Differentiation of related events in hippocampus supports memory reinstatement in development

Nicole L. Varga*[1,2], Hannah E. Roome[3], Robert J. Molitor[4], Lucia Martinez[1], Elizabeth M. Hipskind[5], Michael L. Mack[6], Alison R. Preston[¯1,2,4], & Margaret L. Schlichting[¯6]

Affiliations

[1]Department of Neuroscience; University of Texas at Austin; [2]Center for Learning & Memory; University of Texas at Austin; [3]Department of Psychology; Newcastle University; [4]Department of Psychology, University of Texas at Austin; [5]Department of Neuroscience; Baylor College of Medicine; [6]Department of Psychology; University of Toronto

Corresponding Author

*Nicole L. Varga (nvarga@austin.utexas.edu); 856-304-6771

Author Note: ¯Reflects equal contributions

Abstract

Memories for similar experiences can be either differentiated or integrated in the adult brain to support later memory retrieval and memory-based decision making. Yet how children represent related memories remains unknown. Here, children (7-10 years) and adults formed memories for separate yet overlapping events. We then measured how successfully remembered events were represented and reinstated using functional magnetic resonance imaging (fMRI). We found that children formed differentiated representations in hippocampus—such that related events were stored as less similar than unrelated events. Conversely, adults formed integrated representations, including in medial prefrontal cortex (mPFC). Furthermore, hippocampal differentiation among children and mPFC integration among adults tracked neocortical reinstatement of the specific features associated with the individual events. Together, these findings reveal that the same memory behaviors are supported by different underlying representations across development. Specifically, whereas differentiation underlies memory organization and retrieval in childhood, integration exhibits a protracted developmental trajectory.

Introduction

In everyday and educational contexts alike, we regularly encounter overlapping information. In the adult hippocampus, such overlap can yield representations for related events that are alternately less[1-3] or more[4-6] similar to one another relative to unrelated events. While these discrepant differentiated and integrated traces may both enable later neocortical reinstatement of event features, they may ultimately support different behaviors: namely memory specificity and flexibility, respectively. Yet, and despite much theoretical interest in this topic[7-8], how children represent related experiences to enable successful memory reinstatement and memory-guided behavior remains unknown. Here, we measure such representations directly to test the hypothesis that children and adults achieve behavioral success through fundamentally different representational schemes due to ongoing maturation of the hippocampus across childhood and adolescence[8-12]. As children's ability to successfully encode and later access memories for related events predicts their ability to use memories flexibly[13] and in turn academic success[14], characterizing the representations they leverage during behavior is key.

While children can form and retrieve memories for overlapping experiences[12-13,15-16], the neural organizational scheme that supports these behaviors has not been characterized. For example, children can draw novel connections across memories to make inferences[12,17] and generate new knowledge[13,18] at above-chance levels (albeit not as well as adults[12,17-18]). However, behavioral success on such tasks should not be taken as evidence of any particular neural organizational scheme, as it can be supported by a variety of different underlying representational codes[19,20-22]. For instance, inference behavior may be underpinned by either directly accessing integrated neural representations[5-6,23-25] or recombining separate memories[26-28] as needed. Moreover, integration in the adult brain arises even in the absence of any behavioral requirement to make inferences[5-6,23-25], further evidence of the behavior-representation distinction. Current developmental theories stemming from behavioral work alternately suggest that children rely more on either generalization[8] or storage of distinct memory traces[7,27] to represent related events. The coexistence of such perspectives underscores the challenges associated with drawing strong conclusions about representation from behavior alone, as well as the need to measure underlying neural representations directly. Here we address this challenge, notably using the terms "integrated" and "differentiated" to refer strictly to neural organization as distinct from behavior.

One brain region that might be the source of developmental differences in neural organization is the hippocampus, which has been shown to store both integrated and

differentiated traces in adults[1-6]. Consistent with the notion that hippocampus might be particularly important for memory for overlapping experiences, children and adults alike engage it during retrieval of overlapping memories[29]. However, such findings cannot tell us how related experiences are coded in this region relative to one another. Asking this latter question requires consideration of distributed, experience-specific activation patterns. Even among the studies that have employed such pattern analyses[30-33], few have quantified representations of related events as they are experienced[34], let alone how related events are stored and retrieved. However, one recent study suggests that younger participants are less likely to form integrated memories relative to adults[27]. That study examined how participants reinstated a previous memory as they encoded a new, overlapping event. While adults reinstated related memories during encoding, in younger participants such reinstatement was reduced or even absent. This study implies a developmental shift from non-overlapping to increasingly integrated representations, which further complements recent data showing that hippocampal differentiation is operational in childhood[34]. However, neither of these studies measured neural representational schemes for related events that were newly learned, and therefore cannot address how overlapping events are organized in the developing brain to support memory-guided behavior. Here, we measure the organizational schemes of such newly formed hippocampal memories directly in children and adults as they make memory-based decisions, and how those representations relate to successful behavior at different ages.

