Classification: Biological Science – Neuroscience

Aging and the encoding of changes in events:

The role of neural activity pattern reinstatement

David Stawarczyk[1, 2], Christopher N. Wahlheim[4], Joset A. Etzel[1], Abraham Z. Snyder[3,5], Jeffrey M. Zacks[1,3]

[1] Department of Psychological & Brain Sciences, Washington University in St. Louis, St. Louis, MO 63105, USA.

[2] Department of Psychology, Psychology and Neuroscience of Cognition Research Unit, University of Liège, 4000 Liège, Belgium.

[3] Mallinckrodt Institute of Radiology, Washington University in St. Louis, St. Louis, MO 63110, USA.

[4] Department of Psychology, University of North Carolina at Greensboro, Greensboro, NC 27412, USA.

[5] Department of Neurology, Washington University in St. Louis, St. Louis, MO 63105, USA.

Correspondence concerning this article should be addressed to David Stawarczyk, Department of Psychology, University of Liège, Quartier Agora, place des Orateurs 1, 4000 Liège 1, Belgium. E-mail: d.stawarczyk@uliege.be. Tel.: 0032 4 3665979.

Abstract (250 words)

When encountering new events, memories of relevant past experiences can guide expectations about what will happen. When unexpected changes occur, this can lead to prediction errors with consequences for comprehension and subsequent memory. Current models of memory updating propose that people must first detect and register that features of the environment have changed, then encode the new event features and integrate them with relevant memories of past experiences to form configural memory representations. Aging could potentially impair each of these steps. Using functional MRI, we investigated these mechanisms in healthy young and older adults. In the scanner, participants first watched a movie depicting a series of everyday activities in a day of an actor's life. They next watched a second movie of the day's events in which some scenes ended differently. Crucially, before watching the last part of each scene, the movie stopped, and participants were instructed to mentally replay how the activity previously ended. Three days later, participants were asked to recall the activities. Individual differences in neural activity pattern reinstatement in the medial temporal lobes (MTL) and posteromedial cortex (PMC) during the mental replay phases of the second movie were associated with better memory for changed features in young adults. This was not the case in older adults for the MTL. This finding suggests that the MTL and PMC contribute to the comprehension of perceived changes by reinstating previous event features and that older adults are less able to use reinstatement to update memory for changed features.

Keywords: Representational similarity analysis, cognitive aging, event cognition, episodic memory, change comprehension

Significance Statement (120 words max)

The capacity to detect, encode, and later remember changes in perceived events is crucial for everyday functioning and, like many cognitive processes supported by long-term memory, could be particularly vulnerable to aging. Here, we examined the role of neural reinstatement of episodic memories in processing changes using multivariate pattern-based functional MRI during viewing of movies of everyday activities. Detecting and remembering event changes were both associated with reinstatement in medial temporal and posteromedial brain areas of the episode-specific neural activity pattern present while viewing the original event. This effect was smaller in older adults. These results show how episodic retrieval can inform ongoing event comprehension and how this process is more effective for young than older adults.

Aging and the encoding of changes in events:

The role of neural activity pattern reinstatement

Why do humans and other animals remember? One important reason is that features of past experiences can guide current behavior. Recent proposals suggest that a critical function of event memory (1)—also referred to as episodic memory (2–4)—is to guide anticipation of upcoming events (5–9). In most cases, using event representations of past experiences facilitates predictions in similar new situations. However, when events unexpectedly change, memory-based predictions can lead to errors. Such errors impose a short-term cost but may have long-term benefits for detecting and registering that features of the environment have changed, as well as for encoding the new event features and integrating them with relevant memories of past experiences to form configural memory representations (10). To reap these benefits, memory systems must update representations when things change.

Memory updating when experiencing changes has been found to depend crucially on interactions between the hippocampus, the surrounding medial temporal lobes (MTL), and the rest of the cortex (11, 12). Memory updating comprises several computational operations with different neural correlates and behavioral signatures (13, 14). These include *pattern completion*, which is the prediction function that activates relevant prior memories and knowledge based on environmental cues, *pattern separation* and *differentiation*, which keep features of the two experiences separate, and *integration*, which captures the relationships between different features of similar events. In order to integrate memory representations of events that are similar but include discrepant features, the brain needs to register the discrepancy and use it to prompt new learning. Computational models of memory updating propose that when things change, pattern completion

based on previous memories leads to prediction errors, which can drive new learning including integration processes to form configural memory representations (10, 15).

These proposals have been supported by behavioral and neuroimaging studies of the learning of word pairs and sequences of words or pictures (13, 14, 16–19). However, whereas most research on memory updating has used simple laboratory materials, real-world memory updating depends crucially on additional constraints and demands (20). Naturalistic event comprehension relies on a large set of processes working in parallel, including object recognition, interpretation of biological motion, spatial orienting, and the integration of multiple sensory modalities. Event comprehension is also constrained by specific knowledge about particular classes of events and how the world works—for example, when eating a banana, one peels it before eating it, and when a rock is dropped, it rolls downhill not uphill. These aspects of naturalistic activity mean that it exhibits correlations across features and across time that are more complex than those in laboratory materials. Memory systems that capitalize on this richer structure can predict more effectively, but prediction errors and updating of naturalistic activity may function quite differently than the updating of stimuli with simpler temporal and correlational structures. It is therefore important to characterize memory updating in the context of complex, naturalistic activity.

Memory-based prediction and updating may be selectively affected by healthy aging. Compared to younger adults, older adults are sometimes less able to use episodic memory to make explicit judgments about previous events and to predictively guide action (21, 22), and they are particularly prone to error when confronted with events that include overlapping features (23, 24). Such patterns could result from changes to any of several components of memory updating, including retrieval of relevant previous event features, prediction (pattern completion), registering

discrepancies between predicted and actual features (prediction error), or integration across similar events with discrepant features.

Behavioral experiments using movie stimuli suggest that when older adults encounter an event that begins similarly to a previous event but ends differently, they are less able than young adults to perform memory updating resulting in effective formation of configural memory representations (10). In these studies, older and young adults watched movies that included pairs of events that began identically but could end in one of two ways. For example, the actor might unroll a yoga mat and then perform either stretches or abdominal crunches. For both young and older viewers, the ability to detect change, and to remember it later along with the original activity feature, was associated with better memory for the changed features. Older adults detected and remembered fewer event changes, and this was associated with greater memory disruption when a change occurred. These results suggest that, when change is experienced, prediction based on episodic retrieval can drive new learning through the formation of configural memory representations. The data further suggest that this mechanism is less functional in older adults, but the behavioral data alone leave some uncertainty about where this breakdown occurs.

Neuroimaging data indicate that patterns of brain activity present while encoding new information are reinstated when this information is recollected, both for simple laboratory materials (e.g., 25–27) and for more complex stimuli such as movies of everyday activities (28–30). This effect is usually the strongest in the posterior areas of the default network (DN; 31, 32), more specifically part of the posteromedial cortex (PMC) that includes the posterior cingulate cortex (PPC) and retrosplenial cortex (Rsp), and in the medial temporal lobe (MTL) including the parahippocampal cortex (PHC) and hippocampus. These regions are sometimes referred to as the posterior medial system (33, 34) or contextual association network (35–37) due to their strong

involvement in long-term memory recollection, particularly when episodic representations of everyday events must be remembered from visual cues (20, 38).

The hippocampus shows large metabolic alterations and volume loss in aging (39), making functional change in the MTL a potential source of age-related differences in the episodic memory processes that enable the formation of configural memory representations when change is experienced. In addition, the PMC undergoes substantial metabolic and structural change in aging (40), with the integrity of its functioning related to better cognitive abilities in older adults (41). These considerations make the MTL and the PMC strong candidates for supporting the reinstatement of event features when a new event that is encountered is similar to a previous one. However, there is no evidence to date establishing whether the reinstatement of brain activity patterns facilitate the processing of changes in naturalistic events, nor is there evidence regarding how pattern reinstatement in these regions differs between older and young adults.

In the current study, we aimed to directly assess the role of retrieving episode-specific event features when encoding a new event that was similar to an earlier event. To do so, we used functional MRI (fMRI) in combination with representational similarity analysis (RSA; 42) to assess whether the reinstatement of brain activity patterns associated with past events can facilitate the processing of changes during the perception of new events in older and young adults. We used a task adapted from (10). During fMRI scanning, healthy young and older adult participants viewed two movies of discrete everyday activities, described as two days of an actor's life (hereafter referred to as Day 1 and Day 2). Together, the activities formed a narrative of the actor's day. Each activity was made up of two segments: an initial "cue" segment (6 s) that was always the same on Day 1 and Day 2, and an ending "post-divergence" segment (12 s) that either repeated or changed on Day 2, followed by a fixation interval of 4 to 5.3 s (see **Figure 1**). The Day 1 movie

(810 s) consisted of 45 activities. The Day 2 movie depicted activities that were either repeated exactly (15 activities) or began the same but ended differently (30 activities). We stopped each Day 2 movie activity after the initial cue segment (i.e., before any change had begun) and asked participants to mentally replay the activity ending (seen in the Day 1 movie) for 12 s. After this "reinstatement" phase, participants viewed the Day 2 ending of the activity. Participants were asked after each activity if they remembered what happened in the Day 1 movie when the Day 2 movie stopped, and whether the end of the clip that followed included a repeated or changed feature (a measure of change detection). We then used RSA to determine the extent to which brain activity patterns in PMC and MTL were similar between the Day 1 viewing and Day 2 reinstatement by computing a reinstatement score for each activity and participant.

Three days later, participants were given an unscanned cued recall test for the features of activities viewed in the Day 2 movie (see **Figure 1**). In this memory test, participants were first asked about a feature of the activity (e.g., "What did the actor do on the exercise mat?"), and then were asked whether that activity repeated between Day 1 and Day 2 or had changed, and their confidence in their answers. If they reported that the activity had changed, they were asked to recall the Day 1 feature and their confidence in the recollection.

We hypothesized that stronger Day 2 reinstatement of the MTL and PMC activity patterns that occurred while watching the Day 1 movie would be associated with better change detection while viewing the Day 2 movie, better subsequent recall of the changed features, and better subsequent recollection that a change had occurred (including what the changed feature had been in the Day 1 movie). Furthermore, given the decline of MTL and PMC functional integrity with aging, we expected that these associations between reinstatement and change detection, memory for changed activity features, and memory for change itself would be stronger for young than older

adults and that this difference might be partly explained by older adults' diminished ability to retrieve event memories in the service of creating configural memory representations.

Results

All analyses of memory performance and neural pattern reinstatement effects were computed using mixed effects models with subjects and activities as random effects. Logistic models were used when the dependent variable was binary (see the SI appendix, Section 1.8 for more details).

Memory performance and change classifications

During Day 2 viewing in the scanner, participants were asked after each activity whether they had successfully reinstated the Day 1 activity ending and whether its Day 2 ending was the same as on Day 1 or had changed (Figure 1). Older adults reported higher rates of reinstatement success than did younger adults but were less accurate at detecting that the activities had changed. In younger but not older adults, self-reported reinstatement was associated with change detection accuracy. These results suggest that older adults were overconfident compared to younger adults and less able to use subjective features of reinstatement to detect changes (see SI Appendix section 2.1.1 for a detailed statistical report).

We examined performance in the cued recall task three days later to determine whether remembering change and the original Day 1 activity was associated with better memory for the changed Day 2 features, and whether this was affected by age. In the memory test, participants attempted to recall event features from the Day 2 movie, and then were asked whether that event had changed from Day 1 to Day 2. When participants indicated that an event had changed, they were asked to recall the Day 1 feature. Change could therefore be remembered with recall of Day

1 features (change recollected), remembered without recall of Day 1 features (change remembered, but not recollected), or not remembered at all.

Models including fixed effects of age group and activity type indicated that older adults recalled fewer Day 2 features than young adults [$\chi^2(1) = 9.61$, $p = .002$], and that this effect of age did not differ between repeated and changed activity types [$\chi^2(1) = 0.37$, $p = .54$] (see **Figure 2**). To examine the association between change recollection and Day 2 recall, we fit another model with fixed effects of age group and activity type but with additional levels for changed activity corresponding to each type of memory for change. Both age groups recalled changed activities less accurately than repeated activities when change was not remembered at all or remembered, but not recollected (*smallest z* ratio = -7.64, $p < .001$). However, when participants could recollect change, recall performance was higher than for repeated activities (*z* ratio = 5.95, $p < .001$). Older adults showed lower estimated probabilities of change recollection than younger adults (Older = .27, *95% CI* = [.19, .36]; Younger = .38, *95% CI* = [.29, .48]), [$\chi^2(1) = 4.70$, $p = .03$]. Thus, although both age groups showed proactive facilitation in recall of Day 2 features when change was recollected, older adults experienced this benefit on fewer trials.

During Day 2 viewing, each activity was paused, and participants were asked to try to reinstate how it had ended on Day 1. After seeing the changed or repeated activity ending, they reported whether they had been able to reinstate successfully. For both young and older adults, self-reported successful reinstatement of Day 1 features was associated with better recall of the changed Day 2 features. However, this effect did not remain significant after controlling for the ability to recollect change, indicating that self-reported successful reinstatements predicted more correct recall of changed Day 2 activities solely because they were also predictive of a better memory for change and the Day 1 activity feature during the cued recall test (see SI appendix,

Section 2.1.1). When participants misremembered a changed activity, they were likely to intrude features from the previous corresponding Day 1 activity. Analyses of intrusion rates generally mirrored the rates of correct Day 2 recall (see SI appendix, Section 2.1.2).

Representational similarity measures of neural memory reinstatement

To assay the neural reinstatement of activity-specific Day 1 features during Day 2 we compared the pattern of activity in the MTL and PMC during the viewing of each activity on Day 1 to the pattern during its attempted reinstatement (see **Figure 1**). Patterns were analyzed at the level of parcels in the 17 networks/300 parcels cortex parcellation map from Schaefer et al. (43), and results averaged across the parcels making up the PMC and MTL (see **Figure S1**).

To extract patterns for each Day 1 post-divergence segment and Day 2 reinstatement attempt for each participant we averaged the BOLD signal for each voxel over the ninth to fourteenth scans (11.97 to 18.62 s) after the beginning of the cue segment for each activity; this corresponds to the interval of the segment plus a shift to account for hemodynamic lag. We then computed the Pearson correlation between the averaged patterns for every possible pair of activity segments within each parcel, both matching (e.g., viewing activity #2 and reinstating activity #2) and not (e.g., viewing activity #2 and reinstating activity #3) to form a similarity matrix. Finally, we calculated "*reinstatement scores*" for each activity and parcel as the difference between the correlation on each matching activity minus the mean correlation on mis-matching activities (see **Figure 3**). Reinstatement scores represent represents the degree to which reinstatement activation patterns are more similar to their *matching* Day 1 activity encoding activation pattern than to the others (30, 44). Reinstatement scores for each parcel within each ROI were then averaged, resulting in two scores for each participant, one for the MTL and one for the PMC.

These reinstatement scores tended to be positive, indicating that participants were able to reinstate activity-specific brain activation patterns (see **Figure 4A**). Collapsing across age groups, one-sample *t*-tests compared against zero indicated significant reinstatement effects in both MTL [$t(61) = 3.83$, $p < .001$] and PMC [$t(61) = 6.25$, $p < .001$]. To determine whether there were age differences in reinstatement scores, we fitted models with age group as a fixed effect. Neither region showed significant age differences [PMC: $\chi^2(1) = 0.55$, $p = .56$; MTL: $\chi^2(1) = 1.29$, $p = .26$]. Follow-up analyses on the individual parcels within the Schaefer et al. (43) cortex parcellation map indicated significant reinstatement effects in many parcels for both age groups (see **Figure S2**), only one parcel within our two ROIs showed a significant age difference in the reinstatement effect (Left MTL parcel 144; see **Table S1**).

We then fitted models with between- and within-participant reinstatement scores for changed activities, as well as their respective interaction with age group, as fixed effects. Between-participant reinstatement scores were mean reinstatement scores for the changed activities of each participant replicated across all changed trials for that participant. These allowed us to examine whether participants who reinstated Day 1 neural activity patterns more strongly also had better recall of Day 2 features. Within-participant reinstatement scores were computed by centering reinstatement scores for the changed activities within each participant (i.e., the mean of reinstatement scores from all changed trials of that participant was subtracted to the reinstatement score of each changed trial). These allowed us to examine whether Day 2 trials with higher reinstatement (independently of the mean reinstatement of the participants) were associated with better subsequent memory for Day 2 features. In addition, because reinstatement scores across parcels were only moderately correlated (**Table S2**), we performed similar analyses but examined the effect of each individual PMC and MTL parcel on Day 2 recall performance above and beyond

the effects of all the other parcels in the ROI. This procedure allowed us to examine whether reinstatement in specific parts of the PMC or MTL (such as the hippocampus) predicted behavioral performance. For conciseness, only summarized results of these latter analyses are presented below; see the SI Appendix, Sections 2.2.2 and 2.2.3 for a more detailed statistical report.

Across participants, differences in neural pattern reinstatement were associated with successful encoding of subsequent event endings, and with the subjective experience of reinstating a previous event ending.

We first examined whether RSA reinstatement scores were correlated with self-reported reinstatement success across participants (see **Figure 4B**). Participants with higher mean PMC reinstatement scores rated more of their reinstatements as successful [$\chi^2(1) = 5.00$, $p = .03$], an effect that did not interact with age group [$\chi^2(1) = 2.11$, $p = .15$], though it was driven by different PMC parcels in each age group (see SI Appendix, section 2.2.2). There was also a significant effect of mean reinstatement score at the between-participant level for the MTL [$\chi^2(1) = 8.40$, $p = .004$] that did not interact with age group [$\chi^2(1) = 1.32$, $p = .25$]. This effect was not driven by specific MTL parcels. Further, RSA reinstatement scores were correlated with immediate change detection performance (asked after every trial in the scanner) in younger adults but not in older adults (see SI Appendix section 2.2.1), suggesting that older adults were less able to use reinstated event features to detect that a change had occurred.

To examine between-participant differences in the relationship between reinstatement and memory accuracy during the cued recall task, we correlated the mean RSA reinstatement score for the changed activities with subsequent Day 2 recall performance (see **Figure 4B**). Participants with higher mean PMC reinstatement scores had higher rates of correct Day 2 recall [$\chi^2(1) = 4.72$ $p = .03$] and this effect did not interact with age group [$\chi^2(1) = 2.01$ $p = .16$]. For the MTL, between-

participant mean reinstatement scores did not predict Day 2 Recall collapsed across age groups [$\chi^2(1) = 3.54$ $p = .06$], but reinstatement scores did interact with age group [$\chi^2(1) = 4.09$ $p = .04$]. This showed that the association between reinstatement scores and recall of changed features was stronger in young than older adults. Follow-up tests showed that higher mean MTL reinstatement scores were associated with better recall for young adults [$\chi^2(1) = 8.61$ $p = .003$] but not for older adults [$\chi^2(1) = 0.18$ $p = .67$]. Examining the effects of individuals parcels revealed that these between-participant effects were not driven by specific PMC or MTL parcels. See SI Appendix, Section 2.2.1 for the analyses on Day 1 intrusions during Day 2 recall of changed activities.

We next examined whether RSA reinstatement scores predicted change recollection (see **Figure 4B**). Given that Day 2 recall performance was comparable when change was remembered without Day 1 recall and when change was not remembered at all (see **Figure 2**), we collapsed these cells here and in subsequent analyses into the category *change not recollected*. Consistent with Day 2 recall, mean between-participant PMC reinstatement was positively associated with change recollection [$\chi^2(1) = 4.17$, $p = .04$], but the effect was qualified by an interaction with age group [$\chi^2(1) = 5.26$, $p = .02$]. There was a significant positive association for young adults [$\chi^2(1) = 8.36$, $p = .003$] but not for older adults [$\chi^2(1) = 0.31$, $p = .86$]. Analyses of individual parcel scores revealed that the effect in the young was driven by specific parcels (see SI Appendix, section 2.2.2). Similar between-participant associations were found for the MTL as there was a significant effect of mean reinstatement score [$\chi^2(1) = 4.98$, $p = .03$] that was qualified by a significant interaction with age group [$\chi^2(1) = 3.86$, $p = .049$]. There was a significant positive association for young adults [$\chi^2(1) = 7.24$, $p = .007$] but not for older adults [$\chi^2(1) = 0.54$, $p = .46$]. Analyses of change detection during Day 2 viewing and Day 1 intrusions during Day 2 recall of changed activities are reported in the SI appendix, Section 2.2.1.

Finally, because between-participant differences in mean PMC and MTL reinstatement scores for young adults were positively related to both Day 2 recall of changed features and to change recollection, we examined whether the association between reinstatement scores and Day 2 recall accuracy for the changed activities could be explained by change recollection. These models included mean between-participant change recollection and reinstatement scores as fixed effects (see **Figure S3** for the associations between mean reinstatement scores and Day 2 recall, and for the unique contribution of change recollection to Day 2 recall). With both reinstatement scores and change recollection in the model as fixed effects, neither PMC nor MTL reinstatement significantly predicted Day 2 recall [PMC: $\chi^2(1) = 1.33$, $p = .25$; MTL: $\chi^2(1) = 3.25$, $p = .07$]. However, change recollection was still positively associated with Day 2 recall [PMC: $\chi^2(1) = 147.90$, $p < .001$; MLT: $\chi^2(1) = 147.65$ $p < .001$]. Neither interaction was significant (both *ps* > .05). This is consistent with the possibility that the effect of reinstating of Day 1 activities during Day 2 viewing on Day 2 recall is mediated by recollecting that the activity changed. However, the current design does not allow for strong causal conclusions about this potential mediation.

Within participants, event-by-event differences in brain activity pattern reinstatement were associated with successful encoding of subsequent event endings only for a subset of cortical parcels

Between-participant analyses informed us about how individual differences in reinstatement scores were related to behavioral performance. To assay whether activity-to-activity variation in reinstatement within an individual was related to subsequent memory, we conducted a second set of analyses. Within each participant, we mean-centered the reinstatement scores for the changed activities and correlated these with Day 2 recall for individual items. Neither the effect of mean reinstatement scores in the PMC and MTL nor the interaction with age group was

significant (all *p*s > .05; see **Figure S4**). However, analyses of individual parcels revealed that reinstatement scores in two PMC parcels were positively associated with Day 2 recall whereas two other parcels showed the opposite effect (see SI Appendix, Section 2.2.3). None of the effects of individual parcels in the PMC were qualified by an interaction with age group. There was no significant effect of individual parcel reinstatement scores nor was there an interaction with age group in the MTL. No significant effects were found for either the mean PMC or MTL or individual parcel reinstatement scores regarding change recollection, self-reported reinstatement success, change detection accuracy, or Day 1 intrusion during Day 2 recall (all *p*s > .05; see **Figure S4**). Thus, apart from Day 2 recall of changed activities there was no evidence for a within-participant relationship between neural pattern reinstatement and behavioral performance in either age group.

