to interpret. Table S10 displays the model-estimated values.

The model for correct cued recall of changed Day 2 features (top rows) indicated significant effects of Memory Accuracy, $\chi^2(1) = 92.72$, $p < .001$, and Prediction, $\chi^2(1) = 67.74$, $p < .001$, and a significant Age × Memory Accuracy interaction, $\chi^2(1) = 7.01$, $p = .008$. However, both the Memory Accuracy × Prediction, $\chi^2(1) = 0.12$, $p = .73$, and the Age × Memory Accuracy × Prediction, $\chi^2(1) = 0.03$, $p = .86$, interactions were not significant. Pairwise comparisons showed that the different in confidence for correct and incorrect recalls was smaller for older adults, *t.ratio*(2540) = 5.08, $p < .001$, than younger adults, *t.ratio*(2535) = 8.62, $p < .001$.

The model for recognition of changed Day 2 features (bottom rows) showed a similar pattern: There were significant effects of Memory Accuracy, $\chi^2(1) = 81.90$, $p < .001$, and Prediction, $\chi^2(1) = 66.89$, $p < .001$, and a significant Age × Memory Accuracy, $\chi^2(1) = 21.54$, $p < .001$, but no significant Memory Accuracy × Prediction, $\chi^2(1) = 1.36$, $p = .24$, or Age × Memory Accuracy × Prediction interaction, $\chi^2(1) < 0.01$, $p = .96$. Pairwise comparisons showed that the different in confidence for correct and incorrect recalls was smaller for older adults, *t.ratio*(27.21) = 3.89, $p < .001$, than younger adults, *t.ratio*(2722) = 9.10, $p < .001$.

In sum, for both cued recall and recognition, both age groups made higher confidence judgments when Day 1 consistent predictions had occurred, regardless of memory accuracy. Also, confidence judgments discriminated better between correct and incorrect memory for changed Day 2 activity features for younger than older adults regardless of whether Day 1 consistent predictions had occurred.

Table S10. *Model-Estimated Standardized Confidence in Memory Accuracy for Changed Day 2 Features on Cued Recall and Recognition Tasks Conditioned on Prediction Type*

Cued Recall Task		Prediction	
Memory Accuracy	Age	Day 1 Consistent	Not Day 1 Consistent
Correct	Younger	0.31 [0.20, 0.41]	-0.08 [-0.21, 0.05]

	Older	0.32 [0.19, 0.46]	0.02 [-0.11, 0.15]
Incorrect	Younger	-0.20 [-0.33, -0.06]	-0.54 [-0.68, -0.41]
	Older	0.04 [-0.08, 0.16]	-0.25 [-0.36, -0.14]

Recognition Task		Prediction	
Memory Accuracy	Age	Day 1 Consistent	Not Day 1 Consistent
Correct	Younger	0.29 [0.19, 0.39]	-0.01 [-0.13, 0.10]
	Older	0.28 [0.17, 0.40]	0.01 [-0.10, 0.12]
Incorrect	Younger	-0.24 [-0.43, -0.06]	-0.63 [-0.79, -0.48]
	Older	0.13 [-0.02, 0.28]	-0.24 [-0.37, -0.13]

Note: 95% confidence intervals are displayed in brackets.

6.2. Resolution for Day 2 Recall and Recognition of Changed Activities is Associated with Memory for Change

Finally, we examined if memory for changes is associated with how well confidence judgments for memory measures discriminate between correct and incorrect Day 2 recall and recognition responses (i.e., monitoring resolution) for changed activities. We used a 2 (Age: younger vs. older) × 2 (Memory Accuracy: correct vs. incorrect) × 3 (Classification: recollected vs. vs. not recollected) models. Table S11 displays the model-estimated values. We do not describe effects redundant with those above.

The model for cued recall (top rows) indicated a significant Classification × Memory Accuracy interaction, $\chi^2(1) = 27.89$, $p < .001$. Pairwise comparisons showed that confidence judgments were significantly higher for correct than incorrect recalls when change was recollected, *t.ratio*(183) = 7.44, $p < .001$, but not when change was not recollected, largest *t.ratio*(101) = 1.18, $p = .24$. The Age × Memory Accuracy × Classification interaction was not significant, $\chi^2(1) = 2.28$, $p = .13$.

The model for recognition (bottom rows) indicated a significant Classification × Memory Accuracy interaction, $\chi^2(1) = 13.95$, $p < .001$. Pairwise comparisons showed that confidence judgments were significantly higher for correct than incorrect recognitions when change was recollected, *t.ratio*(2714) = 9.75, $p < .001$, and when it was not recollected, *t.ratio*(2697) = 2.42, $p = .02$, but the association was weaker when change was not recollected. The Age × Memory Accuracy × Classification interaction was not significant, $\chi^2(1) = 0.54$, $p = .46$. Together, these results show that recollecting change in both memory task was associated with better monitoring resolution abilities in both age groups.

Table S11 *Model-Estimated Standardized Confidence in Memory Accuracy for Changed Day 2 Features on Cued Recall and Recognition Tasks Conditioned on Change Classifications*

Cued Recall Task		Change Classification	
Memory Accuracy	Age	Change Recollected	Change not Recollected
Correct	Younger	0.24 [0.12, 0.36]	-0.22 [-0.40, -0.03]
	Older	0.36 [0.23, 0.49]	-0.09 [-0.24, 0.06]
Incorrect	Younger	-0.49 [-0.69, -0.28]	-0.32 [-0.45, -0.18]
	Older	-0.04 [-0.22, 0.13]	-0.13 [-0.25, -0.02]
Recognition Task		**Change Classification**	
Memory Accuracy	Age	Change Recollected	Change not Recollected
Correct	Younger	0.23 [0.14, 0.32]	-0.20 [-0.38, -0.03]
	Older	0.23 [0.13, 0.33]	-0.15 [-0.30, -0.007]
Incorrect	Younger	-0.41 [-0.56, -0.26]	-0.60 [-0.80, -0.41]
	Older	-0.11 [-0.25, 0.02]	-0.12 [-0.25, 0.02]

Note: 95% confidence intervals are displayed in brackets.