We also interrogate the relationship between hippocampal representation and neocortical reinstatement of associated memory features. In addition to organizing related events through differentiation or integration, hippocampal representations also bind distributed neocortical traces together, enabling high-fidelity reinstatement of individual event features at retrieval[35]. The way in which hippocampal representations are organized may thus have downstream consequences for neocortical reinstatement, and in turn, memory behavior. Past work has shown that hippocampus plays a key role in guiding neocortical reinstatement in adults, with hippocampal BOLD activation tracking the degree to which memory elements are reinstated in neocortex[36-37]. Stronger neocortical reinstatement, in turn, has been associated with successful and faster memory decisions[38-39]. Here, we ask whether memory reinstatement in higher-order perceptual regions ventral temporal (VTC) and parietal cortex[37-38] predicts upcoming decision success in children as it does in adults[39]. Furthermore, we interrogate whether this degree of neocortical reinstatement tracks how related events are represented relative to one another in hippocampus. We test whether different representational schemes—differentiation and integration–lead to high-fidelity neocortical reinstatement in children and

adults, respectively. Linking successful memory reinstatement to distinct memory organization schemes in children and adults has important implications not just for foundational theories of memory development, but also for educational interventions that seek to maximize knowledge retention and flexibility at different ages[40].

In the present study, children and adults learned to associate different objects with the same person or place. To ask how those memories were represented, we leveraged high-resolution fMRI to compare neural activation patterns for objects that were related through a common person or place versus an unrelated baseline. We tested whether children form either differentiated or integrated traces in hippocampus, adjudicating alternate perspectives stemming from the cognitive developmental literature[7-8]. Prior adult data further implicate medial prefrontal cortex (mPFC) in integration of related events[19,21-22,25], showing that mPFC represents connections across learning experiences[6]. Given PFC[11,41-43] and its connectivity to hippocampus[44] continues to develop into adulthood, we anticipated that adults but not children would show memory integration in mPFC.

We further tested whether these hippocampal and mPFC organizational schemes were related to neocortical reinstatement by using neural pattern analyses to measure reactivation of associated memory features (i.e., the related person/place). Our goal was to test whether differentiation in children's hippocampus tracks reinstatement on a trial-by-trial basis, and whether the same is true in adults for integration in hippocampus and/or mPFC. To this end, we focused on middle childhood (7-10 years), as children can retrieve associative features[45] and overlapping[13] events by this age. Although such behavioral evidence suggests that the developing brain can support successful recollection of specific details, here we ask a more detailed question: Namely, whether even similar memory behaviors in children and adults might be underpinned by either different hippocampal or mPFC memory representations or different capacities for high-fidelity neocortical reinstatement.

Results

Children and adults learned overlapping pairs. Twenty-five adults (*M*=19.06, *SD*=3.36) and twenty-seven children between the ages of 7 and 10 years (*M*=9.10, *SD*=1.11) learned to associate individual objects with a person or place (Figure 1a). The same person or place (six each) was paired with three different objects, for a total of 36 overlapping pairs. We used stimuli familiar to participants to encourage high-level performance (Methods), which in turn allowed us to ask about the neural organizational schemes that enabled successful memory for related events at different ages. Moreover, before MRI scanning (Figure 1b),

participants completed between two and five learning/retrieval repetitions until a 90% performance criterion was reached (Figure 1c; Supplementary Data).