Mass univariate analyses

To further investigate age differences in the processing of changed Day 2 activities during viewing, we performed mass univariate fMRI analyses using general linear models, as described in the SI appendix, Section 1.6. Many models of memory updating propose that novelty detection and prediction error are critical components of memory updating when encountering changes (10, 15), so we specifically examined the difference in the neural response between viewing of changed and repeated activity endings (the "post-divergence" segments), and whether age group moderated this response. We were particularly interested in neural responses in the anterior hippocampus because this region is commonly associated with the processing of change and novelty (46). Across all participants, we found more activity for changed than repeated endings in clusters mainly located the lateral prefrontal cortex and, at a lower threshold, in the bilateral anterior hippocampus (see **Figures S5 and S6**). A two-sample *t*-test showed that this neural response did not differ

between age groups (no cluster was activated above threshold). In addition, neural activity was not parametrically modulated by RSA reinstatement scores during the changed event epochs in any brain region for either age group (again, no cluster was activated above threshold).

To further assess the proposal that hippocampal responses to novelty are a critical component of memory updating when experiencing changes in events (10, 15), We next examined whether between-participant differences hippocampal response intensity while viewing changed versus repeated activity endings predicted reinstatement scores, behavioral memory measures, self-reported reinstatement, and change detection accuracy during Day 2 viewing, and whether age group moderated any of these possible relationships. Results showed no association between hippocampal response intensity and either PMC or MTL reinstatement scores. At the behavioral level, hippocampal responses did not predict any of the measures of memory for change, nor self-reported reinstatement success, but were positively associated with change detection during Day 2 viewing. This latter effect did not interact with age (see SI Appendix, Section 2.3 for details).

Finally, an unexpected finding of the RSA analyses was that reinstatement scores did not differ across age groups. To examine whether there might nonetheless have been age-related differences in how participants initially perceived the activities, we examined whether there were group differences in brain activity while encoding the activity features during the Day 1 viewing. Across all participants, these analyses revealed the large set of brain regions typically activated while watching movie-stimuli (see **Figure S7**). A two-sample *t*-test revealed that in no region did mass univariate activation level differ between the two groups (no cluster was activated above threshold); however, a pattern classifier indicated that the voxelwise distribution of activation in the PMC and MTL differed between young and older adults (see SI Appendix section 1.7 and 2.4). Thus, although young and older adults did not differ in the extent to which they were able to

reinstate the neural activity patterns formed during the Day 1 viewing during the Day 2 reinstatement phases (see Figure 4A), these Day 1 neural activity patterns during viewing nonetheless differed between the two groups.

Discussion

In this study, attempting to reinstate features of a relevant previous event during comprehension of a new one was associated with widespread reinstatement of fMRI activity patterns corresponding to anticipated features of the event. As hypothesized, reinstatement was associated with better subsequent memory for changed event features. This was true at the between-participant level for the PMC and MTL regions as wholes, and at the within-participant level, it held for only a subset of PMC regions. In the PMC, it did not differ significantly between young and older adults; in the MTL it was significantly stronger in young adults and was not robust in older adults. When participants attempted to recall changed event features, the ability to recall *what* the feature had changed to was associated with being able to recollect *that* the feature had changed and to report what it had been previously. In addition, the positive association between neural pattern reinstatement and memory for changed activity features was statistically explained by the ability to recollect the fact that the activity had changed and the original event feature. A similar pattern was seen for participants' introspective judgments about whether they had successfully reinstated event features before viewing each event's ending.

Neural measures of pattern reinstatement in the PMC and MTL were associated with a more accurate recollection of new information presented *after* the end of the reinstatement phase that *conflicted* with the previously encoded features that were reinstated. This finding is consistent with previous studies showing that reinstatement, during recollection, of the brain activity pattern present in posterior DN areas while watching movies can predict memory for the movie content

up to one week later (28–30), and with evidence for neural pattern reinstatement when rehearsing learned associations to pictures (14). The current results indicate for the first time that such reinstatement is related to the encoding of novel, unexpected event features. One possibility is that reinstatement leads to predictions, which in turn lead to a prediction error signal when events change, leading to memory updating (16). In a previous study using sequences of pictures, updating manifested as selective forgetting, or *pruning*, of previously encoded features (16), whereas here memory updating was associated with better memory for previous as well as new event features. We attribute this difference to the formation of a configural representation composed of the original activity features, the changed features, and their temporal relations (10). Consistent with this idea, both neural and behavioral measures of reinstatement before encoding the change were associated with being able to recollect how the activity had changed, and neither self-reported nor neural pattern reinstatement remained significant predictors of memory accuracy for the changed features after controlling for change recollection accuracy. Forming a configural representation is related to *integration* of separated experiences into a common context, which is also associated with cortical reinstatement (14). It can be contrasted with *differentiation*, in which overlapping features are selectively deleted; differentiation acts to make experiences more distinctive rather than to merge them into a complex (19).

The PMC may play a key a role in supporting the event model representations from which predictions are generated. The PMC is part of the DN, and it was initially thought to be primarily involved in generating internal mentation that stands in opposition with attention to the external world (47). However, there is now substantial evidence that the PMC also supports externally-directed attention and event comprehension when task performance and the processing of perceptual inputs can benefit from relevant information stored in memory (44, 48). In addition,

recent studies have revealed that the transitions between activity patterns within the PMC while watching movies follow timescales ranging from seconds to minutes, closely matching how people segment movie content into distinct events (49). This supports the view that event model representations might be the means by which PMC facilitates interactions with the external world. As for the MTL, there is extensive evidence that this region is involved in the relational binding of information stored in memory and how it relates to perceptual inputs in order to form associative memory representations of everyday experiences (50, 51). Consistent with these findings, recent research suggests that peaks of activity in the hippocampus at the transition between perceived events can predict neural reinstatement in the PMC during recall (49, 52). Pattern reinstatement within the MTL in the current study might therefore reflect the relational binding of information stored in memory—whose retrieval is triggered by the cue segment—in order to form the event model that is supported by the PMC. This event model would then help to detect and encode the changes that are experienced when ongoing experiences do not match model-based predictions.

Older adults were less likely than their younger counterparts to detect changes and to recall the original activity features during the cued recall task. Older adults also showed a reduced association between neural activity pattern reinstatement in the MTL and subsequent memory. One possibility is that older adults are less able to use retrieved event features to form an updated configural representation that includes the changed event features and their relationship to the previous features. Previous studies of age-related deficits in change comprehension (10, 53) and in associative memory (54, 55) are consistent with this idea. A possible explanation for these findings is suggested by behavioral studies showing that the individuation of events during perception is impaired in aging (56, 57): The event representations formed by the older adults during the original viewing of the activities may have been less detailed than those of younger

adults. The finding that the spatial pattern of activation during movie viewing differed across groups is consistent with this possibility. As a result, they may have been less successful in forming configural representations when confronted with changed activities, and thus less successful in encoding and remembering changes. Although speculative, this proposal might explain why self-reported reinstatement accuracy in older adults did not predict better change detection during Day 2 viewing and was not associated with fewer Day 1 intrusions (See SI Appendix, Section 2.1.2) nor correct Day 2 recognition of the changed activity features (See SI Appendix, Section 3.2), as was the case for the young adults.

In any functional neuroimaging study comparing young and older adults, it is important to consider potential sources of artifact; these include group differences in neurovascular coupling, in head motion, and in how the tasks are approached (58). In the present case, the fact that older adults showed robust overall neural reinstatement renders their significantly weaker relationships between neural reinstatement and behavioral memory measures particularly striking. Another caveat that participants were instructed to attempt to reinstate previous event features while Day 2 activities were paused. Cognitive age differences are often larger when under time pressure and when self-initiated processing is required (59); it may be that the time to deliberately reflect ameliorated age differences. To generalize to naturalistic comprehension, it will be important to use converging measures that do not depend on strategic, interruptive task instructions.

In conclusion, the present results showed that the reinstatement of previously seen related events can facilitate the encoding of new, changed event features. This is particularly striking because the new features conflict with the just-retrieved features of the previous activity. We propose that retrieving activity features facilitates encoding precisely because it enables the system to register discrepancy between the predicted event features and the encountered features. That

discrepancy can drive formation of a configural representation that includes the old features, the new ones, and the relationship between the two. This process was impaired in older adults; the pattern of impairment suggested that deficits in encoding a detailed memory representation of the original event might reduce older adults' ability to encode a configural representation of the changed event that includes its relationship to the previous event.

Method

The full stimulus sets for the materials used in the present experiments, anonymized data files, coded data, and R Markdown files (60) containing the analysis scripts are available on the Open Science Framework: (https://osf.io/v3dqg/).

Participants

This study was approved by the IRB of Washington University in St. Louis, and all participants gave their written informed consent before participating in the study. Participants were recruited from the Washington University School of Medicine Research Participant Registry, flyers posted on campus, and word of mouth. Potential participants were initially contacted by phone for a prescreening interview. The sample included 62 healthy right-handed participants: 34 young adults (mean age 22.85 years, *SD* = 2.71, range: 18-27 years, 22 females) and 28 older adults (mean age 69.86 years, *SD* = 5.01, range: 65-84 years, 20 females). All older adults had a score of 25 or above (*M* = 29.25, *SD* = 0.87, range: 27-30) on the Mini-Mental State Exam (MMSE; 61). For more details about recruitment and exclusion criteria, see the SI Appendix Section 1.1.

Materials, Design, and Procedure

The materials were movies of a female actor performing daily activities on two fictive days in her life, which were described to participants as "Day 1" and "Day 2" (10). There were two versions (A and B) of each activity that differed on a thematically central feature (e.g., doing

stretching or sit-ups on a yoga mat, see **Figure 1**). Each activity began with a 6 s initial cue segment that was consistent across versions. Those segments were followed by a post-divergence segment (12 s) that included features from one of two endings (version A or B). The complete stimulus set consisted of 45 pairs of activities. The sequence of activities during each fictive day was fixed and included all 45 activities—beginning with the actor waking up and ending with her going to bed. The version of the activity that participants saw in the Day 1 movie (A or B) and whether the activity repeated or changed in the Day 2 movie were both counterbalanced across participants. This resulted in 12 experimental formats.

Participants completed the task in two sessions separated by three days (see **Figure 1**). During the first session, participants watched both movies while lying in the scanner. As noted in the Introduction, activities in the Day 1 movie (810 s duration) appeared as a continuous 18 s clips that were each followed by a fixation interval of 4 to 5.3 s (M = 4.6 s). After participants viewed the Day 1 movie in a first functional run, field map images and a high-resolution anatomical image were collected, taking approximately 10 minutes. Participants then watched the Day 2 movie (1350 s) during a second functional run. All Day 2 activities were stopped for 12 s after their initial cue segments. During that time, participants were asked to mentally replay the Day 1 ending for each activity. Participants were then shown the Day 2 ending, which was either identical to the Day 1 ending (repeated activities) or included novel features (changed activities). They were then asked to indicated using a button box whether they successfully replayed the Day 1 features when the movie was stopped, and whether the features changed from Day 1 to Day 2. After Day 2 viewing, we collected a second set of field map images and a high-resolution T2-weighted image, taking approximately 6 minutes. We then informed participants that that the session was over.

During Session 2, which took place outside of the scanner, we first tested participants' memory for the previously viewed activities using a cued recall task (see **Figure 1**). We first asked participants to recall the Day 2 activity feature (e.g., "What did the actor do on the exercise mat?"), and then to indicate whether that feature changed from the Day 1 to Day 2 movie. When participants indicated that the activity had changed, they were asked to recall the Day 1 feature. Participant were asked to rate how confident they were in their responses for Day 2 recall, change classification, and Day 1 recall for activities classified as changed. The recall cues for each activity appeared in the same order as the activities during each movie. Following the cued recall task, we administered a recognition test of Session 1 activities (see the SI Appendix, Section 3). After the recognition test, all participants completed a vocabulary test (62), and older adults completed the MMSE (61) last. The three-day retention interval was chosen based on pilot testing, to avoid floor and ceiling effects on the memory measures. All stimulus materials for the movies, cued recall task, and recognition task were presented using E-Prime 2 software (Psychology Software Tools, Pittsburgh, PA, USA; 3). A more detailed description of the procedure, materials, and instructions given to the participants is provided in the SI Appendix (Sections 1.2 and 1.3).

fMRI Data Analyses

Because we had strong *a priori* hypotheses regarding the brain regions that would be relevant in the RSA, we employed a ROI-based analytic strategy. Specifically, we selected the PMC and MTL parcels of the DN subsystems from the 17 networks/300 parcels cortex parcellation map of Schaefer et al. (43). We resampled the parcellation to atlas space using the 3dresample AFNI command and warped it from MNI to our template using the "Old Normalise" command in SPM 12. The resulting ROIs are presented in **Figure S2**. In addition, as the parcellation of Schaefer et al. (43) does not map subcortical areas, we added two parcels for the left and right hippocampus

using the Oro.nifti package (64) in R. These hippocampal parcels were obtained from the minimal preprocessing pipelines for the Human Connectome Project (65) using the "Cifti Separate" workbench command.

Following spatial preprocessing (see the SI Appendix, section 1.5), data were normalized and detrended using second-order polynomials with the 3dDetrend AFNI command, spatially smoothed (66) with a Gaussian kernel of 3-mm full-width at half maximum in SPM, and z-scored with PyMVPA (67). To summarize the activity within each voxel during the period of interests in each run, we performed temporal compression by averaging the ninth to fourteenth scans (11.97 to 18.62 s) after the beginning of each activity using the 3dTstat AFNI command. This averaging window corresponded to the post-divergence segments in the first run and reinstatement phases in the second run. It was chosen to account for the hemodynamic delay, and to reduce the chance of contamination from the blood-oxygen-level-dependent (BOLD) response associated with the fixation cross of the first run and post-divergence segment of the second run, since these started 18 s after the onset of the cue segment. This temporal compression procedure resulted in one brain image for each activity, run, and participant.

We then compared the similarity of the brain activity patterns in each parcel between the two runs (see **Figure 3**). Specifically, a similarity matrix was constructed for each parcel within each participant by placing the 45 activities from the first run along one axis and the same 45 activities from the second run along the other axis, then computing all possible pairwise Pearson correlations (i.e., each Day 1 post-divergence segment correlated with each Day 2 reinstatement). If participants successfully reinstated the pattern of neural activity experienced while watching the post-divergence segments of the first run during the reinstatement phases of the second run, then we would expect the correlations along the diagonal (i.e., watching and reinstating same activity

pairs) to be higher than off-diagonal correlations (i.e., watching and reinstating different activity pairs). This tendency was quantified by computing a "reinstatement score" for each activity: the on-diagonal cell minus the average of off-diagonal cells (after Fisher's r to z transformation; 31, 56) that reflects neural reinstatement specificity of post-divergence segments from the first run during the reinstatement in the second run.

Cued Recall Response Coding

All open-ended responses made during the cued recall task (i.e., answers to the questions about Day 2 activity features and about Day 1 activity features for activities classified as changed) were coded by the first author; a second independent rater also coded the responses, which allowed us to compute interrater agreement. Participants' Day 2 recall responses were classified as one of three types. *Correct Day 2 recall* responses were correct descriptions of the criterial Day 2 activity feature. *Day 1 intrusions* were responses that included the Day 1 criterial feature that did not appear on Day 2. All remaining responses were considered as *Incorrect*. The Cohen's κ (68) between raters was .85. When participants classified activities as changed on the Day 2 recall test, they were asked to recall the Day 1 feature. The same raters coded these responses as *correct Day 1 recalls* and *incorrect Day 1 recalls* depending on whether the participants recalled the critical Day 1 activity feature or not, respectively. Cohen's κ for Day 1 recalls was .86.

Acknowledgements

We thank Ryan Kahle, Madeleine Schroedel, and Priscilla Mei for assistance with data collection and coding. We also thank Aaron B. Tanenbaum for his help with spatial preprocessing and selection of the fMRI sequences. This project was funded by NIH grant R21AG05231401 and supported in part by the Neuroimaging Informatics and Analysis Center (1P30NS098577). David Stawarczyk is currently supported by a Marie Curie Global Fellowship grant from the European Commission (Grant Agreement number: 798109). Joset A. Etzel was partially supported by the National Institutes of Health, grant number R37MH066078.

Author contribution

DS, CW, JE, and JZ designed the study. DS performed the study. DS, JE, and JZ analyzed the data. DS, CW, and JZ wrote the paper with all other co-authors providing critical inputs. AS helped with the selection of the fMRI sequence and provided the spatial preprocessing pipeline used in the present work. All authors approved the final version of the manuscript.

References

1. D. C. Rubin, S. Umanath, Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychol Rev* **122**, 1–23 (2015).

2. E. Tulving, Episodic memory: from mind to brain. *Annu Rev Psychol* **53**, 1–25 (2002).

3. A. Baddeley, J. P. Aggleton, M. A. Conway, *Episodic memory: New directions in research* (Oxford University Press; US, 2002) https:/doi.org/10.1093/acprof:oso/9780198508809.001.0001.

4. E. Dere, A. Easton, L. Nadel, J. P. Huston, Eds., *Handbook of Episodic Memory* (Elsevier, 2008).

5. M. Bar, E. Aminoff, M. F. Mason, M. Fenske, The units of thought. *Hippocampus* **17**, 420–428 (2007).

6. L. W. Barsalou, Simulation, situated conceptualization, and prediction. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **364**, 1281–1289 (2009).

7. D. L. Schacter, *et al.*, The future of memory: remembering, imagining, and the brain. *Neuron* **76**, 677–94 (2012).

8. S. B. Klein, The temporal orientation of memory: It's time for a change of direction. *Journal of Applied Research in Memory and Cognition* **2**, 222–234 (2013).

9. D. Stawarczyk, M. A. Bezdek, J. M. Zacks, Constructing event representations: The role of the midline default network core. *Topics in Cognitive Science* (Submitted).

10. C. N. Wahlheim, J. M. Zacks, Memory guides the processing of event changes for older and younger adults. *Journal of Experimental Psychology: General* **148**, 30–50 (2019).

11. L. Nadel, A. Hupbach, R. Gomez, K. Newman-Smith, Memory formation, consolidation and transformation. *Neuroscience & Biobehavioral Reviews* **36**, 1640–1645 (2012).

12. M. L. Schlichting, A. R. Preston, "The Hippocampus and Memory Integration: Building Knowledge to Navigate Future Decisions" in *The Hippocampus from Cells to Systems: Structure, Connectivity, and Functional Contributions to Memory and Flexible Cognition*, D. E. Hannula, M. C. Duff, Eds. (Springer International Publishing, 2017), pp. 405–437.

13. D. Clewett, L. Davachi, The ebb and flow of experience determines the temporal structure of memory. *Current Opinion in Behavioral Sciences* **17**, 186–193 (2017).

14. F. R. Richter, A. J. H. Chanales, B. A. Kuhl, Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. *NeuroImage* **124**, 323–335 (2016).

15. K. A. Norman, E. L. Newman, G. Detre, A neural network model of retrieval-induced forgetting. *Psychological Review* **114**, 887–953 (2007).

16. G. Kim, J. A. Lewis-Peacock, K. A. Norman, N. B. Turk-Browne, Pruning of memories by context-based prediction error. *Proceedings of the National Academy of Sciences* **111**, 8997–9002 (2014).

17. G. Kim, K. A. Norman, N. B. Turk-Browne, Neural differentiation of incorrectly predicted memories. *J. Neurosci.* **37**, 2022–2031 (2017).

18. S. DuBrow, L. Davachi, Temporal memory is shaped by encoding stability and intervening item reactivation. *Journal of Neuroscience* **34**, 13998–14005 (2014).

19. S. E. Favila, A. J. H. Chanales, B. A. Kuhl, Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nature Communications* **7** (2016).

20. K. L. Bottenhorn, *et al.*, Cooperating yet distinct brain networks engaged during naturalistic paradigms: A meta-analysis of functional MRI results. *Network Neuroscience* **3**, 27–48 (2018).

21. R. T. Zacks, L. Hasher, K. Z. H. Li, "Human memory" in *The Handbook of Aging and Cognition (2nd Ed)*, (Lawrence Erlbaum Associates, 2000), pp. 293–357.

22. D. A. Balota, P. O. Dolan, J. M. Duchek, "Memory changes in healthy older adults" in (Oxford University Press, 2000), pp. 395–410.

23. M. K. Johnson, Memory and reality. *American Psychologist* **61**, 760–771 (2006).

24. S. R. Old, M. Naveh-Benjamin, Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging* **23**, 104–118 (2008).

25. J. F. Danker, J. R. Anderson, The Ghosts of Brain States Past: Remembering Reactivates the Brain Regions Engaged During Encoding. *Psychological Bulletin* **136**, 87–102 (2010).

26. M. E. Wheeler, S. E. Petersen, R. L. Buckner, Memory's echo: vivid remembering reactivates sensory-specific cortex.[Erratum appears in Proc Natl Acad Sci U S A. 2004 Apr 6;101(14):5181]. *Proc Natl Acad Sci U S A* **97**, 11125–9 (2000).

27. A. M. Gordon, J. Rissman, R. Kiani, A. D. Wagner, Cortical reinstatement mediates the relationship between content-specific encoding activity and subsequent recollection decisions. *Cereb Cortex* **24**, 3350–64 (2014).

28. J. Chen, *et al.*, Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience* **20**, 115–125 (2017).

29. C. M. Bird, J. L. Keidel, L. P. Ing, A. J. Horner, N. Burgess, Consolidation of Complex Events via Reinstatement in Posterior Cingulate Cortex. *J Neurosci* **35**, 14426–34 (2015).

30. C. S. H. Oedekoven, J. L. Keidel, S. C. Berens, C. M. Bird, Reinstatement of memory representations for lifelike events over the course of a week. *Scientific Reports* **7**, 14305 (2017).

31. M. E. Raichle, *et al.*, A default mode of brain function. *Proc Natl Acad Sci U S A* **98**, 676–82 (2001).

32. D. A. Gusnard, M. E. Raichle, Searching for a baseline: functional imaging and the resting human brain. *Nat Rev Neurosci* **2**, 685–94 (2001).

33. C. Ranganath, M. Ritchey, Two cortical systems for memory-guided behaviour. *Nature Reviews Neuroscience* **13**, 713–726 (2012).

34. M. Ritchey, M. E. Montchal, A. P. Yonelinas, C. Ranganath, Delay-dependent contributions of medial temporal lobe regions to episodic memory retrieval. *elife* **4**, e05025 (2015).