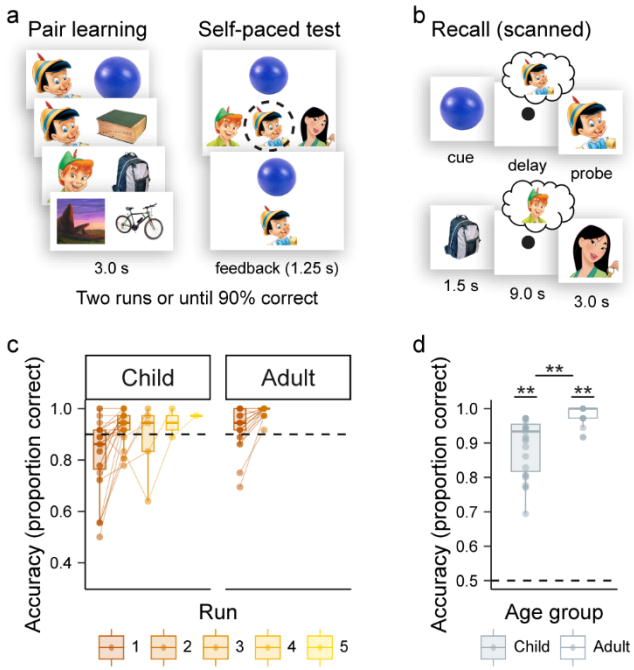

Figure 1. **Procedure schematics and corresponding performance on the behavioral tasks. a** Initial learning/retrieval task. On each run, participants studied the same 36 person-object and place-object pairs and then completed a self-paced three-alternative forced choice memory test with corrective feedback. **b** Scanned recall task. High-resolution fMRI activity was measured while participants viewed previously learned objects and retrieved the target associates (person or place). Associative memory recall was considered successful based on performance on a subsequent mnemonic decision task, in which participants made a button-press response to indicate whether a same-category memory probe (person or place) matched or mismatched the item retrieved. We measured the degree to which neural representations differentiated or integrated overlapping memories during the object cue period, as well as the specificity and degree of ensuing neocortical reinstatement of the associated elements during the delay period. **c** Initial learning/retrieval accuracy. The proportion of correctly selected forced-choice associates was averaged across trials, separately by age group. Dots reflect individual participant means. The lines connecting dots across runs depict within-participant learning trajectories. The number of participants varied by run, depending on how many learning/retrieval repetitions were required to reach the 90% learning criterion (depicted by the dashed black line): Runs 1 and 2: n = 52 (27 children; 25 adults); Run 3: n = 5 children; Run 4: n = 2 children, Run 5: n = 1 child. **d** Scanned recall accuracy. The proportion of correct match/mismatch decisions was averaged across runs and participants, separately for each age group. Dots reflect individual participant means. The dashed black line reflects chance-level performance (0.50). Asterisks above each bar reflect significant above-chance performance at the individual group level while asterisks between bars indicate a significant group difference at a threshold of $p < .001$ (**). For **c** and **d**, box plots depict the median (middle line), 25th and 75th percentiles (the box), and the largest individual participant means no greater than the 5th and 95th percentiles (the whiskers). Individual participant means that extend beyond the whiskers are considered outliers, defined as values that were 1.5 times greater than the interquartile range (IQR).

Following learning, participants recalled the pairs during MRI scanning. Participants were cued with an object followed by a 9-second delay, during which time they were instructed to hold the associated person or place in mind for an upcoming match/mismatch memory decision (Figure 1b). Both children and adults retrieved the overlapping memories, as evidenced by high and above-chance performance in both groups (all *t*s>25.39, all *ps*<.001; Figure 1d). Yet children still performed worse than adults ($t(31.49)=5.62$, $p<.001$, d=1.51, 95% CI=[0.06,0.12]; Figure 1d). Thus, as a further step to isolating the neural organizational schemes and memory reinstatement mechanisms that enabled successful memory, unless otherwise noted, subsequent analyses were restricted to trials in which retrieval was successful during final learning and scanned recall (Methods; Supplementary Table 1 for trial counts).

Children differentiate related memories. To assess organization of related memories in anatomical hippocampus, we used representational similarity analysis (RSA)[46] to quantify the similarity among activation patterns evoked by object cues that were never directly observed together, but that shared an overlapping (person/place) memory element (Figure 2a). In such cases, the related objects might evoke patterns that are significantly either more or less similar to one another as compared to unrelated objects. For such unrelated events, there are no shared links through which to organize the events, thus serving as an effective baseline.

There was a trending effect between related and unrelated objects ($F(1,50)=3.64$, $p=.06$, $\eta_p^2=.07$) that was qualified by an interaction with age ($F(1,50)=4.65$, $p=.04$, $\eta_p^2=.09$), such that children ($F(1,26)=6.06$, $p=.02$, $\eta_p^2=.19$, 95% CI=[-0.02,-0.002]) but not adults ($F(1,24)=.05$, $p=.82$, $\eta_p^2=.002$, 95% CI=[-0.005,0.007]) exhibited hippocampal representations for related objects that were *less* similar than those for unrelated objects (Figure 2b). Notably, this differentiation pattern was absent in visual regions that may instead reflect perceptual similarities among objects (Supplementary Figure 1a) and in neocortical regions implicated in reinstatement of associated memory elements (Supplementary Figures 1b-c). Moreover, this effect remained even when we controlled for hippocampal BOLD activation during the cue (Supplementary Data). Our approach thus enabled us to uncover subtle differences in the organization of hippocampal memory traces not reflected in neocortex and that were above and beyond mean hippocampal activation.

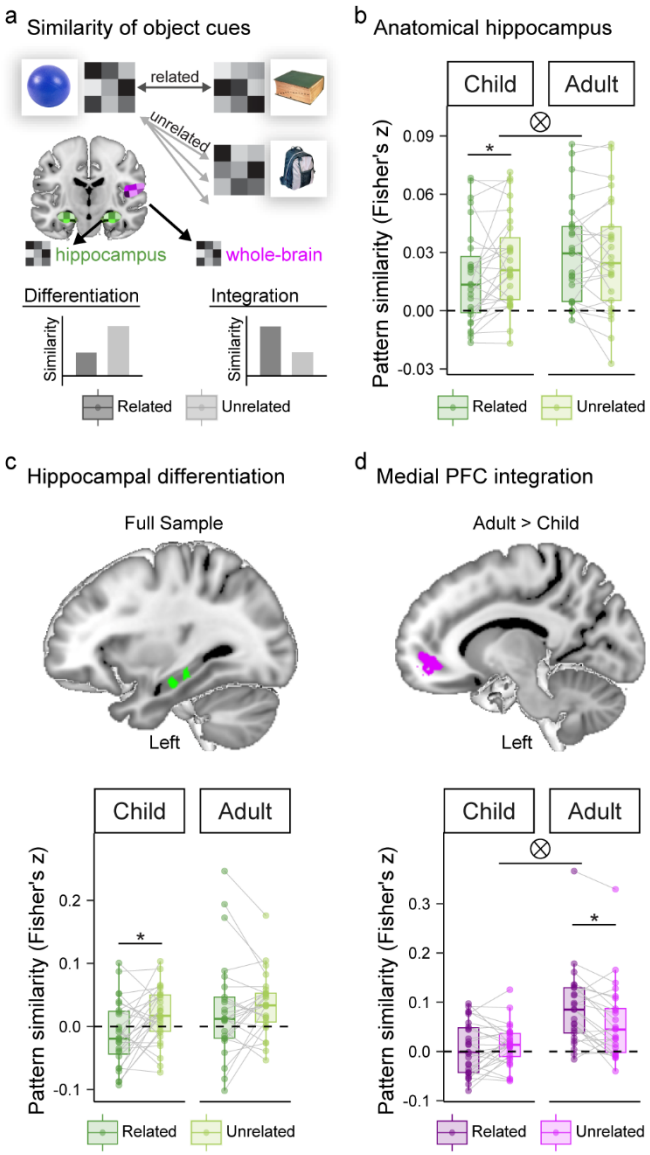

Figure 2. **Children differentiate related object cues, while adults integrate. a** Depiction of representational similarity analyses. We compared the similarity of fMRI patterns for related as compared with unrelated objects to assess evidence for differentiation (related < unrelated) and integration (related > unrelated). **b** Neural patterns extracted from bilateral anatomical hippocampus. Box plots reflect pairwise similarity values averaged across object cues that were related or unrelated, separately for participants within each age group. **c** Hippocampal searchlight regions showing a significant main effect of differentiation when examined across the full sample displayed on the 1-mm MNI template brain. Clusters are significant after small-volume correction for multiple comparisons within anatomical hippocampus. Interrogating effects within these clusters in each age group separately revealed significant differentiation in children but not adults (boxplot). **d** Whole-brain searchlight region showing a significant effect of integration in medial PFC which interacted with age group, with adults showing significantly greater integration than children. Cluster is significant after correction for multiple comparisons across the whole brain. Interrogating effects within this cluster in each age group separately