35. A. W. Gilmore, S. M. Nelson, K. B. McDermott, The Contextual Association Network Activates More for Remembered than for Imagined Events. *Cereb Cortex* **26**, 611–7 (2016).

36. A. W. Gilmore, S. M. Nelson, H.-Y. Chen, K. B. McDermott, Task-related and resting-state fMRI identify distinct networks that preferentially support remembering the past and imagining the future. *Neuropsychologia* **110**, 180–189 (2018).

37. M. Bar, The proactive brain: memory for predictions. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **364**, 1235–1243 (2009).

38. D. Stawarczyk, O. Jeunehomme, A. D'Argembeau, Differential contributions of default and dorsal Attention networks to remembering thoughts and external stimuli from real-life events. *Cereb Cortex* **28**, 4023–4035 (2018).

39. N. Raz, K. M. Rodrigue, Differential aging of the brain: Patterns, cognitive correlates and modifiers. *Neuroscience & Biobehavioral Reviews* **30**, 730–748 (2006).

40. R. Leech, D. J. Sharp, The role of the posterior cingulate cortex in cognition and disease. *Brain* **137**, 12–32 (2014).

41. F. Lin, *et al.*, The cingulate cortex of older adults with excellent memory capacity. *Cortex* **86**, 83–92 (2017).

42. N. Kriegeskorte, M. Mur, P. A. Bandettini, Representational similarity analysis - connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* **2** (2008).

43. A. Schaefer, *et al.*, Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb Cortex* **28**, 3095–3114 (2018).

44. D. Stawarczyk, M. A. Bezdek, J. M. Zacks, Event Representations and Predictive Processing: The Role of the Midline Default Network Core. *Topics in Cognitive Science* (In press) https:/doi.org/10.1111/tops.12450 (September 6, 2019).

45. L. Hoffman, On the Interpretation of Parameters in Multivariate Multilevel Models Across Different Combinations of Model Specification and Estimation. *Advances in Methods and Practices in Psychological Science* **2**, 288–311 (2019).

46. P. Zeidman, E. A. Maguire, Anterior hippocampus: the anatomy of perception, imagination and episodic memory. *Nat Rev Neurosci* **17**, 173–82 (2016).

47. M. D. Fox, *et al.*, The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *PNAS Proceedings of the National Academy of Sciences of the United States of America* **102**, 9673–9678 (2005).

48. D. Vatansever, D. K. Menon, E. A. Stamatakis, Default mode contributions to automated information processing. *PNAS* **114**, 12821–12826 (2017).

49. C. Baldassano, *et al.*, Discovering Event Structure in Continuous Narrative Perception and Memory. *Neuron* **95**, 709-721.e5 (2017).

50. R. K. Olsen, S. N. Moses, L. Riggs, J. D. Ryan, The hippocampus supports multiple cognitive processes through relational binding and comparison. *Front. Hum. Neurosci.* **6** (2012).

51. A. Mayes, D. Montaldi, E. Migo, Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences* **11**, 126–135 (2007).

52. A. Ben-Yakov, R. N. Henson, The Hippocampal Film Editor: Sensitivity and Specificity to Event Boundaries in Continuous Experience. *J. Neurosci.* **38**, 10057–10068 (2018).

53. C. N. Wahlheim, Proactive effects of memory in young and older adults: the role of change recollection. *Mem Cognit* **42**, 950–64 (2014).

54. B. Levine, E. Svoboda, J. F. Hay, G. Winocur, M. Moscovitch, Aging and autobiographical memory: dissociating episodic from semantic retrieval. *Psychol Aging* **17**, 677–689 (2002).

55. T. A. Allen, A. M. Morris, S. M. Stark, N. J. Fortin, C. E. L. Stark, Memory for sequences of events impaired in typical aging. *Learn. Mem.* **22**, 138–148 (2015).

56. J. M. Zacks, N. K. Speer, J. M. Vettel, L. L. Jacoby, Event understanding and memory in healthy aging and dementia of the Alzheimer type. *Psychology and Aging* **21**, 466–482 (2006).

57. C. A. Kurby, J. M. Zacks, Age differences in the perception of hierarchical structure in events. *Mem Cognit* **39**, 75–91 (2011).

58. A. Gazzaley, M. D'Esposito, *BOLD functional MRI and cognitivec aging* (Oxford University Press., 2005).

59. J. L. Earles, A. Kersten, B. Berlin Mas, D. M. Miccio, Aging and Memory for Self-Performed Tasks: Effects of Task Difficulty and Time Pressure. *J Gerontol B Psychol Sci Soc Sci* **59**, P285–P293 (2004).

60. J. J. Allaire, *et al.*, rmarkdown: Dynamic Documents for R. R package version 1.13 (2019).

61. M. F. Folstein, S. E. Folstein, P. R. McHugh, "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189–198 (1975).

62. W. C. Shipley, C. C. Burlingame, A convenient self-administering scale for measuring intellectual impairment in psychotics. *AJP* **97**, 1313–1325 (1941).

63. W. Schneider, A. Eschman, A. Zuccolotto, *E-Prime user's guide* (Psychology Software Tools Inc., 2012).

64. B. Whitcher, V. J. Schmid, A. Thorton, Working with the DICOM and NIfTI Data Standards in R. *Journal of Statistical Software* **44**, 1–29 (2011).

65. M. F. Glasser, *et al.*, The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80**, 105–124 (2013).

66. M. H. A. Hendriks, N. Daniels, F. Pegado, H. P. Op de Beeck, The Effect of Spatial Smoothing on Representational Similarity in a Simple Motor Paradigm. *Front. Neurol.* **8** (2017).

67. M. Hanke, *et al.*, PyMVPA: a unifying approach to the analysis of neuroscientific data. *Front. Neuroinform.* **3** (2009).

68. J. R. Landis, G. G. Koch, The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174 (1977).

Figures

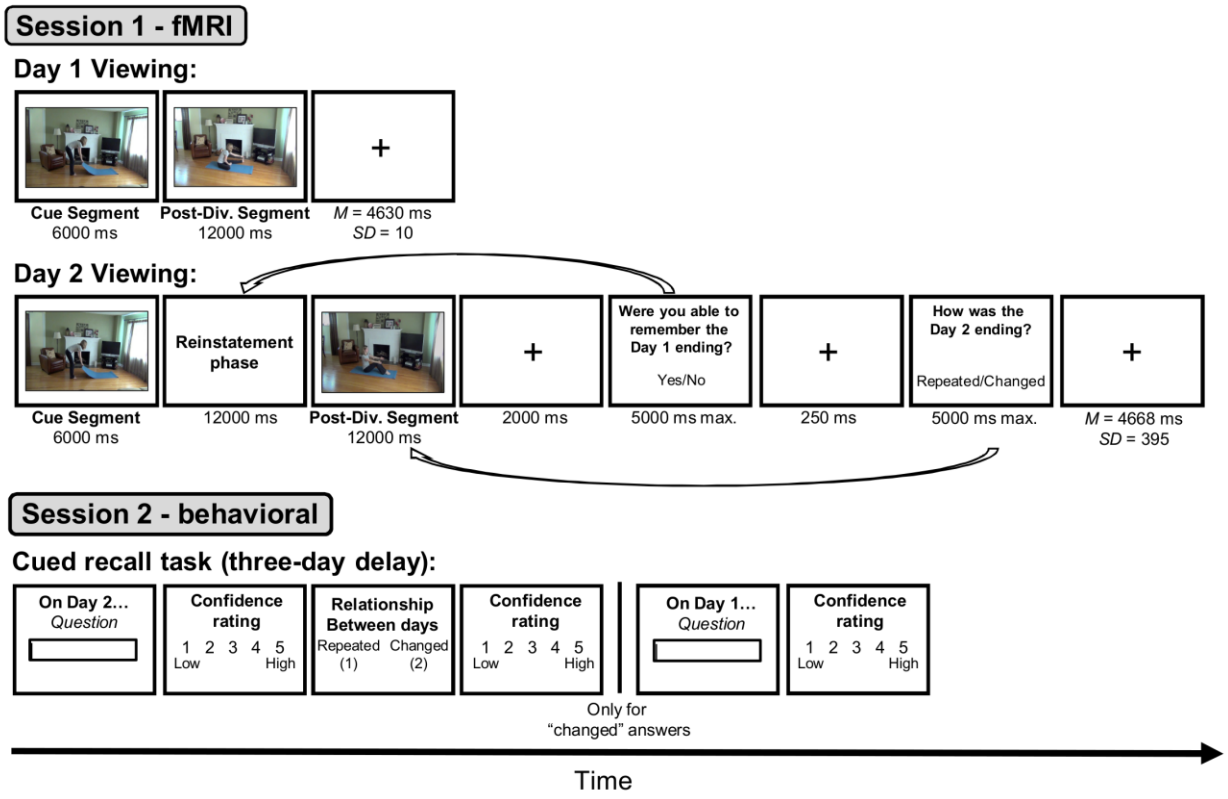


Figure 1. Trial structure of the tasks. For a more detailed description of the materials and procedure, see "Materials, Design, and Procedure," in the main manuscript and the SI appendix, Section 1.2 and 1.3.

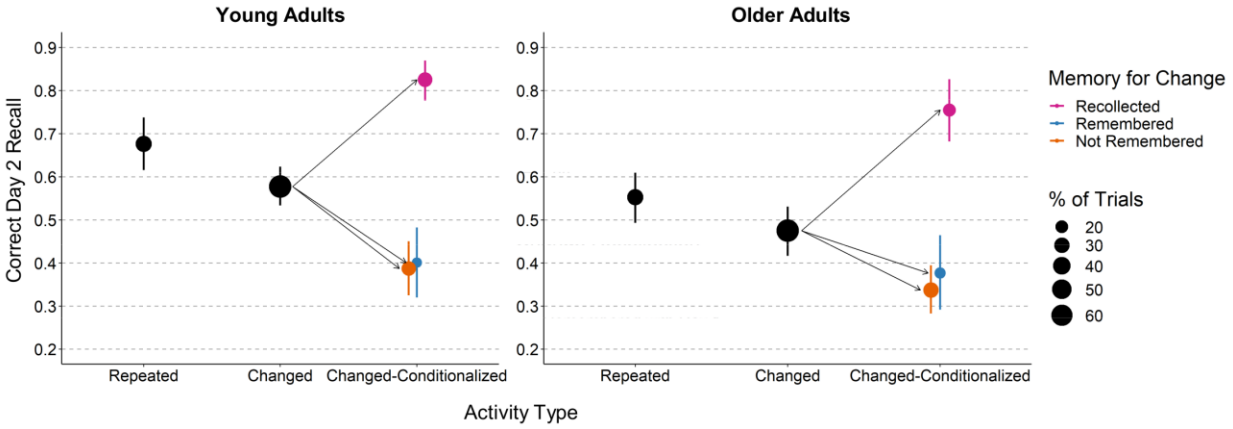


Figure 2. Mean probabilities of correct Day 2 recall for younger adults (left panel) and older adults (right panel). Error bars are bootstrap 95% confidence intervals. Change could be remembered with recall of Day 1 features (Change Recollected; purple points), remembered without recall of Day 1 features (Change Remembered; blue points), or not remembered at all (Change Not Remembered; orange points). Point areas for conditional cells reflect the relative proportions of observations in each cell.

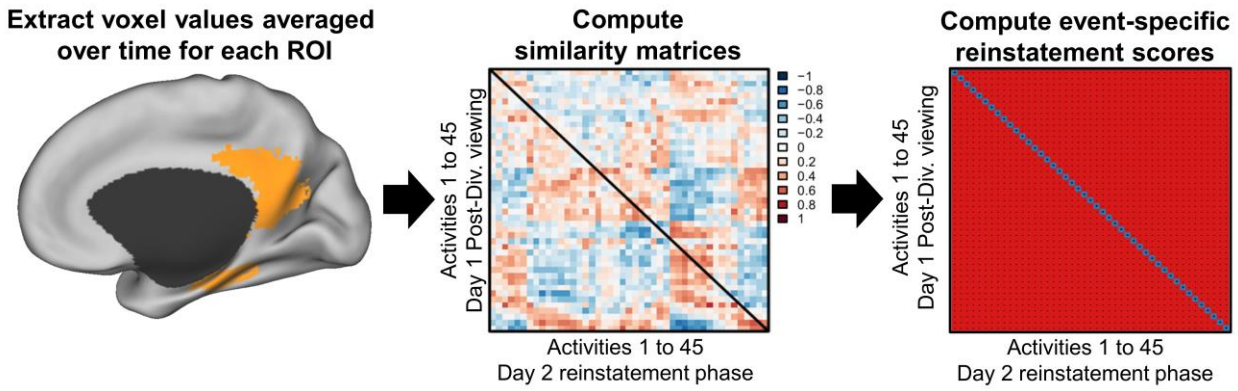


Figure 3. Multivariate voxel analyses. We first averaged voxel values over time in the parcels of interest during the Day 1 post-divergence viewings and Day 2 reinstatement phases for each activity and participant. Next, we correlated voxel values in the parcels of interest for each activity during Day 1 viewing with the corresponding voxel values for all the activities during Day 2 reinstatement. The resulting correlation coefficients can be plotted in similarity matrices, as illustrated above. We then computed event-specific reinstatement scores by calculating the differences between each of the on-diagonal values indicated in blue and the off-diagonal values indicated in red. For a more detailed description of this procedure, see "fMRI Data Analyses."

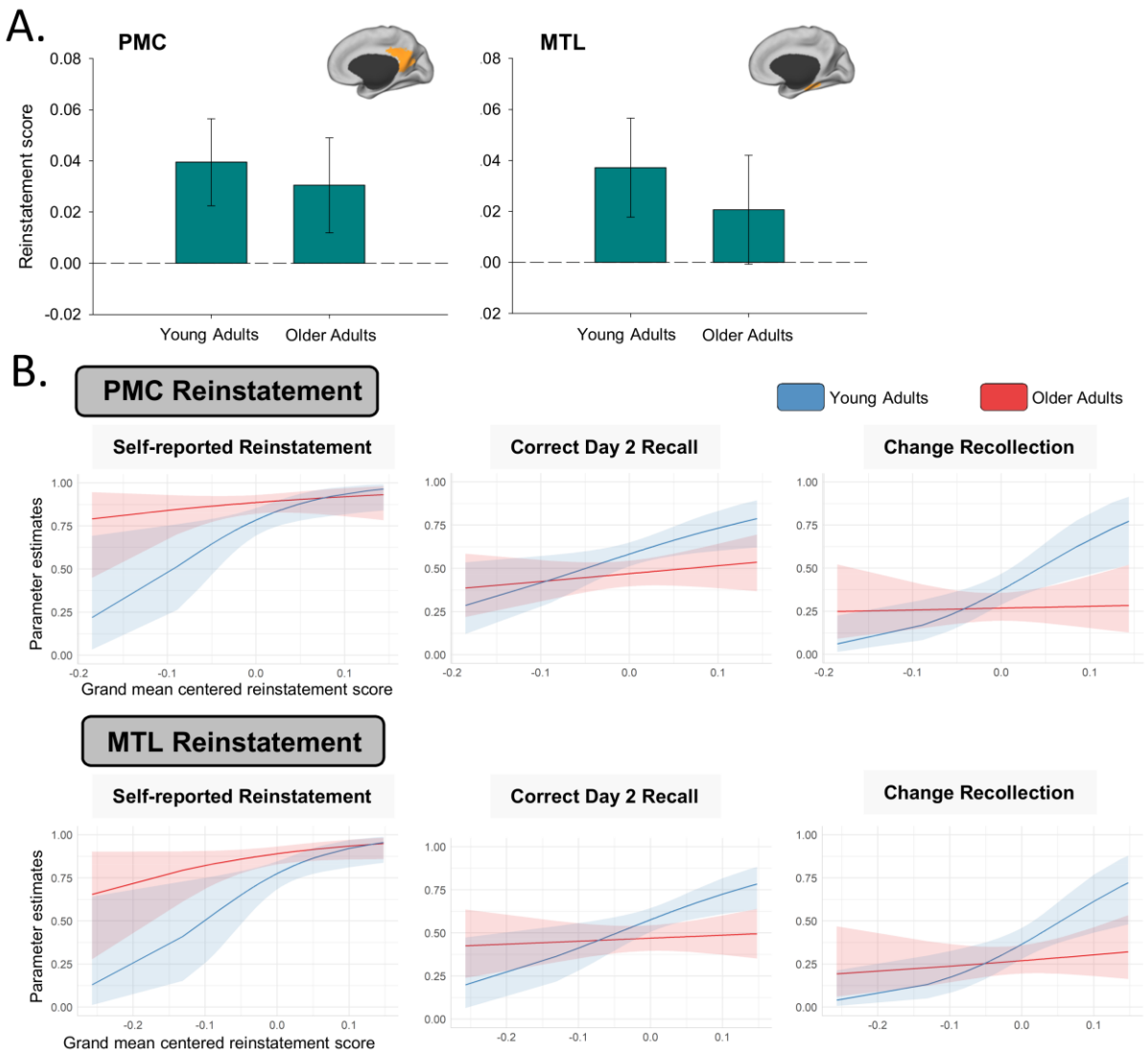


Figure 4. Panel A. Parameter estimates for mean reinstatement scores in the MTL and PMC by age group. The shaded regions represent 95% confidence intervals. Panel B. Parameter estimates and 95% confidence intervals for the between-participant associations between mean PMC/MTL reinstatement scores and behavioral memory measures.

Aging and the encoding of changes in events:

The role of neural activity pattern reinstatement

Supporting Information

David Stawarczyk[1, 2], Christopher N. Wahlheim[4], Joset A. Etzel[1], Abraham Z. Snyder[3,5], Jeffrey M. Zacks[1,3]

[1] Department of Psychological & Brain Sciences, Washington University in St. Louis, St. Louis, MO 63105, USA.

[2] Department of Psychology, Psychology and Neuroscience of Cognition Research Unit, University of Liège, 4000 Liège, Belgium.

[3] Mallinckrodt Institute of Radiology, Washington University in St. Louis, St. Louis, MO 63110, USA.

[4] Department of Psychology, University of North Carolina at Greensboro, Greensboro, NC 27412, USA.

[5] Department of Neurology, Washington University in St. Louis, St. Louis, MO 63105, USA.

Correspondence concerning this article should be addressed to David Stawarczyk, Department of Psychology, University of Liège, Quartier Agora, place des Orateurs 1, 4000 Liège 1, Belgium. E-mail: d.stawarczyk@uliege.be. Tel.: 0032 4 3665979.

Supporting Information Table of Contents

a. Participants 39

b. Materials 39

c. Procedure and task description 40

d. fMRI data acquisition 42

e. Spatial preprocessing of fMRI data 43

f. Mass univariate fMRI modeling 43

g. Classifier MVPA 44

h. Behavioral performance modeling 45

2. Results and Discussion 45

a. Behavioral performance 45

i. Self-reported reinstatement success and change detection during Day 2 viewing 45

ii. Day 1 intrusions during Day 2 recall 48

b. **RSA analyses** 50

i. Day 1 intrusions and change detection accuracy: Between-participant associations with mean PMC and MTL reinstatement scores 50

ii. Between-participant analyses of individual parcel scores 50

iii. Within-participant analyses of individual parcel scores 52

iv. Conjoint associations of self-reported reinstatement and RSA reinstatement scores with Day 2 recall accuracy for changed activities 54

c. Hippocampal response for the difference between changed and repeated activities, reinstatement scores, and behavioral performance 55

d. Classifier MVPA results 57

3. Recognition task 57

a. Task description 57

b. Behavioral results 58

c. RSA results 59

d. Hippocampal response for the difference between changed and repeated activities, reinstatement scores, and behavioral performance 61

4. References 62

5. Figures 64

6. Tables 82

Material and Methods

a. Participants

Participants were recruited from the Washington University School of Medicine Research Participant Registry, flyers posted on campus, and word of mouth. Potential participants first completed a prescreening interview over the phone. In addition to MRI contraindications, we excluded anyone who reported a history of neurological or psychiatric disorders, taking medication that could affect their cognitive functioning, or not having normal or corrected to normal vision and audition. The two age groups did not significantly differ in years of education [$t(60) = -1.51$, $p = .14$, $d = .38$; $M = 15.85$, $SD = 1.71$, for the young adults; $M = 16.57$, $SD = 2.56$, for the older adults] or percentage correct on the Shipley Institute of Living Scale (1) vocabulary test [$t(60) = -1.87$, $p = .07$, $d = .48$; $M = 83.90\%$, $SD = 8.33$ for the young adults; $M = 87.88\%$, $SD = 8.28$ for the older adults]. Each participant received $25.00 per hour for participating in the study. Nine other participants took part in the study but were excluded from the analyses: Two young adults interrupted the study during scanning, technical issues lead to unusable data for two older adults, and five more older adults did not comply with the task instructions.

b. Materials

Each of the 45 activities comprised two parts: an initial *cue* segment that was identical for the A and B versions and lasted 6 s followed by a *post-divergence* segment that sometimes included the changed feature and lasted 12 s with the last second including a fade to black transition. For some of the activities, the changed feature was an object that the actor contacted (e.g., pouring a glass of milk or a glass of water). For other activities, the changed feature was the action itself (e.g., doing leg stretches or sit-ups on a yoga mat). In all cases the changed feature was central to the activity performed. The critical manipulation was whether the post-divergence segment was the same in both movies (repeated activities), or whether that segment changed from the Day 1 to Day 2 movie (changed activities; e.g., the A version in Day 1 and the B version in Day 2). We included twice as many changed as repeated activities (i.e., 30 vs. 15) to increase the power to detect differences in neural activity associated with change processing.

Colored square-wave gratings were overlaid on the movies during the post-divergence segment with a spatial frequency of approximately half a cycle per degree of visual angle and an opacity of 60%. The gratings gradually appeared over the movies during the first 1.5 s of the post-divergence segment (see **Figure S8**). Half of the gratings were red and vertical and half were green and

horizontal. Gratings were presented in a fixed pseudo-random sequence with the constraint that they did not repeat for more than four consecutive activities. For clips that included a changed feature in the Day 2 movie, the gratings also changed; for clips that included a repeated feature in the Day 2 movie, the gratings remained the same[1].