showed that integration was significant in adults, but children showed no evidence of integration nor differentiation. This analysis also revealed a significant age interaction, in line with the fact that this searchlight cluster was defined based on this effect (marked here simply for clarity). Additional whole-brain results related to **c** and **d** are reported in Supplementary Table 2 and Supplementary Figures 2 & 3. For **b**-**d**, box plots depict the median (middle line), 25th and 75th percentiles (box), and the largest individual values no greater than the 5th and 95th percentiles (whiskers). Dots reflect individual participant means; lines connecting dots depict within-participant differences; n = 52 (27 children; 25 adults). Individual participant values (dots) that extend beyond the whiskers are considered outliers, defined as values that were 1.5 times greater than the interquartile range (IQR). Asterisks reflect a significant related/unrelated difference within an age group (*$p < .05$) while tensor product symbols indicate a significant interaction between age group and related/unrelated conditions at **b** $p < .05$ or **d** $p < .001$.

The findings across our *a priori* anatomical hippocampal region of interest (ROI) establish that children, but not adults, differentiate related memories. However, we also know that the same memories can be represented as both differentiated and integrated within the same person, within different voxels across hippocampus and neocortex[5-6], raising the possibility that different organizational schemes might be evident within different parts of hippocampus. In separate searchlight analyses, we therefore looked for voxels exhibiting differentiation or integration of related object cues in hippocampus and neocortex (Figure 2a). We searched for voxels that showed developmental differences in representational scheme (two-sample t-test: adult>child; child>adult), as well as voxels that showed a neural scheme across both groups (one-sample t-test of full sample vs. 0). To verify that significant coding effects were evidenced in both age groups separately for clusters identified across the full sample (>0), and for the group showing enhanced organization for clusters showing developmental differences (e.g., adults for adult>child contrast), for all significant clusters we additionally considered whether effects were present in each age group separately.

Results in hippocampus showed differentiation in children but not adults, but no evidence of integration. Specifically, we found no evidence of hippocampal integration in adults nor children when considered together; nor any clusters that showed developmental differences in either hippocampal integration or hippocampal differentiation. By contrast, clusters in left anterior hippocampus (Figure 2c; cluster center of gravity in the Montreal Neurological Institute (MNI) template coordinates (mm): x, y, z = -32.1, -18.7, -17.5) and left hippocampal body (Figure 2c: x, y, z = -29.7, -27.2, -12.8) showed significant differentiation across the whole group. However, it is noteworthy that when we interrogated these hippocampal clusters within the two age groups separately, we found that this effect was driven by significant or trend-level differentiation in the child group (anterior: $t(26)=-2.27$, $p=.03$, d=-0.44, 95% CI=[-0.076,-0.004]; body: $t(26)=-1.99$, $p=.06$, d=-0.38, 95% CI=[-0.080,0.001]; average across clusters: $t(26)=-2.90$, $p=.01$, d=-0.56, 95% CI=[-0.068,-0.012]), with no evidence of differentiation just among adults (ps>.36) (Figure 2c; averaged across clusters). This pattern hints that while differentiation is the primary scheme for memory organization in children, it is not invoked in the adult group in the present task.

Integration is more pronounced in adults. We next asked whether integration was more pronounced in the adult mPFC or elsewhere. Because different representational schemes have been previously observed in different subregions of mPFC[6], instead of using an anatomical ROI, here we chose a searchlight approach. In particulate, we conducted a whole-brain searchlight, asking whether mPFC or any other regions exhibited integration. As anticipated, the searchlight revealed that adults showed greater integration in mPFC (Figure 2d; two-sample t-test:

adult>child; *x*, *y*, *z* = -15.30, 44.80, -4.43), as well as in superior parietal cortex and posterior cingulate (Supplementary Table 2; Supplementary Figure 2c). Within the regions showing integration in adults, children showed either no reliable/significant representational scheme (mPFC) or differentiation (parietal; posterior cingulate; Supplementary Figures 3c). In addition to revealing a general lack of evidence for integration in children (except for caudate; Supplementary Table 2 & Supplementary Figures 2b & 3b), this whole-brain searchlight further revealed that children (but not adults) showed differentiation in a number of other regions (Supplementary Table 2), with several of these clusters showing developmental differences, such that differentiation was enhanced in children relative to adults (Supplementary Figures 2 & 3). Together, these findings suggest that the neural substrates implicated in memory organization in adulthood may serve a different representational function earlier in life.