To counterbalance the assignment of activities to conditions, we created 12 experimental formats. We divided the 45 activities into 3 groups of 15 activities and rotated them through conditions across participants, such that each participant viewed two groups of changed activities and one group of repeated activities. The sequence of activities was fixed, beginning with the actor waking up and ending with her going to bed. The assignment of activities to experimental conditions (which ending was presented on Day 1, and whether the item was repeated or changed) was fixed in a pseudo-random sequence such that each third of the task (15 activities) contained five repeated activities with no more than six consecutive changed activities throughout the task. We also alternated the gratings and whether the A or B version of each activity appeared in the Day 1 movie. All stimulus materials were presented using E-Prime 2 software (Psychology Software Tools, Pittsburgh, PA, USA; , 2).

c. Procedure and task description

Participants completed the task in two sessions separated by three days (**Figure 1**). During the first session, participants watched both movies while lying in the scanner. Before the first functional run, we informed participants that they would watch a movie of an actor performing a series of everyday activities throughout the course of her day. We instructed participants to pay attention to her actions and the objects that she contacted. We called this functional run the "Day 1" viewing. To encourage attention during encoding of post-divergence segments, we instructed participants to associate the movie content with the colored gratings that we told them would also appear briefly after the onset of each activity. In the Day 1 movie, the initial cue segment of each activity was followed immediately by the post-divergence segment. Following each post-divergence segment, a fixation cross appeared in the middle of the screen for a mean duration of 4630 ms (*SD* = 10, range = 4585-4693 ms) with the onset of the cue segment for the next activity

[1] The aim of these gratings was to provide an alternative means of analyzing neural pattern reinstatement that allowed for repeated presentations stimulus features as a complement to identifying unique event features. Initial analyses indicated that reinstatement of event features was more robust than reinstatement of grating features, so the effects of the gratings were not analyzed further.

being synchronized with the onset of the next scan. This fixation cross provided a demarcation between events, which was necessary for the subsequent reinstatement task. It also allowed the blood-oxygen-level-dependent (BOLD) response to decrease before the onset of the next activity and permitted us to use the same time window across activities regarding the onsets of the video clips in the representational similarity analyses (RSA; see the fMRI analyses section for details). We presented two example activities before the beginning of the run.

After participants viewed the Day 1 movie, field map images and a high-resolution anatomical image were collected, taking approximately 10 minutes. We then told participants that their next task would be to watch another movie that depicted the same actor performing activities on another fictive day that occurred one week later in her life. We called this second functional run the "Day 2" viewing (see **Figure 1**). We explained to participants that the activities would appear in the same order as in the first movie, but that the clips would stop after a few seconds so that they could mentally replay the upcoming action from the Day 1 movie. Participants were told that during this "reinstatement phase" they should imagine the event ending from the Day 1 movie in as much detail as possible. During the Day 2 viewing, each clip stopped after the cue segment, and a question appeared for 12 s (to match the duration of the Day 1 post-divergence segment) asking participants what happened next during Day 1. After the 12 s reinstatement phase, participants watched the post-divergence segment. For this segment, participants were told to pay specific attention to whether the activity features repeated or changed from Day 1. After each post-divergence segment, a 2 s fixation cross followed by two successive questions appeared. The first question asked whether participants thought that they correctly remembered the Day 1 post-divergence segment during the reinstatement phase. The second question asked whether the action in the post-divergence segment was repeated or changed. Participants responded "yes" or "no" to the first question and "repeated" or "changed" to the second question, using a button box to respond to both questions. The duration of each question was self-paced with a cut-off of 5 s, and a 250 ms fixation cross appeared between questions. As for the Day 1 movie, a fixation cross appeared at the end of the trial for a mean duration of 4668 ms (*SD* = 395, range = 4021-5345 ms) with the onset of the following cue segment being synchronized with the onset of the next scan.

Before the task began, participants completed two example trials using the same example activities that appeared before the first run. At the end of the second functional run, we collected a second set of field map images and a high-resolution T2-weighted image. We then informed

participants that the session was over. We did not mention that Session 2 would consist of a memory test, in order to reduce the likelihood that participants would rehearse the activities during the three-day delay.

During Session 2, outside of the scanner, we first tested participants' memory for the previously viewed activities using a cued recall task (see **Figure 1**). During the task, the questions appeared in the same order as the activities during each movie. For each activity, we first asked participants to recall features that appeared in the Day 2 clips by typing their response. For example, for the question, "What did the actor eat for breakfast?", participants might respond "A banana." The tested features were all critical features that varied between the A and B versions of the activities. After each response, we asked participants whether the way that the actor performed the activity changed from Day 1 to Day 2. Participants pressed the "1" key to indicate that the activity repeated exactly and the "2" key to indicate that the activity changed. When participants indicated that the activity had changed, they were prompted to type in the original Day 1 feature. For each of these questions, we asked participants to rate the confidence in their answer on a Likert scale ranging from 1 "Low" to 5 "High" (these confidence ratings are not relevant to the hypotheses investigated here, so we do not discuss them further). All responding during the task was self-paced. Following the cued recall task, we administered a recognition test of Session 1 activities (see **Section 3** of the Supplementary Information for details). After the recognition test, all participants completed a vocabulary test (1), and older adults completed the MMSE (3) last.

d. fMRI data acquisition

All fMRI data were collected using a 3 Tesla scanner (Magnetom Prisma, Siemens Erlangen, Germany) equipped with a 64-channel receiver head coil. Head movement was minimized with foam padding. Anatomical imaging including a multi-echo, T1-weighted MP-RAGE scan (TR = 2500 ms, TE = 1.81, 3.60, 5.39, and 7.18 ms, FOV 256 × 240 mm, matrix 320 × 300 × 208, voxel size 0.8 × 0.8 × 0.8 mm) and a high-resolution T2-weighted SPACE scan (TR = 3200 ms, TE = 564 ms, FOV 256 × 240 mm, matrix 320 × 300 × 208, voxel size 0.8 × 0.8 × 0.8 mm). BOLD fMRI was acquired using a T2*-weighted, multi-band accelerated EPI pulse sequence developed at the Center for Magnetic Resonance Research (CMRR) at the University of Minnesota (MB factor = 4, TR = 1330 ms, TE = 38.8 ms, FA 63°, matrix size 110 × 110 × 60, voxel size 2.4 × 2.4 × 2.4 mm, A >> P phase encoding). Whole brain coverage was obtained with 60 2.4 mm slices without in-plane acceleration (iPAT = 0). Two fMRI runs were acquired in each participant.

A mean of 782 (*SD* = 3.83, range: 775-800) and 1343.50 (*SD* = 29.78, range: 1285-1420) volumes were acquired in the first and second runs, respectively. Spin echo field maps were acquired after each fMRI run. Stimuli were displayed on a screen positioned at the rear of the scanner that participants viewed via a mirror mounted on the head coil.

e. Spatial preprocessing of fMRI data

The functional data were analyzed in $(2\ mm)^3$ 711-B2 atlas space (4). Atlas transformation was initially computed by composition of affine transforms (fMRI functional volume mean → T2w → T1w → 711-2B space representative target image). The last transformation step was refined by non-linear registration of each individual's T1w to the atlas representative target using the FNIRT module in fsl (5, 6). One step final resampling of the functional data in atlas space combined retrospective head motion correction, magnetization inhomogeneity distortion correction via topup (7) and non-linear atlas transformation.

f. Mass univariate fMRI modeling

Following preprocessing, brain images were spatially smoothed with a Gaussian kernel of 6-mm full-width at half maximum in Statistical Parametric Mapping (SPM). For each participant, BOLD responses during the Day 2 run were modeled at each voxel, using a general linear model (GLM). The post-divergence segments of repeated and changed items were modeled separately as epoch-related responses (lasting 12 s) and convolved with the canonical hemodynamic response function to create the regressors of interest. The cue segments and reinstatement phases were also modeled as epoch-related responses (respectively lasting 6 s and 12 s), each with a single regressor across all trials. The 2 questions ending each trial were also modeled as epoch-related responses with a single regressor across all conditions. The design matrix also included the realignment parameters to account for any residual movement-related effect. The canonical hemodynamic response function was used, and a high-pass filter was implemented using a cutoff period of 128 s to remove the low frequency drifts from the time series. Serial autocorrelations were estimated with a restricted maximum likelihood algorithm with an autoregressive model of order 1 (+ white noise). Based on this design, we computed a linear contrast to examine the brain regions more active in changed than repeated post-divergence segments. The contrasts of interest were first computed for each participant and were then entered in a random-effect one-sample *t*-test to assess the overall effect of change across all participants. We then entered the contrasts in a two-sample *t*-test to assess group difference between young and older adults. For both contrasts, we report

activations that were statistically significant using a threshold of $p < .05$, corrected for multiple comparisons (familywise error; FWE) over the whole brain or within a mask of the bilateral hippocampus.

To assess the possibility that brain activity while viewing the changed post-divergence segments could be modulated by event-specific reinstatement scores, we computed the same model as described in the previous paragraph but added a parametric regressor consisting of the trial-wise reinstatement scores computed in the RSA, centered within each participant. Three models were computed with either the mean PMC, MTL, or hippocampus reinstatement scores used as parametric regressor. We next computed for each participant the contrast of this parametric regressor against a baseline that we then entered in a random-effect one-sample *t*-tests to assess its overall effect across all participants. We then entered the contrasts in a two-sample *t*-test to assess the group difference between young and older adults. For both contrasts, we report activations that were statistically significant using a threshold of $p < .05$, corrected for multiple comparisons (FWE) over the whole brain or within a mask of the bilateral hippocampus.

Finally, to assess the neural response to the movies during Day 1 viewing, we also modeled BOLD responses at each voxel for each participant using a GLM. The 18 s of each activity were modeled separately as epoch-related responses and convolved with the canonical hemodynamic response function to create the regressor of interest. The design matrix also included the realignment parameters to account for any residual movement-related effect. The canonical hemodynamic response function was used, and a high-pass filter was implemented using a cutoff period of 128 s to remove the low frequency drifts from the time series. Serial autocorrelations were estimated with a restricted maximum likelihood algorithm with an autoregressive model of order 1 (+ white noise). Based on this design, we computed a linear contrast to examine the brain regions more active while viewing the movies than baseline. The contrasts of interest were first computed for each participant and were then entered in a random-effect one-sample *t*-tests to assess the overall effect movie viewing across all participants. We then entered the contrasts in a two-sample *t*-test to assess the difference between young and older adults. For both contrasts, we report activations that were statistically significant using a threshold of $p < .05$, corrected for multiple comparisons (FWE) over the whole brain. Only clusters with a size of $k > 20$ voxels are reported for each of these analyses.

g. Classifier MVPA

To further assess the effect of age on the neural response to the movies during Day 1 viewing, we computed the same GLM as in the previous sections using SPM, except that images were spatially smoothed with a Gaussian kernel of 3 (rather than 6)-mm full-width at half maximum. We then used the Pronto Toolbox (8) to determine with a leave-one out procedure whether a linear support vector machine classifier ($C = 1$) trained on the parameter estimate images corresponding to the movie regressor of every participant except one could accurately determine the age group of the remaining participant. We ran this analysis twice, using either a mask of the PMC or MTL. To assess whether the classification accuracy between the two groups was above chance, a permutation procedure randomly shuffling the group label of the parameter estimate images was used (1000 permutations using the default Pronto settings). Given the unequal number participants in each age group, we report the balanced accuracy values that give equal weight to the accuracies obtained for each group.

h. Behavioral performance modeling

All analyses of memory performance and reinstatement effects were computed in R software (9) using linear or logistic mixed effect models with subjects and activities as random effects. Logistic models were used when the dependent variable was binary. Models were fitted using functions from the lme4 package, hypothesis tests were performed using the Anova function of the car package (10), and post hoc comparisons using the Tukey method were conducted using the emmeans function from the emmeans package (11). Comparisons between nested models were performed with the anova function of the lme4 package. Finally, the plot_model and get_model_data functions from the sjPlot package (12) were used to compute parameter estimates (i.e., model estimated probabilities) for description and visualization.

2. Results and Discussion

a. Behavioral performance

i. Self-reported reinstatement success and change detection during Day 2 viewing

During Day 2 viewing, each activity was paused after the cue segment, and participants were asked to replay how the activity ended on Day 1. After seeing the changed or repeated post-divergence segment, they reported whether their reinstatement was successful and whether the Day 2 ending was repeated exactly or included a changed feature (see **Figure 1**). To characterize participants' subjective accuracy in reinstating Day 1 features and objective accuracy in activity type classifications (see **Table S3**), we fitted 2 (Activity Type: Repeated vs. Changed) × 2 (Age

Group: Young vs. Older) models to each measure. The model for reinstatement success indicated that older adults reported significantly more successful Day 1 reinstatements than young adults [$\chi^2(1) = 4.55$, $p = .03$]. In contrast, the model for classification accuracy showed the opposite pattern: young adults showed significantly greater classification accuracy than older adults [$\chi^2(1) = 6.68$, $p = .01$]. No other effects were significant [*largest* $\chi^2(1) = 1.10$, $p = .30$]. Together, these results suggest that metacognitive accuracy about memory for Day 1 features was better for young than older adults. Relevant to our interest in age differences in change processing, these findings imply that reinstatement success should be less predictive of change detection accuracy for older than young adults.

To determine the extent to which successful reinstatement of Day 1 activity features contributed to detection of changed Day 2 features for both age groups, we examined the relationship between reinstatement success and classification accuracy for changed activities only. Finding that successful reinstatement predicts more accurate change detection would be consistent with the proposal that prior event retrieval facilitates detection of changed event features. Further, finding that such an effect is greater for young than older adults would suggest that older adults are impaired in their ability to use of prior event retrievals to update their memories. A 2 (Reinstatement: Successful vs. Unsuccessful) × 2 (Age Group: Young vs. Older) model (see **Table S4**) indicated a significant main effect of Reinstatement [$\chi^2(1) = 25.95$, $p < .001$] and a significant Reinstatement × Age Group interaction [$\chi^2(1) = 9.97$, $p = .002$], showing that successful reinstatements were associated with more accurate change detection for young adults (z ratio = 5.94, $p < .001$) but not for older adults (z ratio = 1.00, $p = .32$). These results suggest that prior event retrievals enabled more effective change detection for young than older adults.

Next, we assessed the extent to which reinstatement success and change detection during Day 2 viewing were associated with subsequent memory performance in the cued recall task. Prior behavioral work has shown that accurate change detection is associated with better cued recall for Day 2 changed activity features and predicts overall rates of change recollection for both young and older adults (13). Here, we examined whether those findings would replicate, and we extended this analysis to the self-report measure of reinstatement. Given the age differences in the relationship between reinstatement success and change detection above, we expected age differences in the associations between these measures and behavioral memory performance.

Figure S9 displays Day 2 recall and change recollection measured on the cued recall test conditionalized on reinstatement success (left panels) and change detection accuracy (right panels) measured during Day 2 viewing. To examine how the Day 2 measures predicted performance on behavioral memory measures for changed activities at test, we fitted separate models to each test measure including the Day 2 measure (reinstatement or change detection) and age group as fixed effects. The model testing the association between reinstatement success and Day 2 recall (top left panel) indicated that successful reinstatement was associated with more accurate Day 2 recall [$\chi^2(1) = 12.49$, $p < .001$] and this effect did not interact with Age Group [$\chi^2(1) = 0.14$, $p = .71$]. Reinstatement was also associated with greater change recollection[2] for both age groups (middle left panel) [$\chi^2(1) = 61.93$, $p < .001$] and this effect did not interact with Age Group [$\chi^2(1) = 0.05$, $p = .82$]. The same models including change detection accuracy (instead of reinstatement) as the fixed effect indicated no association between change detection and Day 2 recall (top right panel) [$\chi^2(1) = 3.44$, $p = .07$], and a positive association between change detection and change recollection (bottom right panel) [$\chi^2(1) = 87.85$, $p < .001$]. These effects did not interact with Age Group [*largest* $\chi^2(1) = 0.74$, $p = .39$]. The lack of association between change detection and Day 2 recall accuracy fails to replicate earlier findings (13), but the positive association between reinstatement and Day 2 recall implicates a role for prior event retrieval in the updating of memory to include changed features. Also, the strong positive associations with change recollection for both reinstatement and change detection are not surprising given that all these measures are considered to assay the accessibility of Day 1 activity features during various phases of the experiment.

Together, the findings that self-reported reinstatement and change recollection were both positively associated with Day 2 recall and each other suggest that the association between successful Day 1 reinstatement and change recollection is a consequence of reinstatement enabling change detection and the formation of configural memory representations during Day 2 viewing. We examined this possibility by testing whether the association between reinstatement and Day 2 recall remained when change recollection was also included as a fixed effect. The rationale here is that if change recollection predicts Day 2 recall when controlling for reinstatement, but

[2] Given the lack of difference in Day 2 recall performance when change was not remembered or remembered without the Day 1 feature (see Figure 2), we collapsed across these cells in this and subsequent analyses. Conditional cells for changed activities were classified as *change recollected* (remembered as changed and the Day 1 feature correctly recalled) and *change not recollected* (not remembered as changed or remembered without the Day 1 feature).

reinstatement does not predict Day 2 recall when controlling for change recollection, then this would imply a mediating role for reinstatement during Day 2 viewing in the positive association between change recollection and Day 2 recall at test. **Figure S10** displays Day 2 recall accuracy predicted by reinstatement success and change recollection controlling for the effects of the other variable. A 2 (Reinstatement: Successful vs. Unsuccessful) × 2 (Change Recollection: Recollected vs. Not Recollected) × 2 (Age Group: Young vs. Older) model indicated no significant effect of Reinstatement [$\chi^2(1) = 0.49$, $p = .48$] and a significant effect of Change Recollection [$\chi^2(1) = 219.95$, $p < .001$], and Age Group [$\chi^2(1) = 7.18$, $p = .008$]. There were no significant interactions [*largest* $\chi^2(1) = 3.75$, $p = .052$].

Together, these results indicate that self-reported reinstatement of Day 1 features during Day 2 viewing was associated with better cued recall of the changed Day 2 features in both age groups. However, this effect did not remain significant after controlling for the ability to recollect change. This suggests that the association between reinstatement of Day 1 experiences during viewing of Day 2 activities and recall of those Day 2 activities is mediated through recollecting at test that the activity had earlier changed. However, the current design does not allow for strong causal conclusions about this potential mediation.

ii. Day 1 intrusions during Day 2 recall

Analyses of Day 1 intrusions during Day 2 recall (**Figure S11**) revealed results that mostly mirrored those of correct recall. We fitted a model to Day 1 intrusions that was comparable to the model for Day 2 recall. There was a fixed effect of Activity Type that included one level for all repeated activities and three levels of changed activities that corresponded with each change classification type. There was also a fixed effect of Age Group. The model indicated a significant effect of Activity Type [$\chi^2(3) = 144.00$, $p < .001$], no significant effect of Age Group [$\chi^2(3) = 0.12$, $p = .73$], and no significant Activity Type × Age Group interaction [$\chi^2(3) = 2.26$, $p = .52$]. For both age groups, recollecting change was associated with fewer intrusions compared to baseline rates for repeated items (z ratio = -3.10, $p = .01$). This is not surprising given that this conditional cell for changed activities were instances when participants reported the Day 1 feature twice, presumably because they were guessing. Intrusion rates for changed activities did not differ between when changes were remembered but the Day 1 feature was not recalled and when changes were not remembered (z ratio = -0.92, $p = .79$). Both conditional cells were associated with intrusion rates above baseline (*smallest* z ratio = 7.56, $p < .001$). We also examined whether Day

1 intrusions during Day 2 recall of changed activities were predicted by reinstatement success and change detection during Day 2 viewing (see **Figure S9, bottom panels**). The model with Age Group and Reinstatement as fixed effects revealed no significant effect of Reinstatement [$\chi^2(1) = 0.81$, $p = .37$], nor Age Group [$\chi^2(1) = 1.31$, $p = .25$], but a significant Reinstatement × Age Group interaction [$\chi^2(1) = 4.80$, $p = .03$]. Successful reinstatement was associated with fewer intrusions for young adults (z ratio = 2.01, $p = .04$) but not for older adults (z ratio = -1.23, $p = .22$). The comparable model including Change Detection as a fixed effect indicated no association between change detection and Day 1 intrusions [$\chi^2(1) = 0.01$, $p = .90$], no effect of Age Group [$\chi^2(1) = 1.03$, $p = .31$], and no significant interaction with Age Group [$\chi^2(1) = 0.01$, $p = .91$].

It may seem somewhat surprising that older adults did not produce more Day 1 intrusions than young adults. Several theories of age-related episodic memory deficits assume that older adults' impaired memory stems from a higher susceptibility to interference (e.g., 14). However, there is mixed empirical evidence supporting this claim (for a review, see 15). One possibility is that the lack of an age-related difference in the present study reflects younger adults having access to more activity features and more liberal reporting criterion than older adults (for similar arguments, see 16). Another possibility is that the shorter delay between Day 2 and test used here (72 hours) compared to that used in previous behavioral work (one week; 13) did not allow enough time for older adults' deficit in source monitoring of intrusions to emerge.

To further investigate these results, we also examined the specific kinds of errors that the participants from each group made during the Day 2 recall of changed activities that were not Day 1 intrusions. To do so, we classified the remaining errors as *No Reponse* (i.e., "I don't remember" and similar answers), *Ambiguous* (i.e., the participants gave an answer that was not specific enough to discriminate between the two versions of the activity), *Inter-Activity Intrusion* (i.e., participants recalled a feature from a different activity), and *Other* (i.e., participants recalled an extra-experimental activity feature). As an illustration of these response types, consider when the actor opened the refrigerator to get water on Day 1 and milk on Day 2. An *Ambiguous* response could be "cold drink," an *Inter-Activity Intrusion* could be "Gatorade" if "Gatorade" was the criterial feature of another activity, and *Other* could be "Wine" if "Wine" was not the criterial feature of an activity. For young adults, of the 218 incorrect Day 2 recalls of changed activities that were not Day 1 intrusions, 25% were *No Response*, 21% were *Ambiguous*, 14% were *Inter-Activity Intrusions*, and 40% were *Other*. For older adults, of the 244 incorrect Day 2 recalls to changed

activities that were not Day 1 intrusions, 28% were *No Response*, 14% were *Ambiguous*, 16% were *Inter-Activity Intrusions*, and 43% were *Other*. Thus, it seems that the distributions of error types (other than Day 1 intrusions) were comparable for young and older adults.

b. **RSA analyses**

i. Day 1 intrusions and change detection accuracy: Between-participant associations with mean PMC and MTL reinstatement scores

To examine between-participant differences in the relationship between neural reinstatement during Day 2 viewing and change detection in the scanner, we computed mean between-participant correlations between RSA reinstatement scores and Change detection accuracy. These analyses revealed no significant effect of reinstatement scores in the either the PMC or MTL [*largest* $\chi^2(1) = 1.14$, $p = .23$], but the interactions with Age Group were significant [$\chi^2(1) = 5.60$, $p = .02$, for the PMC; $\chi^2(1) = 9.03$, $p = .003$, for the MTL]. Higher mean reinstatement scores in both regions predicted more accurate change detection for young adults [$\chi^2(1) = 4.79$, $p = .03$, for the PMC; $\chi^2(1) = 9.85$, $p = .002$, for the MTL] but not for older adults [$\chi^2(1) = 0.81$, $p = .37$, for the PMC; $\chi^2(1) = 0.27$, $p = .60$, for the MTL], see **Figure S12** (right panels). These findings are consistent with the behavioral results showing that self-reported reinstatement success predicted better change detection for young but not older adults. The comparable model predicting Day 1 intrusions indicated no significant effects reinstatement scores in either the PMC or MTL and no significant interaction with Age Group [*largest* $\chi^2(1) = 2.16$, $p = .14$], see **Figure S12** (middle panels).

ii. Between-participant analyses of individual parcel scores

Analyses of the relations between the mean PMC and MTL reinstatement scores revealed several significant associations (see main text). However, it is unclear whether (1) these associations were driven by specific parcels above and beyond the effects of the other parcels in each of the two areas and whether (2) some significant associations might have been obscured by different parcels within the PMC or MTL having opposite associations with memory performance. To investigate these possibilities, we used logistic linear models with between- and within-participant reinstatement scores, as in the main manuscript, but added the individual scores of each parcel as fixed effects rather than a single mean value representing the average of all the parcels. We then used the anova function of the lme4 package (9) to determine whether adding these single parcel scores relative to a baseline model comprising only the random effects improved model fits.

We then examined whether these model fits were further improved by adding the interaction terms that included Age Group as a fixed effect.

First, regarding *Day 2 recall accuracy for changed activities*, adding the between-participant reinstatement scores for the PMC parcels did not improve model fit, nor did adding the interaction terms with Age Group [*largest* $\chi^2(12) = 14.00$, $p = .30$]. These results indicate that the significant effect of the mean PMC reinstatement score reported in the main manuscript was not driven by specific parcels. For the MTL parcels, adding the between-participant reinstatement scores improved model fit [$\chi^2(6) = 13.87$, $p = .03$], but the interactions with Age Group did not [$\chi^2(6) = 7.22$, $p = .30$]. Examination of the parameter estimates for the individual parcels showed that higher reinstatement in parcel 143 was associated with worse Day 2 recall accuracy [$\chi^2(1) = 4.76$, $p = .03$], but this effect did not remain significant when the fixed effect of Age Group was included in the model [$\chi^2(1) = 3.72$, $p = .05$]. We next performed the same analyses for *Day 1 intrusions* for changed activities. Adding reinstatement scores for the PMC or MTL parcels did not improve fit and neither did the interaction with Age Group [*largest* $\chi^2(12) = 18.82$, $p = .25$].

Third, regarding *change recollection* for changed activities in the cued recall task, adding the between-participant reinstatement scores for the PMC parcels did not improve model fit [$\chi^2(12) = 17.03$, $p = .15$], but adding the interaction terms with Age Group did [$\chi^2(12) = 22.99$, $p = .03$]. Further analyses showed that individual parcel reinstatement scores improved model fit for the young [$\chi^2(12) = 24.49$, $p = .02$] but not the older adults [$\chi^2(12) = 14.66$, $p = .26$]. For the young adults, higher reinstatement scores were associated with higher change recollection accuracy in parcels 115 [$\chi^2(1) = 18.55$, $p < .001$] and 292 [$\chi^2(1) = 5.41$, $p = .02$]. No significant effect was found in the models including the MTL parcels [*largest* $\chi^2(6) = 11.76$, $p = .07$].

Fourth, for *self-reported reinstatement success* during Day 2 viewing, adding the between-participant reinstatement scores for the PMC parcels did not significantly improve model fit [$\chi^2(12) = 20.71$, $p = .05$], but adding the interaction terms with Age Group did [$\chi^2(12) = 31.40$, $p = .002$]. Further analyses showed that the individual parcel scores improved model fit for both young [$\chi^2(12) = 28.40$, $p = .005$] and older adults [$\chi^2(12) = 23.66$, $p = .02$]. However, reinstatement scores in different parcels for each age group were associated with self-reported reinstatement success: for the young adults, higher reinstatement scores were associated with higher self-reported reinstatement accuracy in parcels 115 [$\chi^2(1) = 19.83$, $p < .001$] and 117 [$\chi^2(1) = 5.30$, $p = .02$], whereas higher reinstatement scores were associated with lower self-reported reinstatement

accuracy in parcel 114 [$\chi^2(1) = 5.83$, $p = .02$]. For the older adults, higher reinstatement scores were associated with higher self-reported reinstatement accuracy in parcels 142 [$\chi^2(1) = 10.42$, $p = .001$] and 276 [$\chi^2(1) = 10.48$, $p = .001$], whereas higher reinstatement scores were associated with lower self-reported reinstatement accuracy in parcel 117 [$\chi^2(1) = 5.12$, $p = .02$]. No significant effects were found in the models including the MTL parcels [*largest* $\chi^2(6) = 8.53$, $p = .20$].

Finally, regarding *change detection accuracy* during Day 2 viewing, adding the between-participant reinstatement scores for the PMC parcels improved model fit [$\chi^2(12) = 21.57$, $p = .04$], as did the interaction terms with Age Group [$\chi^2(12) = 24.97$, $p = .01$]. Further analyses revealed that adding the parcels improved model fit for the older adults [$\chi^2(12) = 32.23$, $p = .001$] but not the younger adults [$\chi^2(12) = 19.36$, $p = .08$]. Examination of the parameter estimates for the individual parcels in the older adults showed that higher reinstatement scores were associated with higher change detection accuracy in parcels 114 [$\chi^2(1) = 5.62$, $p = .02$] and 142 [$\chi^2(1) = 18.72$, $p < .001$], whereas higher reinstatement scores were associated with lower change detection accuracy in parcels 117 [$\chi^2(1) = 9.14$, $p = .002$] and 125 [$\chi^2(1) = 16.88$, $p < .001$]. No significant effect was found in the models including the MTL parcels [*largest* $\chi^2(6) = 11.91$, $p = .06$].

Together, these results show that memory for the original Day 1 activity feature (measured by either self-reported reinstatement success during Day 2 viewing or change recollection accuracy in the cued recall task) was strongly associated with PMC reinstatement in parcel 115 (corresponding to the posterior cingulate cortex) in the young adult group. These results are concordant with past fMRI studies that examined memories for naturalistic stimuli using RSA and other multivariate analyses and showed that activity patterns in these regions are particularly predictive of subsequent memory in this population (for a review, see 17). The analyses also showed that different parcels in the two age groups were related to self-reported reinstatement success during Day 2 viewing. These results might explain the behavioral differences between the two groups regarding this variable: Older adults reported greater reinstatement success but, unlike the young adults, self-reported reinstatement did not predict better change detection during Day 2 viewing and was not associated with fewer Day 1 intrusions in the cued recall task.

iii. Within-participant analyses of individual parcel scores

For correct *Day 2 recall* of changed activities in the cued recall task, adding the fixed effects of the PMC parcels improved model fit [$\chi^2(12) = 29.12$, $p = .004$] but adding the interaction terms with Age Group did not [$\chi^2(12) = 11.08$, $p = .52$]. Examination of the parameter estimates

for the individual parcels showed that higher reinstatement scores were associated with higher Day 2 recall accuracy in parcels 291 [$\chi^2(1) = 7.97$, $p = .004$] and 292 [$\chi^2(1) = 7.67$, $p = .006$], whereas higher reinstatement scores predicted lower Day 2 recall accuracy in parcels 141 [$\chi^2(1) = 3.87$, $p = .049$] and 142 [$\chi^2(1) = 6.20$, $p = .01$]. No significant effect was found for the models including the MTL parcels [*largest* $\chi^2(6) = 7.45$, $p = .28$]. Next, regarding *Day 1 intrusions* for changed activities in the cued recall task, adding the fixed effects of the parcels improved model fit [$\chi^2(12) = 24.68$, $p = .02$] but adding the interaction terms with Age Group did not [$\chi^2(12) = 8.55$, $p = .74$]. Examination of the parameter estimates for the individual parcels showed that higher reinstatement scores were associated with more Day 1 intrusions in parcels 142 [$\chi^2(1) = 6.95$, $p = .008$] and 277 [$\chi^2(1) = 4.29$, $p = .04$], whereas higher reinstatement scores were associated with fewer Day 1 intrusions in parcels 115 [$\chi^2(1) = 7.62$, $p = .006$] and 292 [$\chi^2(1) = 5.93$, $p = .01$]. No effects were significant in the models including the MTL parcels [*largest* $\chi^2(6) = 5.89$, $p = .44$]. No significant effects were found for the models including either the PMC or MTL parcels for *self-reported reinstatement success* [*largest* $\chi^2(12) = 19.21$, $p = .08$], *change detection accuracy* [*largest* $\chi^2(12) = 18.92$, $p = .09$], or *change recollection accuracy* [*largest* $\chi^2(12) = 20.65$, $p = .06$].

In sum, these analyses showed a more complex picture for within- than between-participant reinstatement scores. As described in the main text, there were no positive associations between behavioral performance and mean within-participant reinstatement scores in either the PMC or MTL. However, when examining the contribution of individual parcels above and beyond the others parcels in each area, within-participant differences in reinstatement scores for some PMC parcels were associated with *better* memory for changed features, and other PMC parcels were associated with *worse* memory for change features (e.g., parcel 292 in the right retrosplenial cortex vs. 142 in the left retrosplenial cortex). To the best of our knowledge, these results are the first to reveal that reinstatement in closely located areas within the same cortical region can have opposite effects on subsequent memory performance. Interestingly, reinstatement in some of these parcels (e.g., parcel 142) were both predictive of worse memory for the changed features and a higher probability of committing intrusions by erroneously recalling the original activity feature. These results suggest that reinstatement in these areas did not purely consist of noise and did reflect memory for the Day 1 activity features, but in a way that did not facilitate their association with subsequently presented changed activity features. Together, these results support the view that extended cortical areas such as the PMC are composed of distinct subregions (18–20), and indicate

that within-participant neural pattern reinstatement in some PMC parcels might be detrimental to memory for changed features, possibly by disrupting the creation of configural representations and thus favoring the experience of proactive interference (21). However, this is the first such finding, so further research is needed before firm conclusions can be drawn.

iv. Conjoint associations of self-reported reinstatement and RSA reinstatement scores with Day 2 recall accuracy for changed activities

Because both self-reported and neural measures of reinstatement were positively associated with higher memory accuracy for the changed Day 2 activity features for young adults, and because neither effects remained significant after controlling for change recollection (see main text), we tested whether self-reported and RSA reinstatement were both significantly associated with Day 2 recall of changed features above and beyond each other. These models included both self-reported reinstatement and mean reinstatement scores at the between-participant level as fixed effects for either the PMC or MTL (see **Figure S13** for Day 2 recall estimates by mean neural and self-reported reinstatement above and beyond the other effect). Both the PMC and MTL reinstatement scores [$\chi^2(1) = 5.32$, $p = .02$, for the PMC; $\chi^2(1) = 6.77$ $p = .009$, for the MTL], and self-reported reinstatement were significant in these models [$\chi^2(1) = 10.09$, $p = .001$, for the PMC; $\chi^2(1) = 9.81$, $p = .002$, for the MTL]; neither interaction term was significant [*largest* $\chi^2(12) = 0.49$, $p = .48$].

These results showed that between-participant differences in neural reinstatement scores and self-reported reinstatement accuracy both independently predicted recall of changed Day 2 features in young adults. Although the effects of these two indices of reinstatement on memory for changed features were fully explained by change recollection, these results suggest that these two measures reflect partly distinct but complementary processes. A possibility is that young participants mostly made their reinstatement judgments based on whether they were able to remember the specific features that changed between the two movies, whereas neural reinstatement in the PMC and MTL reflected the retrieval of more abstract features of the event models (e.g., spatio-temporal context within the movies of each activity) but not necessarily the changed features. Another possibility is that seeing the activity endings led to feature activation that participants misattributed to reinstatement that had occurred before seeing the ending. This is a form of source confusion, and older adults are more susceptible to such misattributions (22). This may also explain why older adults indicated greater success in their self-reported reinstatement judgments than did young adults and why this was related to reinstatement scores in different parcels than in the young adults

(see section 2.2.2.). Future studies might illuminate this issue with more objective behavioral measures of reinstatement accuracy (e.g., verbal reports of memory representations retrieved during the reinstatement phase).

c. Hippocampal response for the difference between changed and repeated activities, reinstatement scores, and behavioral performance

Computational models of memory updating propose that pattern completion based on previous memories leads to prediction errors when change is experienced, which can then drive new learning including integration processes to form configural memory representations (13, 23). To assess this possibility, we first examined whether participants who had higher mean reinstatement scores in the MTL and PMC were also those who showed higher neural responses in the anterior hippocampal clusters that we found to be more activated while viewing changed as compared to repeated post-divergence segments (see **Figure S6**), and whether this association differed between age group. Then we examined whether participants with higher hippocampal responses had better behavioral performance on measures of memory for changes.

To do so, we extracted for each participant the parameter estimates of the bilateral hippocampal clusters that were more activated during changed than repeated activity viewing from the random-effect one-sample *t*-test using Marsbar Toolbox (24). Given that these estimates in the left and right hippocampus were highly correlated (r = .78), we combined them into a single mean score for subsequent analyses. We then computed multiple regression models with hippocampal response as the dependent variable and reinstatement scores in either the PMC or MTL, Age Group, and their interaction as predictor variables. As there was no difference in hippocampal activity between young and older adults, we do not report the effect of age here to avoid redundancy. For the MTL model, there was no effect of mean reinstatement score [$F(1, 58) = 0.09$, $p = .76$, $\eta_p^2 = .002$] and no significant interaction with Age Group [$F(1, 58) = 0.20$, $p = .66$, $\eta_p^2 = .003$]. The PMC model indicated a similar absence of effects, as neither the effect of mean reinstatement score [$F(1, 58) = 0.01$, $p = .94$, $\eta_p^2 < .001$], nor the interaction with Age Group [$F(1, 58) = 0.10$, $p = .75$, $\eta_p^2 = .002$], was significant.

We next examined whether the difference in hippocampal response for changed than repeated activities could predict behavioral performance. To do so, we used linear mixed effects models to examine whether adding the hippocampal response and its interaction with Age Group in the previously described models assessing the association between reinstatement scores and

behavioral performance improved model fits. Results showed that this was not the case in either region for Day 2 recall accuracy [$\chi^2(2) = 3.10$, $p = .21$, for the MTL; $\chi^2(2) = 2.66$, $p = .26$, for the PMC], Day 1 intrusions during Day 2 recall [$\chi^2(2) = 0.88$, $p = .64$, for the MTL; $\chi^2(2) = 0.95$, $p = .62$, for the PMC], change recollection [$\chi^2(2) = 4.24$, $p = .12$, for the MTL; $\chi^2(2) = 3.90$, $p = .14$, for the PMC], and self-reported reinstatement [$\chi^2(2) = 3.02$, $p = .22$, for the MTL; $\chi^2(2) = 2.84$, $p = .24$, for the PMC]. Unsurprisingly given that hippocampal responses were not significantly related to reinstatement scores, all the significant effects of reinstatement scores described in previous sections and in the main text remained unchanged. However, participants with larger differences in bilateral hippocampal responses when viewing changed than repeated activity endings were more likely to correctly detect changes during Day 2 viewing. Adding the parameter estimates for the hippocampal response and their interaction with Age Group significantly improved fit for the MTL [$\chi^2(2) = 10.67$, $p = .005$] and PMC [$\chi^2(2) = 9.35$, $p = .009$] models. Results also showed that higher parameter estimates predicted better change detection [$\chi^2(1) = 11.03$, $p < .001$, for the MTL; $\chi^2(1) = 9.01$, $p = .003$, for the PMC] and that these effect did not interact with Age Group [*largest* $\chi^2(12) = 1.01$, $p = .31$]. The interaction between reinstatement scores and Age Group described above (see Section 2.2.1) remained significant [$\chi^2(1) = 11.88$, $p < .001$, for the MTL; $\chi^2(1) = 7.06$, $p = .008$, for the PMC]. Higher mean reinstatement scores still predicted better change detection for young adults in the PMC [$\chi^2(1) = 5.31$, $p = .02$] and MTL [$\chi^2(1) = 11.41$, $p < .001$].

The association between the intensity of the difference in the hippocampal response to changed compared to repeated activities and change detection is concordant with previous studies showing that the anterior hippocampus is one brain region that is most consistently involved in novelty processing (for reviews, see 25, 26). However, this activity was not modulated by Age Group and was unrelated to neural reinstatement scores and several behavioral measures (except for Day 2 recognition accuracy of changed activities, see section 3.5 below). Although these findings do not support our hypothesis that the hippocampal response would (at least partly) explain the association between reinstatement scores and subsequent memory performances (13), they are nonetheless consistent previous studies showing that anterior hippocampal engagement does not necessarily reflect successful encoding (e.g., 27). A possibility is that this anterior hippocampal activity solely reflects scene construction processes that are more strongly engaged during changed than repeated events because of the new perceived event features. The engagement of these scene

constructions processes would depend on change detection but they would not facilitate subsequent episodic recollection nor depend on how well participants were able to remember the original Day 1 event features during the reinstatement phases (25). Further studies should be conducted to further assess this possibility.

d. Classifier MVPA results

An unexpected finding of the RSA analyses was that reinstatement scores in the PMC and MTL did not differ across age groups. To examine whether there might nonetheless have been age-related differences in how participants initially perceived the activities, we examined whether there were group differences in brain activity while encoding the activity features during the Day 1 viewing. To do so, we used a pattern classifier to examine whether the voxelwise distribution of activation in the PMC and MTL differed between young and older adults (see section 1.7 for a methodological description). The balanced accuracy for the classifier ran on activity patterns within the PMC mask was 74.79% (p = .001). The classifier accuracy was 85.29% (p = .002) for young adults and 64.29% (p = .002) for older adults. The balanced classifier accuracy ran on activity patterns within the PMC mask was 87.29% (p = .001). The classifier accuracy was 85.29% (p = .002) for the young adults and 89.29% (p = .001) for the older adults. Thus, although young and older adults did not differ in the extent to which they were able to reinstate during the Day 2 reinstatement phases the neural activity patterns they had during the Day 1 viewing, these Day 1 neural activity patterns nonetheless differed between the two groups.

3. Recognition task

a. Task description

Following the cued recall task, we tested participants' memory for Session 1 activities using a recognition task (see **Figure S14**). We presented the questions above each activity in the same order as the cued recall test (and the Day 2 movie), but we asked participants to choose which of two still shots appeared in the Day 2 movie. Each shot showed a key frame featuring the criterial feature from the A or B version of the activity without gratings. The pictures appeared on opposite sides of the screen, and the assignment of activity version to the left or right side was counterbalanced across such that the same version did not appear on the same side of the screen more than three times consecutively. Participants chose the left shot by pressing the "1" key and the right shot by pressing the "2" key. The unchosen shot then appeared in the center of the screen, and participants were instructed to indicate whether the activity had appeared in the Day 1 movie.

They responded "yes" by pressing the "1" key (to indicate that the activity had changed) and "no" by pressing the "2" key (to indicate that the activity had repeated). Participants rated their confidence in the accuracy of each response on the same Likert scale as in the cued recall task (ranging from 1 "Low" to 5 "High"). All responses were self-paced.

b. Behavioral results

Consistent with the cued recall task, **Figure S15** shows that Day 2 recognition accuracy was higher for repeated than changed activities [$\chi^2(1) = 16.29$, $p < .001$] and that younger adults outperformed older adults [$\chi^2(1) = 9.03$, $p = .002$]. There was no significant Activity Type × Age Group interaction [$\chi^2(1) = 0.50$, $p = .48$]. In addition, changed activities for which participants reported that both versions had appeared in Session 1 (and thus remembered that a feature changed) were associated with higher Day 2 recognition accuracy than changed activities for which participants did not indicate seeing both versions in Session 1 [$\chi^2(1) = 61.95$, $p < .001$]. There was no significant Activity Type × Age Group interaction [$\chi^2(1) = 0.81$, $p = .37$]. Finally, model estimates indicated that older adults reported having seen both versions of the changed activities significantly less often than younger adults (Older = .64, *95% CI* = [.54, .73]; Younger = .76, *95% CI* = [.68, .82]), [$\chi^2(1) = 4.70$, $p = .03$]. Thus, as in the cued recall task, both age group showed similar positive associations between memory for Day 2 activities and remembering change, but older adults experienced this association on fewer trials.

We next examined whether self-reported reinstatement success and change detection were related to better recognition performance for changed activities. To do so, we fitted models to Day 2 recognition accuracy for changed activities with either self-reported reinstatement success or change detection accuracy, along with age group and their interaction as fixed effects. The model for self-reported reinstatement effects on Day 2 recognition (**Figure S16,** top left panel) indicated no significant effect of Reinstatement [$\chi^2(1) = 2.51$, $p = .11$], a significant effect of Age Group [$\chi^2(1) = 7.46$, $p = .006$], and a significant interaction [$\chi^2(1) = 6.44$, $p = .01$] showing that self-reported reinstatement success was associated with higher Day 2 recognition accuracy for young (z ratio = 2.80, $p = .005$) but not older adults (z ratio = 0.10, $p = .31$). **Figure S16** (bottom left panel) shows that self-reported reinstatement success was also associated with a higher probability of participants remembering that both activity versions appeared in Session 1 [$\chi^2(1) = 12.76$, $p < .001$]. This effect did not significantly interact with Age Group [$\chi^2(1) = 0.23$, $p = .63$]. The model for the change detection effect on Day 2 recognition (**Figure S16,** top right panel) indicated no

significant effect of change detection [$\chi^2(1) = 2.63$, $p = .10$], a significant effect of Age Group [$\chi^2(1) = 5.87$, $p = .02$], and no significant Change Detection × Age Group interaction [*largest* $\chi^2(1) = 2.63$, $p = .10$]. However, **Figure S16** (bottom right panel) shows that change detection accuracy was associated with better memory that both activity version appeared in Session 1 [$\chi^2(1) = 39.22$, $p < .001$], a significant effect of Age Group [$\chi^2(1) = 4.92$, $p = .03$], and this effect was not qualified by an interaction with Age Group [$\chi^2(1) = 1.31$, $p = .25$].