Children and adults reinstate specific event features in neocortex. Having established that both children and adults systematically organized overlapping memories, with children relying on differentiation in hippocampus and neocortex and adults showing integration in neocortical regions including mPFC, our second goal was to test whether such discrepant schemes tracked neocortical reinstatement of the associated memory features. To address this question, we measured reinstatement at two levels of specificity: category (people vs. places) and item (Pinocchio vs. Peter Pan). We focused on functionally defined regions of interest (fROIs) within anatomical VTC and parietal cortex (Figure 3a; Methods), guided by prior work[37]. Because we were only interested in effects wherein reinstatement was significantly greater than zero, all tests measuring above baseline reinstatement were one-tailed.

To quantify the degree to which the target category was reinstated during the delay period, we trained a multi-voxel pattern analysis (MVPA) classifier[47] to identify patterns of activity associated with face and scene processing within each fROI (Methods; Supplementary Data: Classifier validation accuracy; for consideration of an RSA-based category measure instead, see Supplementary Figure 4). We then applied the trained classifier to patterns of fMRI activation measured during the 9-second delay interval (Figure 3b), to quantify the amount of face and scene evidence (Methods). Both age groups showed categorical reinstatement in VTC (adult: $t(24)=15.94$, $p<.001$, $d=3.19$, 95% CI=[1.01,1.30]; child: $t(26)=11.24$, $p<.001$, $d=2.16$, 95% CI=[0.76,1.09]; one-tailed) which further differed by age ($t(50)=2.08$, $p=.04$, $d=0.58$, 95% CI=[0.008,0.452]; two-tailed), such that adults showed higher-fidelity reinstatement (Figure 3c). Adults and children additionally showed reinstatement in parietal cortex (adult: $t(24)=5.82$, $p<.001$, $d=1.17$, 95% CI=[0.35,0.73]; child: $t(26)=1.95$, $p=.03$, $d=0.38$, 95% CI=[-0.01,0.53]; one-

tailed; Figure 3c), which did not differ by age ($t(50)=1.71$, $p=.09$, $d=0.48$, 95% CI=[-0.05,0.61]; two-tailed).