Together, these results closely mirror those of the cued recall task, with the main difference being that for older adults, self-reported reinstatement success did not predict correct recognition of changed Day 2 activities, whereas it did predict higher recall accuracy for the changed Day 2 features. As in the cued recall task, change detection accuracy was more strongly related to memory that a change occurred than for the changed Day 2 features, and this effect did not differ between the two groups.

c. RSA results

We next examined whether recognition performance was related to RSA reinstatement scores to test our hypothesis that higher reinstatement scores would predict better memory for changed activity features. As with the cued recall task, we first fitted models with the mean between- and within-participant reinstatement scores in the PMC and MTL, as well as their respective interaction with Age Group as a fixed effect. Recognition performance predicted by mean between- and within-participant neural reinstatement scores appear in **Figure S17.** Analyses of Day 2 recognition accuracy for changed activities (left panels of **Figure S17**) revealed no significant effect of mean PMC or MTL reinstatement scores, either between- or within-participants, and no significant interaction with Age Group [*largest* $\chi^2(1) = 3.42$, $p = .06$]. In addition, analyses of remembering change (right panels of **Figure S17**), indicated by reporting that both activity versions appeared in Session 1, showed no effect of mean PMC reinstatement scores, either between- or within-participants [*largest* $\chi^2(1) = 2.76$, $p = .10$]. However, for the MTL, there were significant Age Group × Reinstatement Score interactions both between-participants [$\chi^2(1) = 4.72$, $p = .03$] and within-participants [$\chi^2(1) = 8.76$, $p = .003$]. Young adults with higher mean MTL reinstatement scores remembered more changes [$\chi^2(1) = 5.85$, $p = .02$], but older adults did not [$\chi^2(1) = 0.08$, $p = .78$]. Surprisingly, at the within-participant level, trials with higher reinstatement scores predicted worse memory for change for young adults [$\chi^2(1) = 5.67$, $p = .02$], but not for older adults [$\chi^2(1) = 3.08$, $p = .08$].

Next, for the individual parcel scores, we used logistic mixed effect models with between- and within-participant reinstatement scores but added the individual scores of each parcel as fixed effects rather than a single mean value representing the average of all the parcels (as with the cued recall task, see sections 2.2.2, and 2.2.3). We then used the anova function of the lme4 package (9) to determine whether adding these single parcel scores relative to a baseline model comprising only the random effects improved model fits. We then examined whether these model fits were further improved by adding the interaction terms with Age Group as a fixed effect. Regarding Day 2 recognition accuracy for changed activities, adding the between-participant reinstatement scores for the PMC parcels did not improve model fit, and neither did the interaction terms with Age Group [*largest* $\chi^2(12) = 9.94$, $p = .62$]. However, at the within-participant level, adding the fixed effects of the parcels improved model fit [$\chi^2(12) = 8.89$, $p = .004$], but the interaction terms with Age Group did not [$\chi^2(12) = 9.35$, $p = .67$]. Examination of the parameter estimates for individual parcels showed that higher reinstatement scores in parcel 291 were associated with higher Day 2 recognition accuracy [$\chi^2(1) = 12.28$, $p < .001$], whereas higher reinstatement scores in parcel 276 showed were associated with lower Day 2 recognition accuracy [$\chi^2(1) = 4.46$, $p = .03$]. No effects were significant for either the PMC or MTL parcels nor the interaction with Age Group when examining the association between reinstatement scores and memory for change [*largest* $\chi^2(12) = 17.67$, $p = .12$].

Surprisingly, these results do not fully replicate those of the cued recall task at the between-participant level, because mean reinstatement scores were not related to Day 2 recognition accuracy in either age group (although MTL reinstatement scores predicted correctly reporting having seen both versions of the changed activities in the young adults). It is likely that these findings result from the different nature of the two memory tasks. Recall accuracy of the changed Day 2 features relied primarily on recollection processes, whereas recognition of changed features relied relatively less on recollection and more on familiarity. Recent reviews of neuroimaging findings indicate that correct recognition performance does not necessarily rely on activity within posterior default network regions, and that this is particularly the case when participants rely on familiarity rather than recollection processes to guide their answers (28). This lower reliance on PMC and MTL during the recognition task might explain why neural activity pattern reinstatements in these regions did not predictive better memory for changed features.

d. Hippocampal response for the difference between changed and repeated activities, reinstatement scores, and behavioral performance

To further assess the proposal that hippocampal responses to novelty are a critical component of memory updating when experiencing changes in events (13, 23), we next examined whether the difference in the hippocampal response for changed compared to repeated activities predicted Day 2 recognition of changed activities. To do so, we used linear mixed effect models to examine whether adding these parameters estimates and their interaction with Age Group improved model fits above and beyond reinstatement scores. Results showed that this was the case for Day 2 recognition accuracy for changed items [$\chi^2(2) = 6.90$, $p = .03$, for the MTL; $\chi^2(2) = 6.59$, $p = .03$, for the PMC]. Results further showed that higher parameters estimates were predictive of better change detection accuracy [$\chi^2(1) = 5.81$, $p = .02$, for the MTL; $\chi^2(1) = 5.60$, $p = .02$, for the PMC] and that these effect did not interact with Age Group [$\chi^2(1) = 1.23$, $p = .27$, for the MTL; $\chi^2(1) = 1.12$, $p = .29$, for the PMC]. The effects of reinstatement and their interaction with Age Group were not significant [*largest* $\chi^2(1) = 3.47$, $p = .06$]. Adding the parameter estimates did not improve model fit for the tendency to remember change in the MTL [$\chi^2(2) = 4.77$, $p = .09$] or PMC [$\chi^2(2) = 3.94$, $p = .14$].

Together, these results complement those observed in the cued recall task (see section 2.3). Although the intensity of difference in hippocampal responses for changed compared to repeated event viewing does not seem to facilitate subsequent episodic recollection, the present results show that it is nonetheless associated with better Day 2 recognition accuracy for the changed event features. A possible explanation for these discrepant results is that the intensity of hippocampal response has a specific effect on correct recognition based on familiarity processes that are less necessary in the cued recall task. Future studies should investigate whether participants based their answers on recollection or familiarity during similar recognition tasks to assess this possibility.

4. References

1. W. C. Shipley, C. C. Burlingame, A convenient self-administering scale for measuring intellectual impairment in psychotics. *AJP* **97**, 1313–1325 (1941).

2. W. Schneider, A. Eschman, A. Zuccolotto, *E-Prime user's guide* (Psychology Software Tools Inc., 2012).

3. M. F. Folstein, S. E. Folstein, P. R. McHugh, "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189–198 (1975).

4. R. L. Buckner, *et al.*, A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *NeuroImage* **23**, 724–738 (2004).

5. J. L. R. Andersson, M. Jenkinson, S. Smith, Non-linear registration, aka spatial normalisation. *FMRIB technical report TR07JA2* (2010).

6. M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, S. M. Smith, FSL. *NeuroImage* **62**, 782–790 (2012).

7. J. L. R. Andersson, S. Skare, J. Ashburner, How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage* **20**, 870–888 (2003).

8. J. Schrouff, *et al.*, PRoNTo: pattern recognition for neuroimaging toolbox. *Neuroinformatics* **11**, 319–337 (2013).

9. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, 1–48 (2015).

10. J. Fox, S. Weisberg, *An {R} Companion to Applied Regression*, Second (Sage, 2011).

11. R. Lenth, emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.3.5.1. (2019).

12. D. Lüdecke, sjPlot: Data Visualization for Statistics in Social Science, R package version 2.7.1 (2019).

13. C. N. Wahlheim, J. M. Zacks, Memory guides the processing of event changes for older and younger adults. *Journal of Experimental Psychology: General* **148**, 30–50 (2019).

14. L. Hasher, R. T. Zacks, "Working memory, comprehension, and aging: A review and a new view" in *The Psychology of Learning and Motivation: Advances in Research and Theory, Vol. 22*, (Academic Press, 1988), pp. 193–225.

15. D. H. Kausler, *Learning and memory in normal aging* (Academic Press, 1994).

16. C. N. Wahlheim, B. H. Ball, L. L. Richmond, Adult age differences in production and monitoring in dual-list free recall. *Psychology and Aging* **32**, 338–353 (2017).

17. D. Stawarczyk, M. A. Bezdek, J. M. Zacks, Event Representations and Predictive Processing: The Role of the Midline Default Network Core. *Topics in Cognitive Science* (In press) https:/doi.org/10.1111/tops.12450 (September 6, 2019).

18. R. Leech, D. J. Sharp, The role of the posterior cingulate cortex in cognition and disease. *Brain* **137**, 12–32 (2014).

19. A. Schaefer, *et al.*, Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb Cortex* **28**, 3095–3114 (2018).

20. R. Leech, R. Braga, D. J. Sharp, Echoes of the brain within the posterior cingulate cortex. *J Neurosci* **32**, 215–22 (2012).

21. M. C. Anderson, J. H. Neely, "Interference and inhibition in memory retrieval." in *Memory*, 2nd Ed., E. L. Bjork, R. A. Bjork, Eds. (Academic Press, 1996), pp. 237–313.

22. W. D. Spencer, N. Raz, Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging* **10**, 527–539 (1995).

23. K. A. Norman, E. L. Newman, G. Detre, A neural network model of retrieval-induced forgetting. *Psychological Review* **114**, 887–953 (2007).

24. M. Brett, J.-L. Anton, R. Valabregue, J.-B. Poline, Region of interest analysis using an SPM toolbox in (2002).

25. P. Zeidman, E. A. Maguire, Anterior hippocampus: the anatomy of perception, imagination and episodic memory. *Nat Rev Neurosci* **17**, 173–82 (2016).

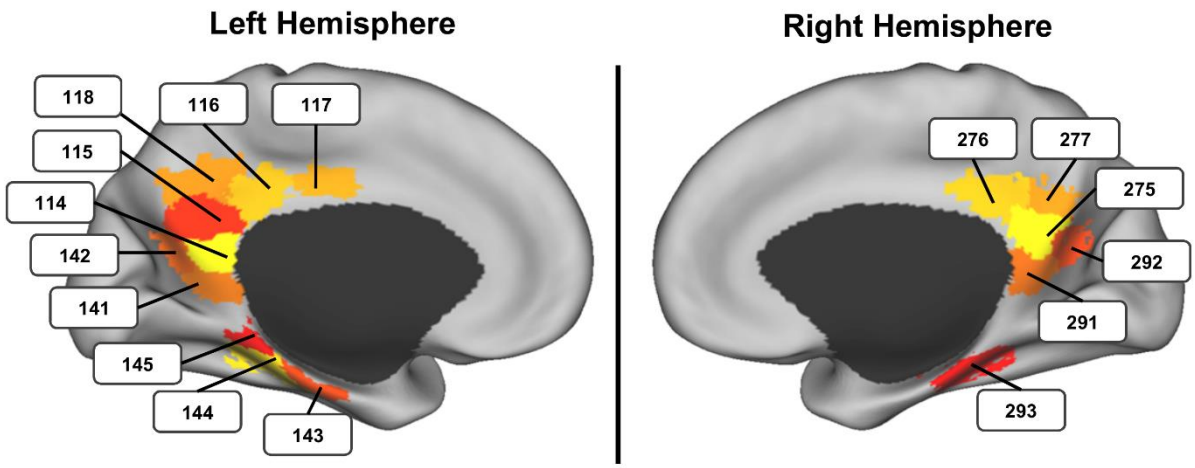
26. A. Kafkas, D. Montaldi, How do memory systems detect and respond to novelty? *Neuroscience Letters* **680**, 60–68 (2018).

27. A. C. H. Lee, K. H. Brodersen, S. R. Rudebeck, Disentangling spatial perception and spatial memory in the hippocampus: a univariate and multivariate pattern analysis fMRI study. *J Cogn Neurosci* **25** (2013).

28. H. Kim, Differential neural activity in the recognition of old versus new events: an activation likelihood estimation meta-analysis. *Hum Brain Mapp* **34**, 814–36 (2013).

5. Figures

Figure S1. PMC and MTL parcels of interest.



Note: PMC and MTL parcels of interest of the DN subsystems from the 300 parcels of the 17 networks cortex parcellation map of Schaefer et al. (19). The numbers attributed to each parcel correspond to the labels in the original parcellation map. Parcels 143-145 and 293 correspond to the MTL, all other parcels correspond to the PMC.

Figure S2. Reinstatement scores across the full set of parcels.

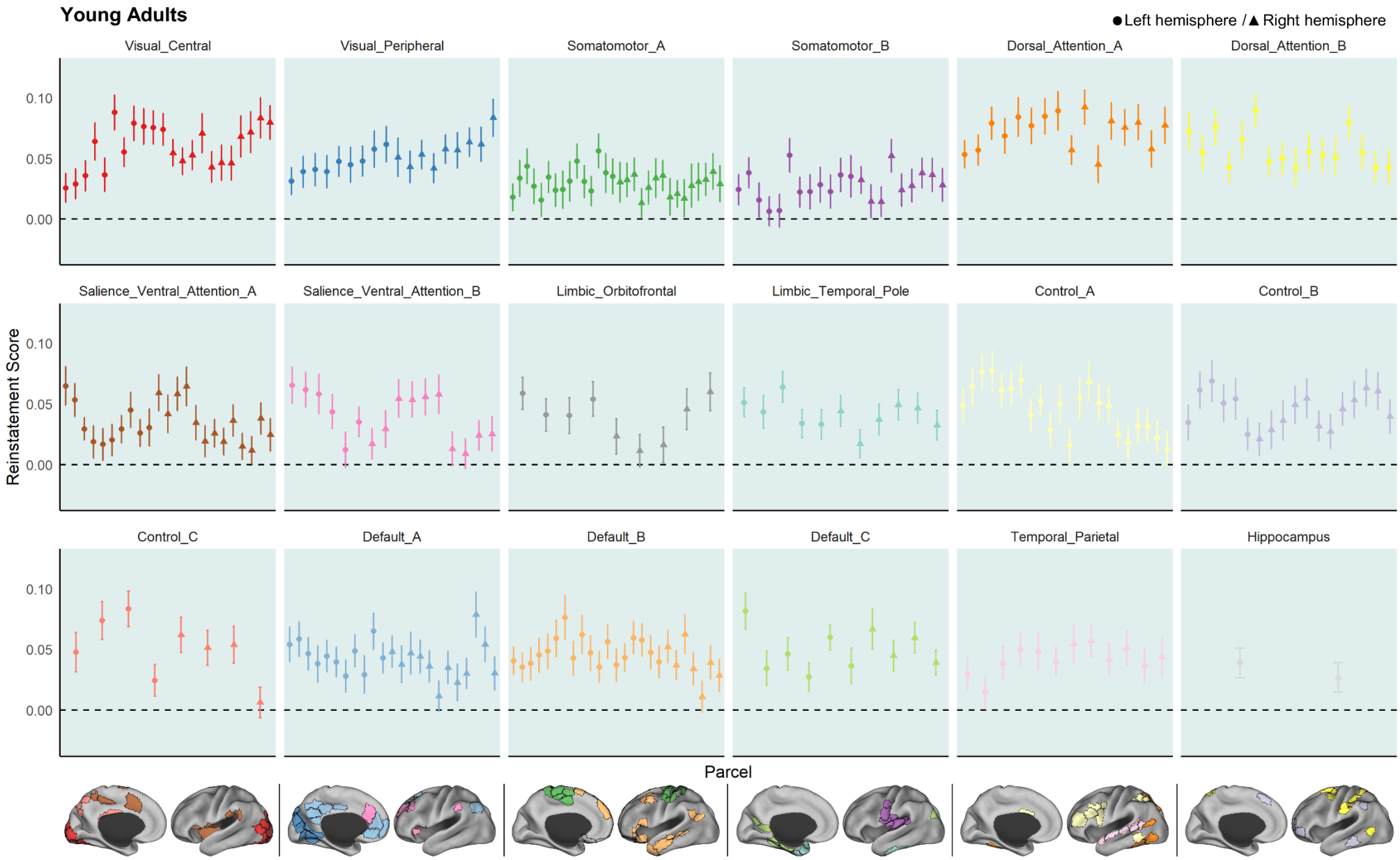
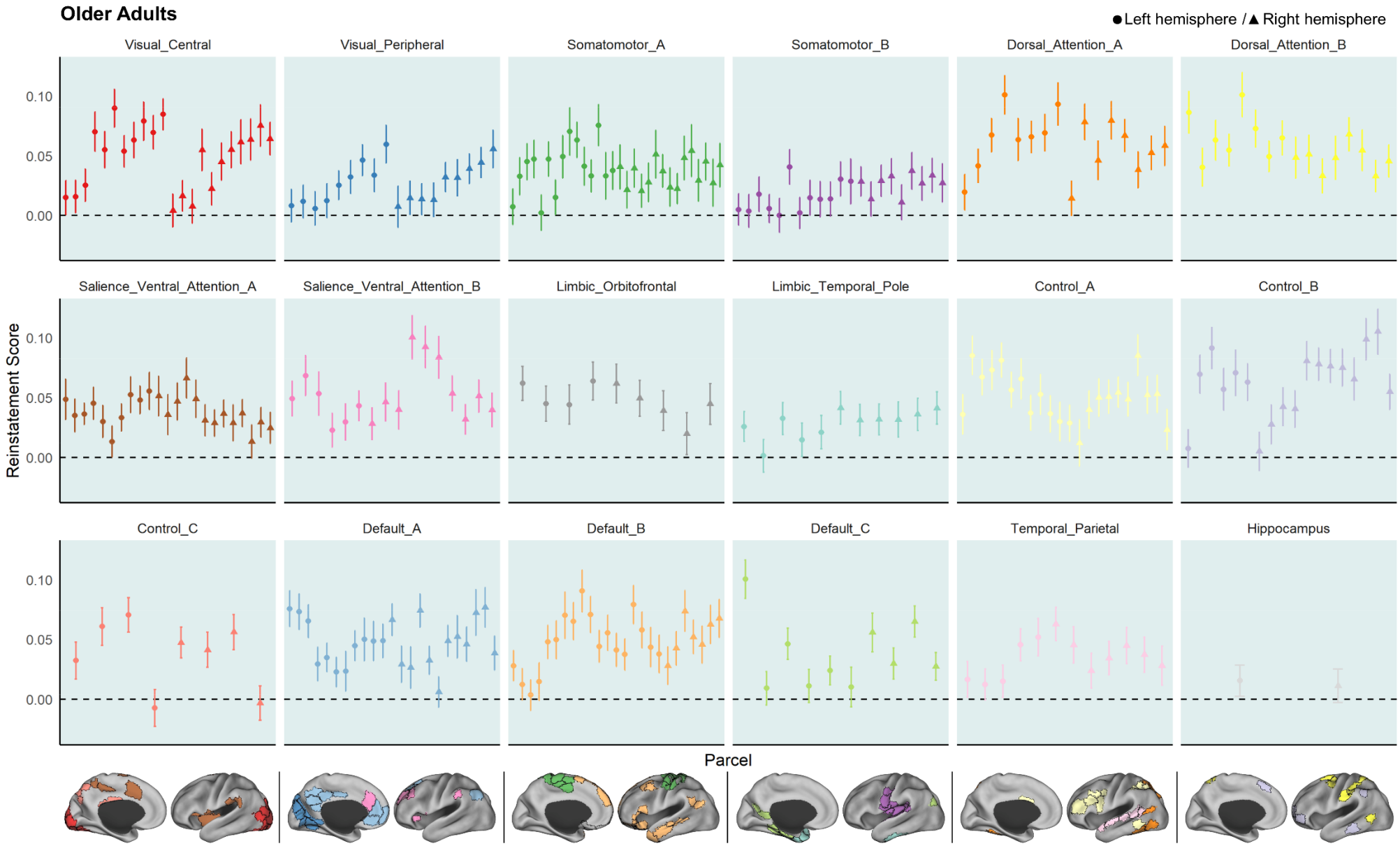
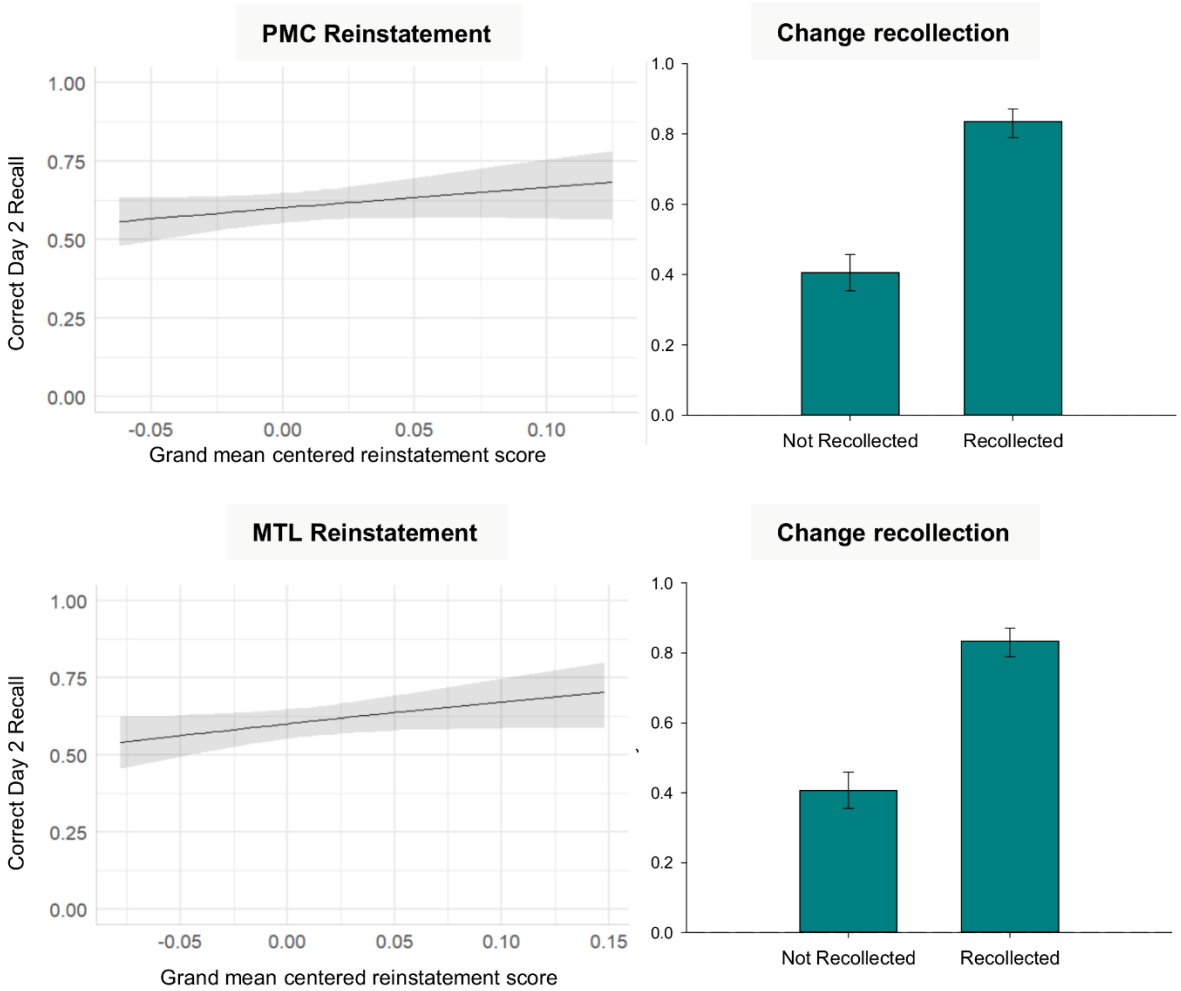


Figure S2. Reinstatement scores (continued).



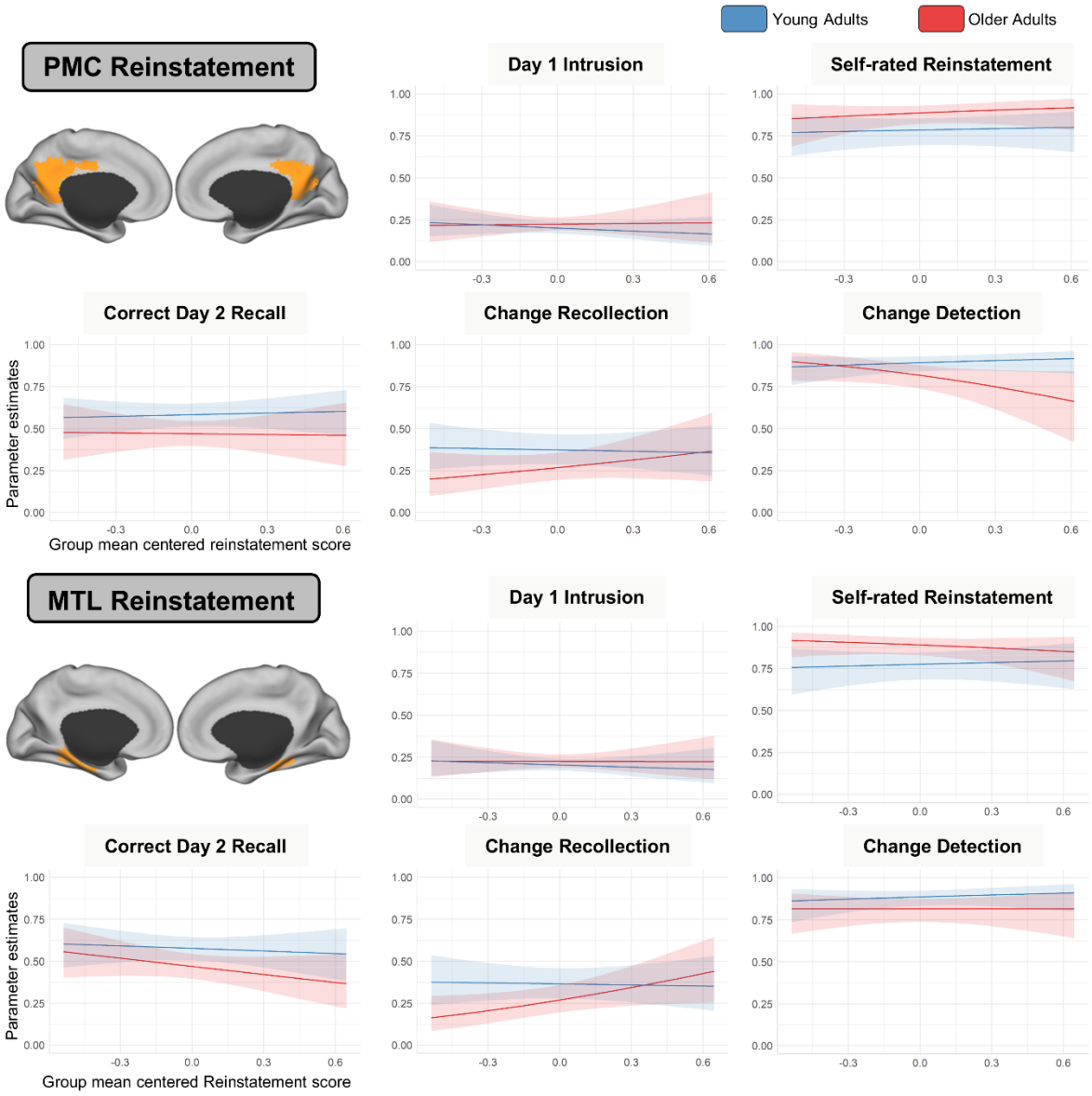
Note: Mean reinstatement scores and 95% confidence intervals computed over all the trials of each group for the 300 parcels of the 17 network parcellation of Schaefer et al. (19). The parcels and networks are illustrated in the bottom row and ordered within each hemisphere following the original parcellation numbering.

Figure S3. Day 2 recall accuracy by mean PMC/MTL reinstatement scores (between-participant level) and change recollection accuracy for young adults.



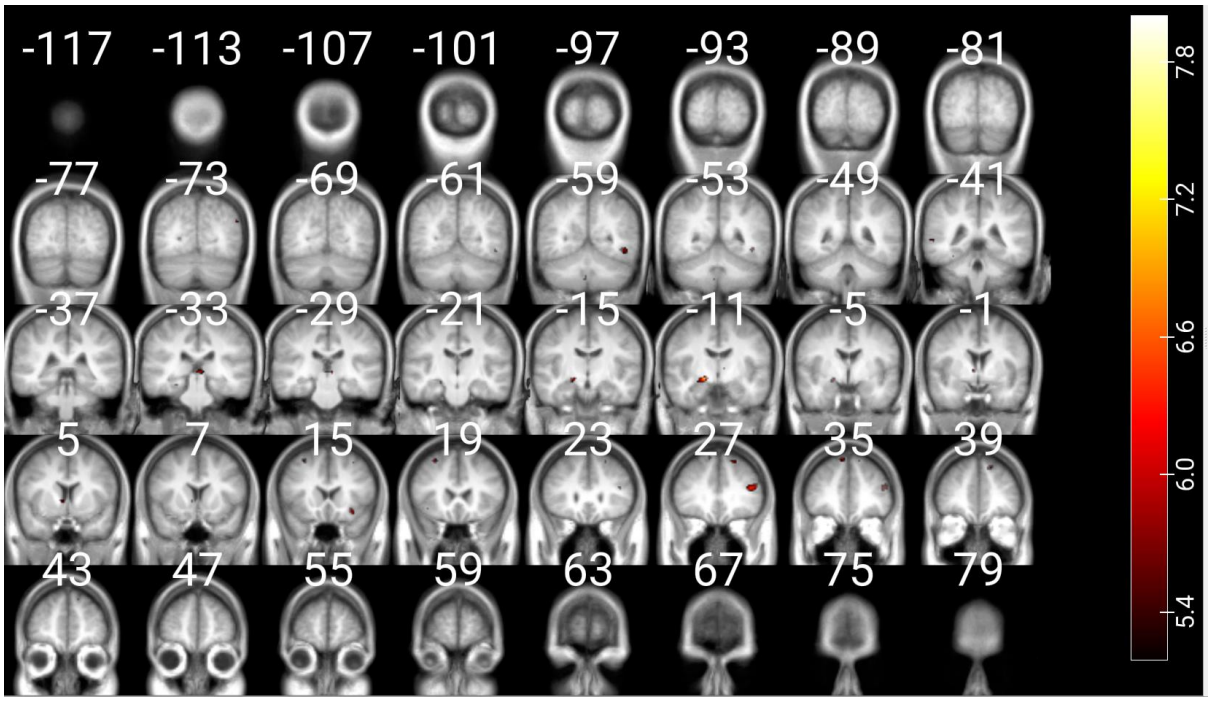
Note. Shaded regions and error bars are 95% confidence intervals.

Figure S4. Within-participant associations between neural reinstatement scores and behavioral performance in the cued recall task measures.



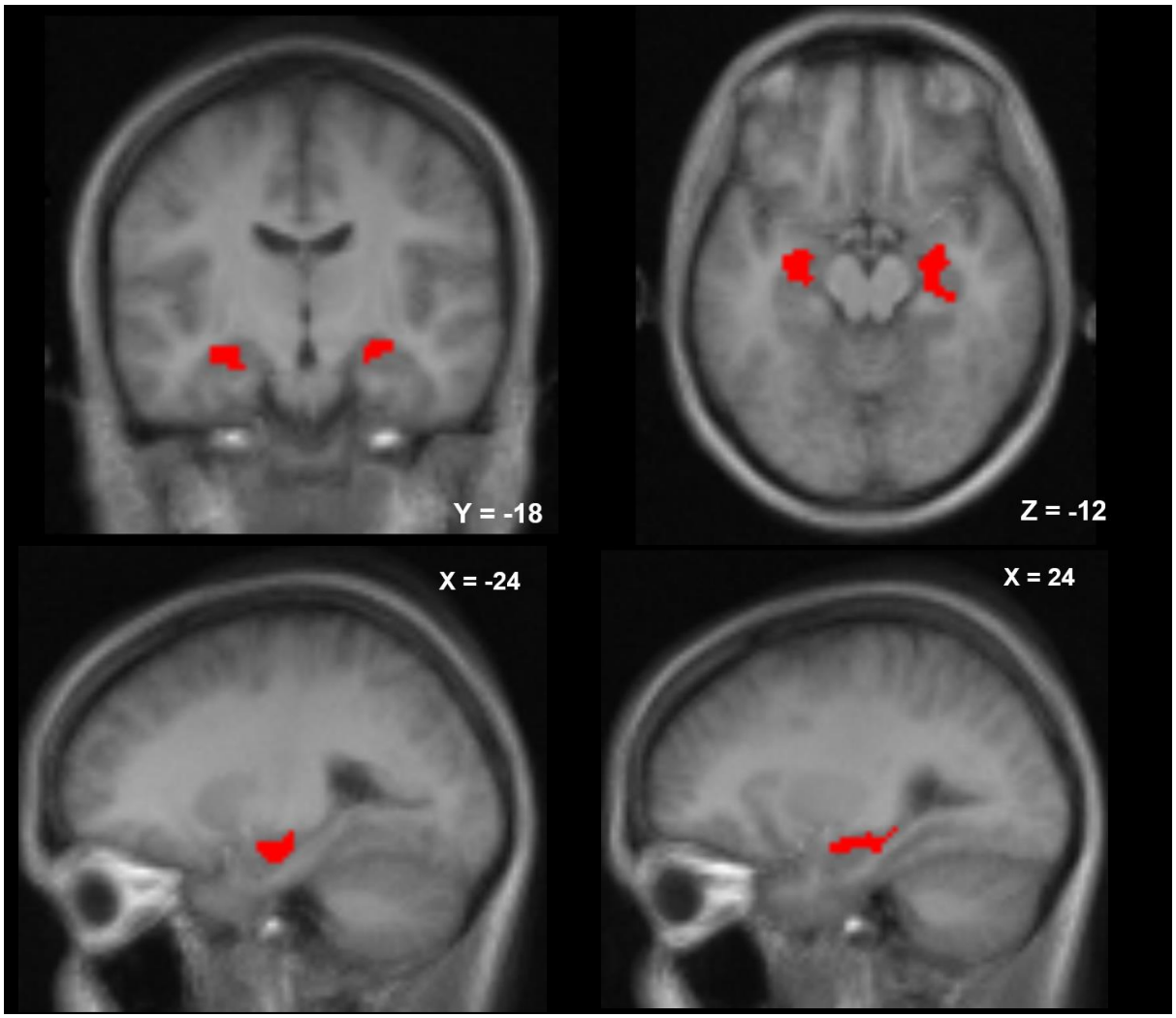
Note: The top two rows display associations with PMC reinstatement, and the bottom two rows display associations with MTL reinstatement. Shaded regions are 95% confidence intervals.

Figure S5. Brain regions more activated while watching changed vs repeated post-divergence segments during the Day 2 viewing across all participants (*N* = 62).



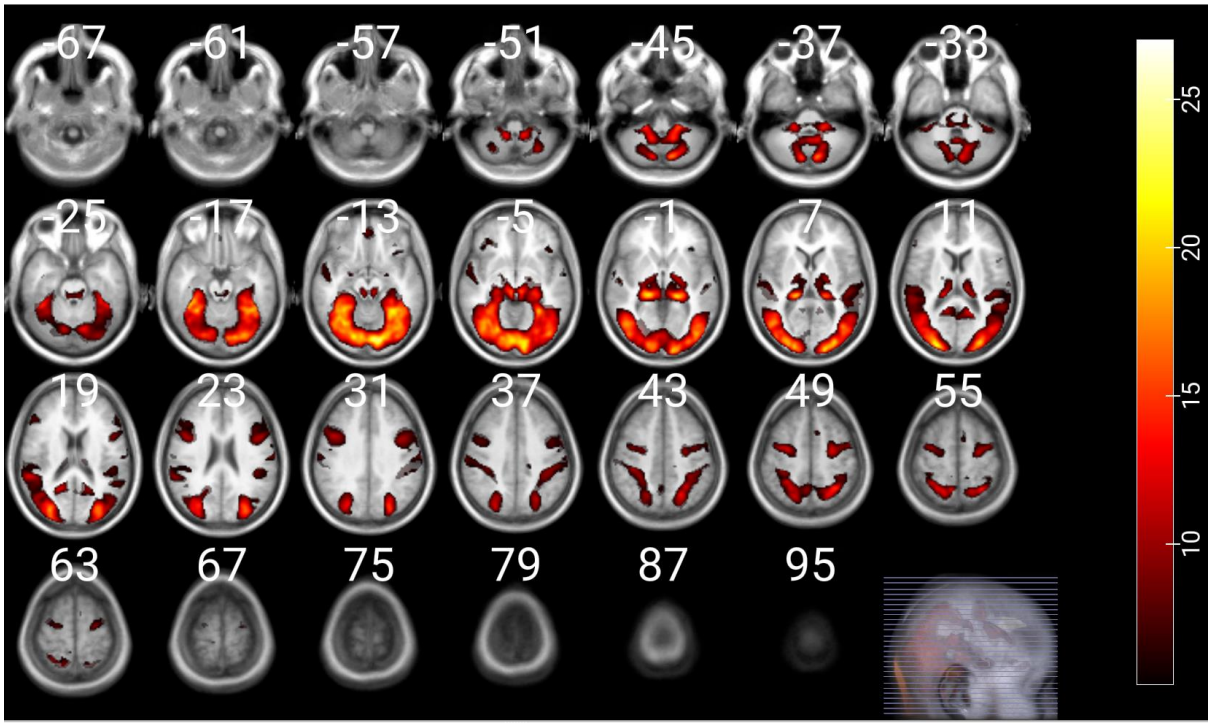
Note: p < .05 Family-Wise Error (FWE) corrected over the whole brain. The color bar represents *t*-values.

Figure S6. Bilateral anterior hippocampal cluster is more activated while watching changed than repeated post-divergence segments during Day 2 viewing across all participants (*N* = 62).



Note: p < .05 FWE small-volume corrected over a mask of the bilateral hippocampus.

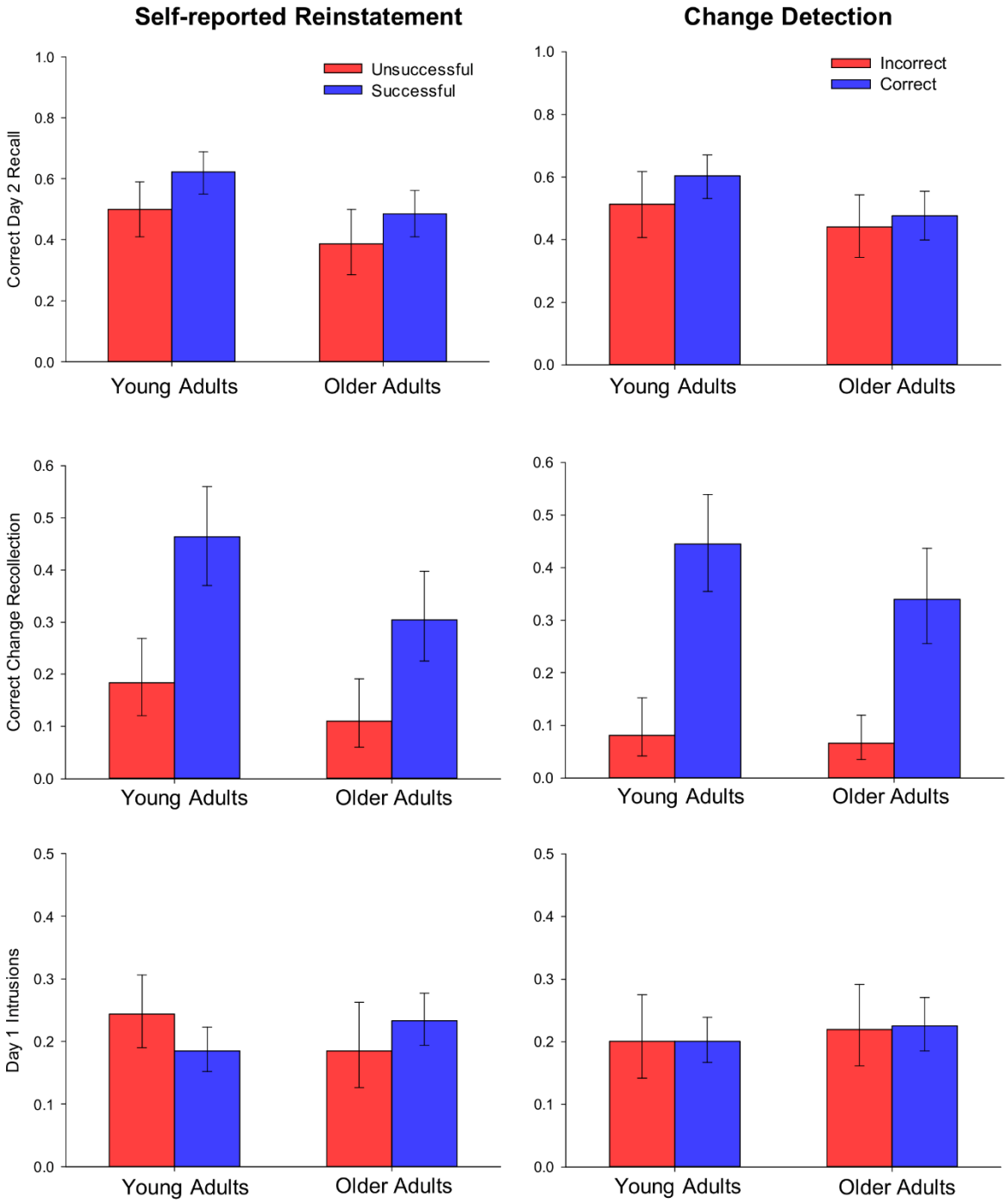
Figure S7. Brain regions that are more activated than baseline while watching activities during the Day 1 viewing across all participants (*N* = 62).



Note: $p < .05$ FWE corrected over the whole brain. The color bar represents *t*-values.

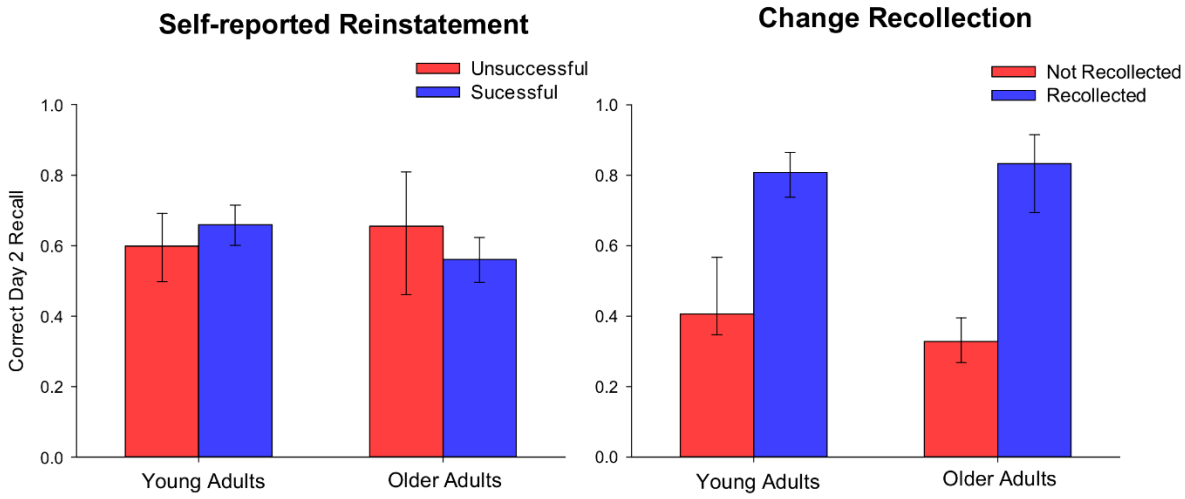
Figure S8. Illustration of the gratings overlayed on the movies during the first 1.5 s of the post-divergence segments of the activities.

Figure S9. Performance in the cued recall task for changed activities for young and older adults as a function of self-reported reinstatement success and change detection accuracy.



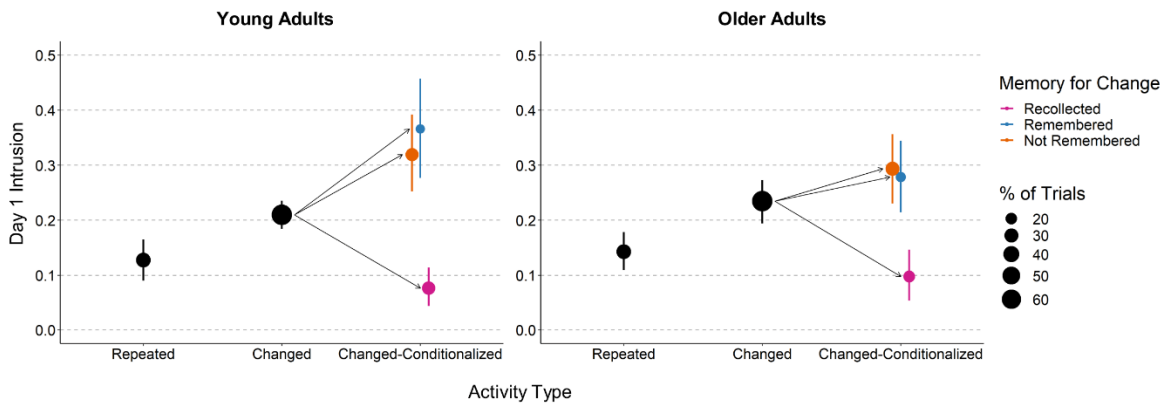
Note: The parameter estimates are derived from logistic mixed models. Error bars are 95% confidence intervals.

Figure S10. Day 2 recall accuracy conditionalized on self-reported reinstatement and change recollection accuracy for young and older adults.



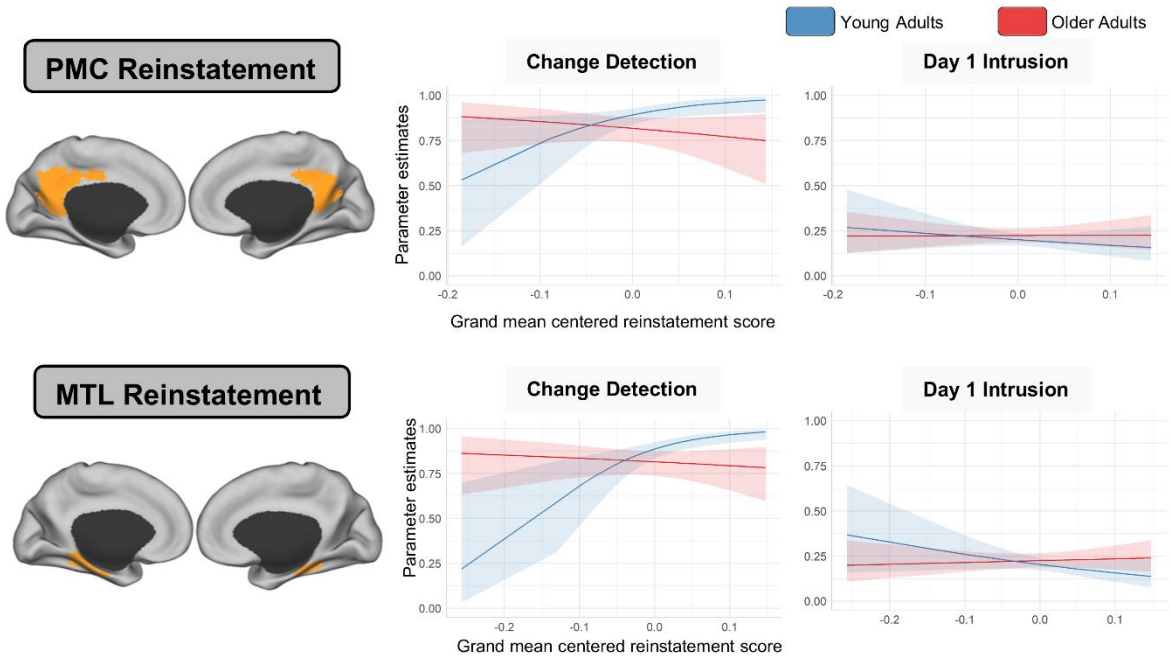
Note: The parameter estimates are derived from a logistic mixed model. Error bars illustrate 95% confidence intervals.

Figure S11. Day 1 Intrusions during Day 2 recall for young and older adults.



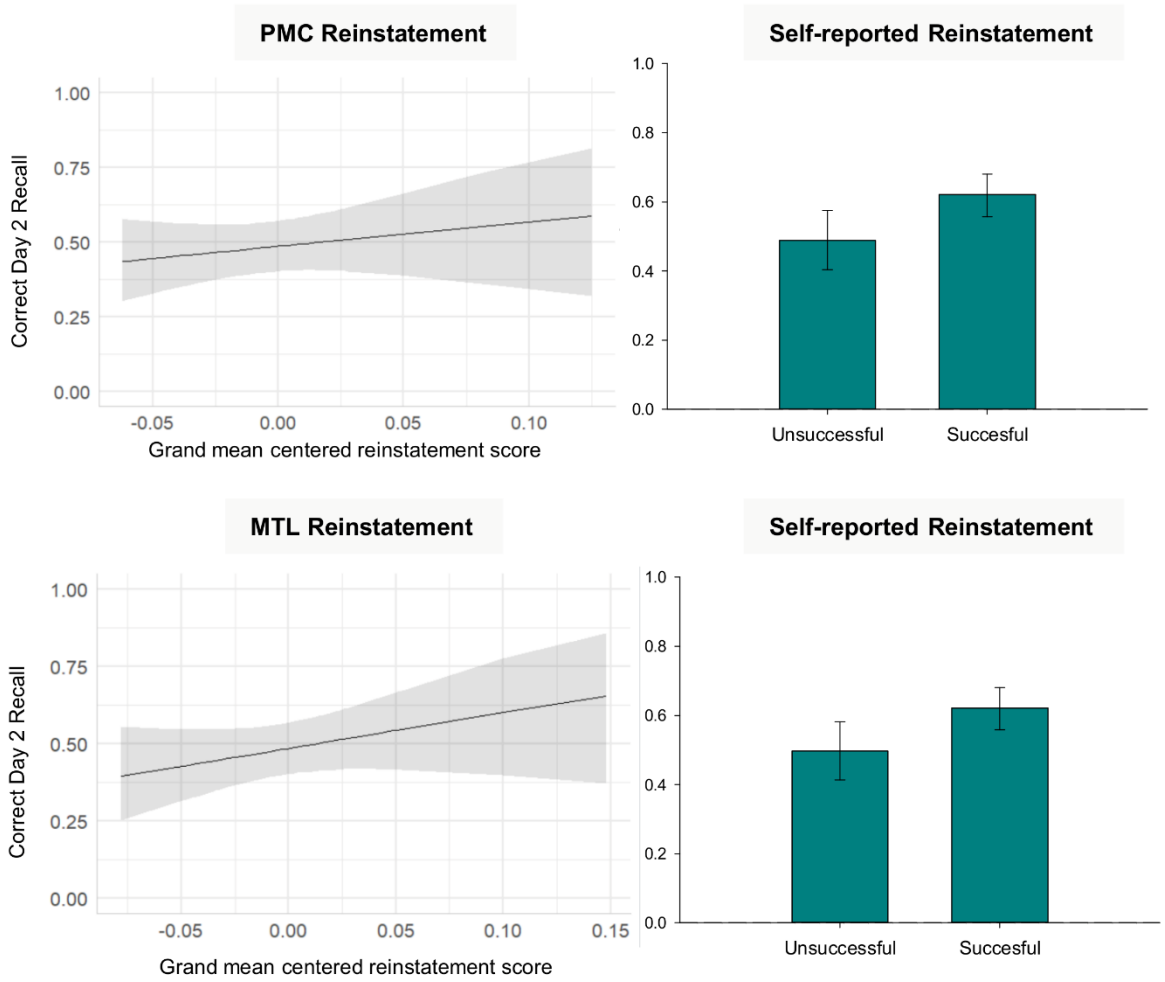
Note: Change could be remembered with recall of Day 1 features (Recollected), remembered without recall of Day 1 features (Remembered), or not remembered at all (Not Remembered). Day 1 intrusions for Repeated activities are baseline estimates of responses that included alternative criterial activity features that did not appear during Day 1 viewing. Error bars are bootstrap 95% confidence intervals.

Figure S12. Between-participant associations between behavioral measures (Day 1 Intrusions and Change Detection) and mean PMC/MTL reinstatement scores.



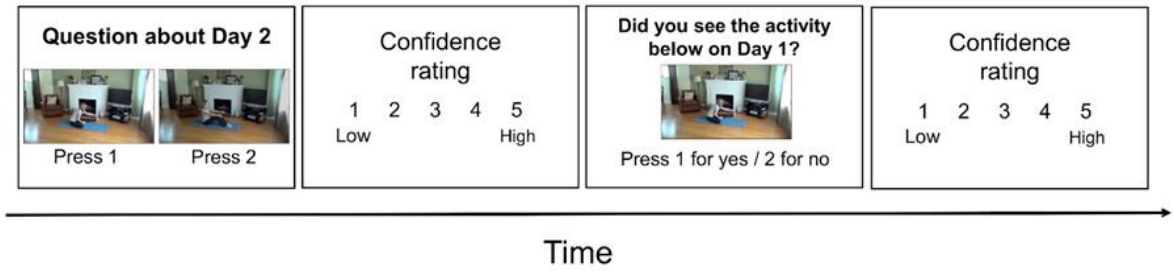
Note: Shaded regions are 95% confidence intervals.

Figure S13. Day 2 recall accuracy by mean PMC/MTL reinstatement scores (between-participant level) and self-reported reinstatement accuracy above and beyond the effect of the other variable for young adults.



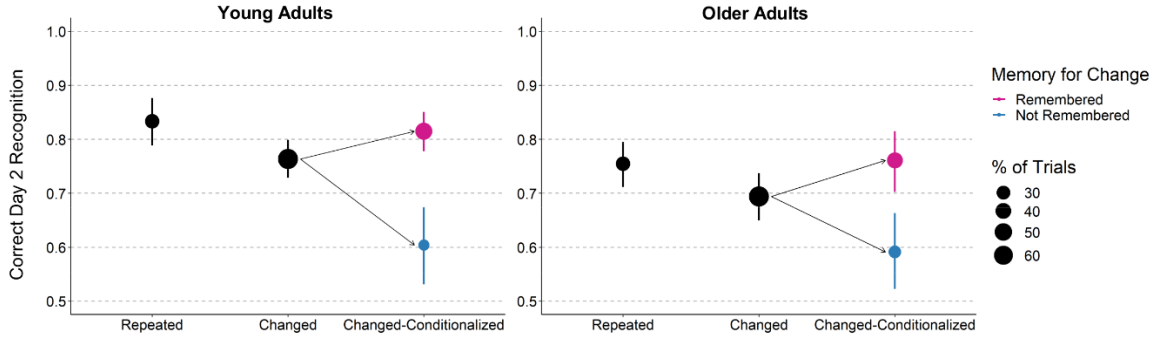
Note: Shaded regions and error bars are 95% confidence intervals.

Figure S14. Schematic of the recognition task procedure.



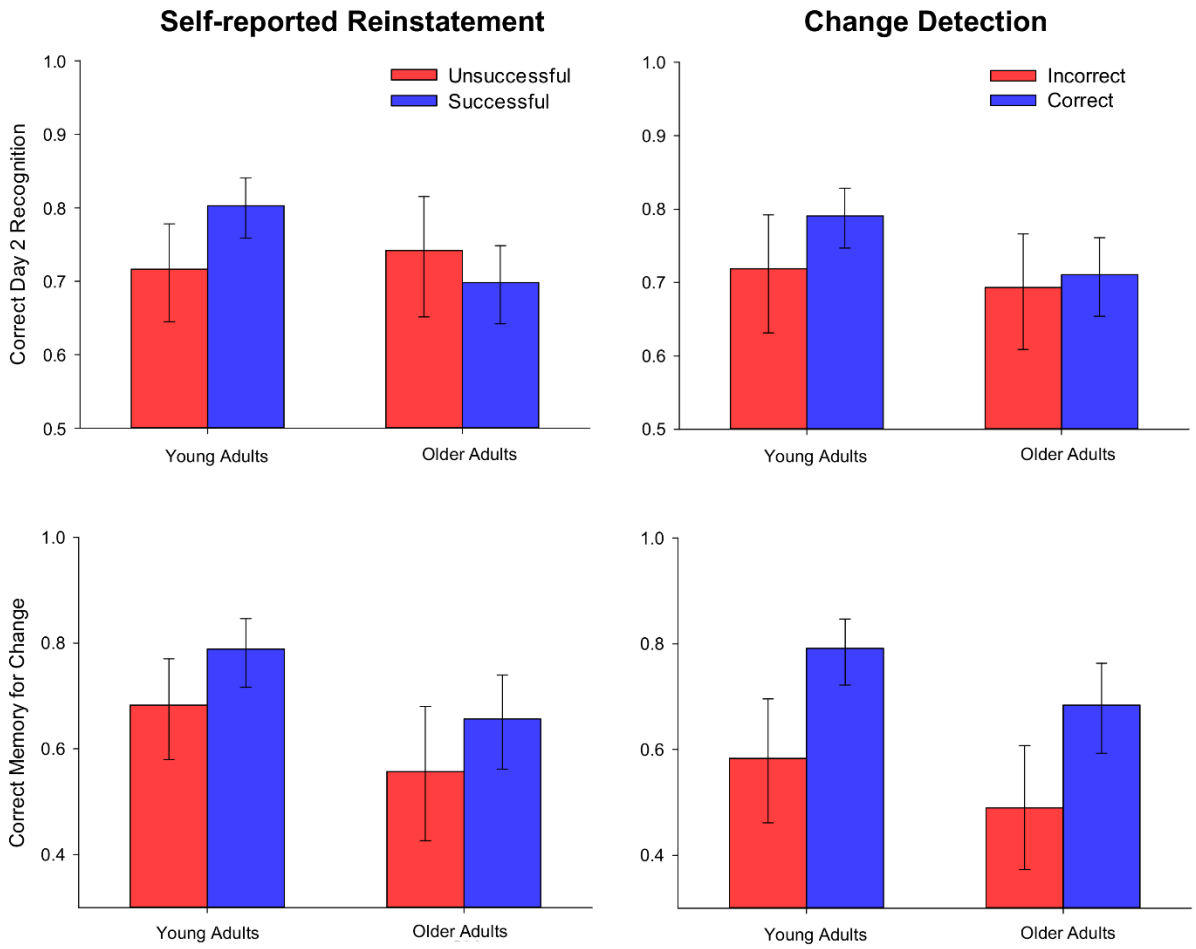
Note: In the recognition task shown above, participants were first asked to choose from two still shots the one depicting the activity performed on Day 2. Next, the still shot that participants did not choose appeared in the center of the screen. Participants were asked whether that activity had appeared on Day 1. We interpreted a "yes" response to that question as indicating that participants had remembered seeing both versions of the activity, and therefore, that they considered the activity to be changed. We interpreted a “no” response to that question as indicating that participants did not remember seeing both activities, and therefore, that they did not consider the activity to be changed (and often considered it to be repeated). Participants rated their confidence in each response. Responses to each step above were self-paced.

Figure S15. Correct recognition of Day 2 activities for young and older adults.



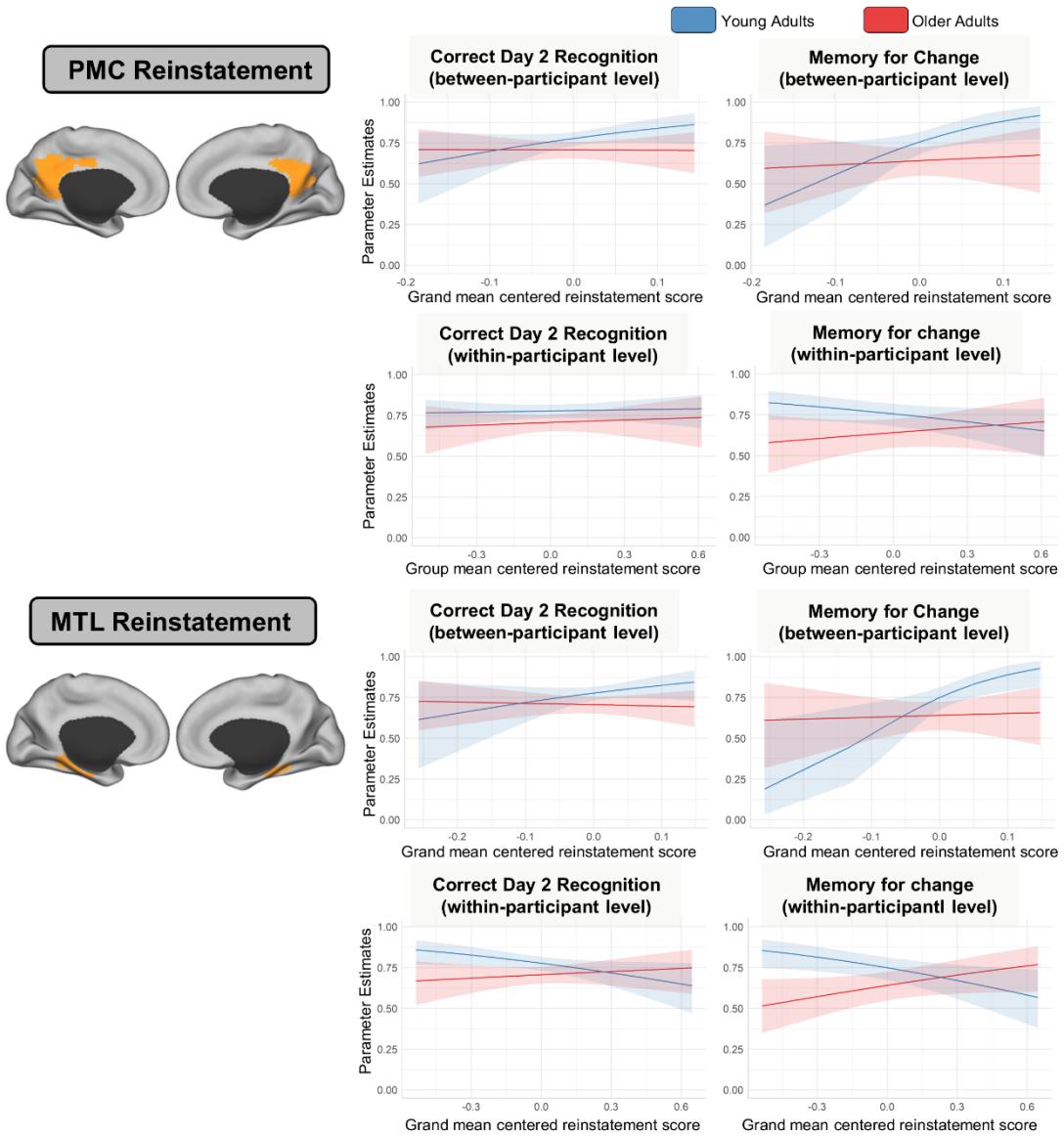
Note: Memory for change is based on the response to the Day 1 recognition question (see Figure S14). Error bars are bootstrap 95% confidence intervals.

Figure S16. Correct recognition of changed Day 2 activities and correct memory for change in young and older adults conditionalized on self-reported reinstatement success and change detection accuracy.



Note: The parameter estimates are derived from logistic mixed models. Error bars illustrate 95% confidence intervals.

Figure S17. Mean PMC/MTL reinstatement scores predicting recognition memory for changed activities and memory for changes in young and older adults.



Note: Shaded regions are 95% confidence intervals.

6. Tables

Table S1. Comparisons of reinstatement scores for young and older adults for each PMC and MTL parcel.

Parcel	$\chi^2(1)$	*p*-value	Parameter estimates YA [95% CI]	Parameter estimates OA [95% CI]
PMC Left				
p.114	1.17	.28	0.04 [0.02; 0.07]	0.02 [-0.001; 0.05]
p.115	0.32	.57	0.05 [0.02; 0.07]	0.04 [0.01; 0.06]
p.116	0.94	.33	0.04 [0.02; 0.06]	0.02 [-0.0001; 0.05]
p.117	0.03	.87	0.03 [0.004; 0.06]	0.03 [0.004; 0.06]
p.118	0.08	.78	0.05 [0.02; 0.08]	0.05 [0.02; 0.07]
p.141	0.81	.37	0.03 [0.004; 0.06]	0.01 [-0.02; 0.04]
p.142	0.02	.90	0.05 [0.02; 0.07]	0.05 [0.02; 0.07]
PMC Right				
p.275	0.32	.57	0.04 [0.02; 0.05]	0.03 [0.01; 0.05]
p.276	0.01	.90	0.005 [-0.02; 0.03]	0.003 [-0.02; 0.03]
p.277	0.52	.47	0.05 [0.02; 0.07]	0.04 [0.01; 0.06]
p.291	0.89	.34	0.04 [0.02; 0.07]	0.03 [0.002; 0.05]
p.292	0.08	.77	0.07 [0.04; 0.09]	0.06 [0.03; 0.09]
MTL Left				
p.143	0.13	.72	0.02 [-0.004; 0.05]	0.02 [-0.01; 0.05]
p.144	4.47	.03	0.06 [0.04; 0.09]	0.02 [-0.002; 0.05]
p.145	1.32	.25	0.03 [0.003; 0.06]	0.01 [-0.03; 0.04]
Hippo.left	1.19	.27	0.04 [0.02; 0.06]	0.02 [-0.01; 0.04]
MTL Right				
p.293	0.24	.62	0.04 [0.02; 0.06]	0.03 [0.01; 0.06]
Hippo.Right	0.36	.55	0.02 [0.004; 0.05]	0.01 [-0.01; 0.04]

Table S2. Pearson correlation matrix of reinstatement scores.

	114	115	116	117	118	141	142	277	275	276	291	292	143	144	145	Hip.L	293
115	.43																
116	.27	.39															
117	.16	.21	.28														
118	.28	.50	.39	.22													
141	.35	.32	.24	.13	.28												
142	.31	.36	.22	.12	.32	.34											
277	.26	.44	.29	.19	.46	.25	.26										
275	.42	.52	.30	.18	.35	.27	.28	.44									
276	.29	.36	.48	.29	.36	.25	.23	.37	.36								
291	.35	.32	.27	.18	.30	.42	.30	.29	.37	.33							
292	.25	.28	.19	.13	.27	.30	.42	.31	.33	.27	.39						
143	.16	.18	.20	.21	.18	.21	.14	.15	.13	.19	.20	.15					
144	.19	.24	.26	.16	.27	.32	.28	.22	.18	.24	.28	.27	.41				
145	.16	.21	.21	.14	.19	.31	.18	.18	.16	.22	.27	.19	.32	.41			
Hip.L	.18	.20	.18	.18	.20	.22	.18	.16	.17	.19	.20	.17	.42	.47	.28		
293	.20	.25	.21	.18	.22	.25	.23	.22	.22	.29	.34	.27	.34	.37	.26	.29	
Hip.R	.13	.16	.17	.16	.18	.16	.14	.17	.18	.24	.24	.18	.26	.24	.14	.34	.54

Note: The coefficients were obtained by correlating the reinstatement score of each trial across all participants.

Table S3: Self-reported reinstatement success (reinstatement success) and activity type classification accuracy (classification accuracy) during Day 2 viewing based on activity type and age group.

		Activity Type	
Measure	**Age Group**	**Repeated**	**Changed**
Reinstatement Success	Young	.80 [.71, .87]	.79 [.70, .86]
	Older	.89 [.83, .94]	.88 [.81, .93]
Classification Accuracy	Young	.87 [.81, .91]	.88 [.83, .91]
	Older	.79 [.71, .85]	.80 [.73, .85]

Note: 95% confidence intervals are displayed in brackets.

Table S4: Classification accuracy for changed activities (change detection) conditionalized on self-reported reinstatement success (reinstatement) for both age groups.

	Reinstatement	
Age Group	**Successful**	**Unsuccessful**
Young	.93 [.89, .95]	.79 [.69, .86]
Older	.82 [.75, .88]	.79 [.67, .86]

Note: 95% confidence intervals are displayed in brackets.