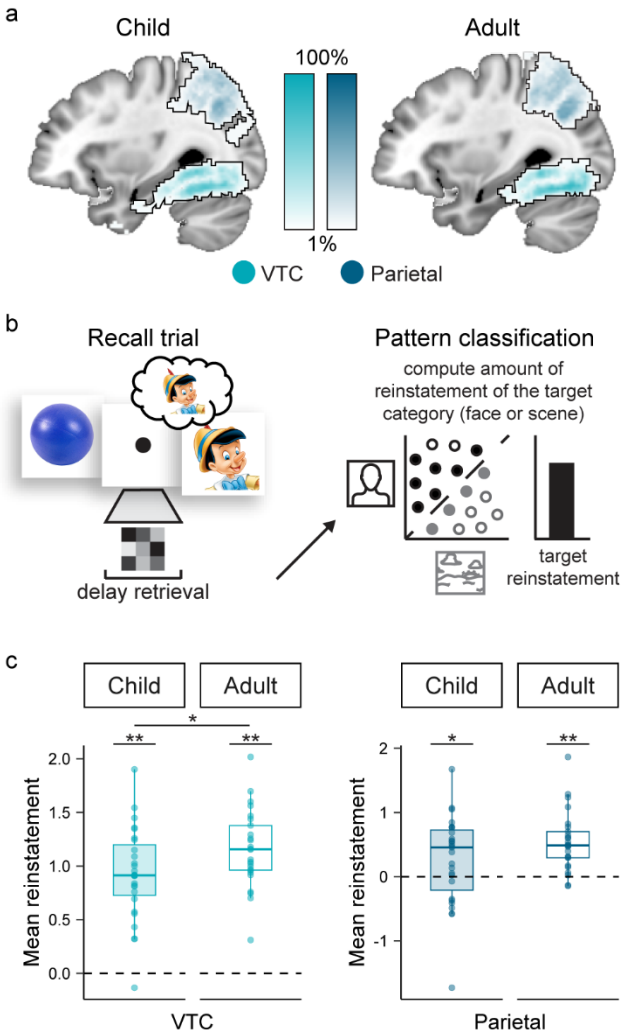

Figure 3. **Children and adults reinstate the target category. a** Voxels included within the VTC and parietal functional regions of interest (fROIs), separately for each age group. For each participant, we tested for reinstatement in the voxels that they engaged most during face and scene perception in a separate localizer task, within VTC and parietal cortex separately. Maps show the percentage of participants who had that same voxel in their fROI, plotted as a color-coded gradient. **b** Category decoding approach. We trained a multivoxel pattern analysis classifier to identify patterns of activity associated with face and scene processing in the VTC and parietal fROIs separately. For each delay retrieval pattern—which was modeled as a single voxelwise beta image across the 9-second retrieval period—we then applied the trained pattern-classification algorithm to estimate the amount of evidence for the target category (i.e., face or scene) reinstated. For each participant, we computed a single target reinstatement index by averaging the target classifier evidence (continuous probabilities) across trials, within each fROI separately. To correct for non-normality, the raw classifier output was transformed to logits. **c** VTC (left) and parietal cortex (right) category reinstatement. A reinstatement index reliably above 0 (i.e., the black dashed line) indicates significant target category evidence. Box plots depict the median (middle line), 25th and 75th percentiles (box), and the largest individual values no greater than the 5th and 95th percentiles (whiskers). Dots reflect individual participant means; n = 52 per plot (27 children; 25 adults), with dots extending beyond the whiskers reflecting

outliers, defined as values that were 1.5 times greater than the interquartile range (IQR). Asterisks above each box reflect a significant within-group effect at a one-tailed threshold of $p < .05$ (*) or $p < .001$ (**), while asterisks between bars reflect a significant between-group effect at a two-tailed threshold of $p < .05$.

We next examined item-level reinstatement. Specifically, we used RSA to quantify the similarity between patterns of fMRI activation measured during retrieval of a target item and viewing of the same item during a separate pre-exposure task (same-item similarity; Figure 4a), as compared to viewing other items from the same visual category (different-item similarity; Methods – general linear model [GLM] approach[48-49]). Item-level reinstatement in VTC was significant in adults ($t(24)=2.32$, $p=.01$, d=0.47, 95% CI=[0.001,0.011]; one-tailed) but only a trend in children ($t(26)=1.52$, $p=.07$, d=0.29, 95% CI=[-0.001,0.009]; one-tailed), and was not significant at either age in parietal cortex (*t*s<1.13; *p*s>.14; one-tailed). Neither VTC nor parietal item reinstatement differed as a function of age (*t*s<0.52; *p*s>.61). Given only modest item reinstatement effects in the category-selective fROIs (being observed only in VTC and only for adults), we additionally conducted a whole-brain item-level reinstatement searchlight analysis, which was small-volume corrected within the VTC and parietal fROIs as well as within hippocampus given prior adult work[39]. For both groups, we found significant item reinstatement in more focal regions within the fROIs as well as elsewhere in the brain (Supplementary Table 3; Supplementary Figure 5), though developmental differences were observed in hippocampus, with the locus of item reinstatement shifting from posterior to anterior between childhood and adulthood (Supplementary Figure 6).

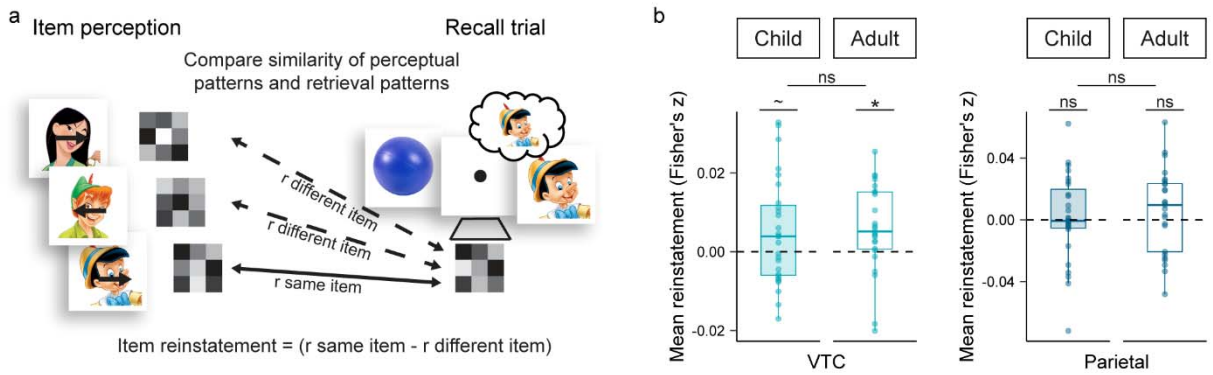

Figure 4. **Children and adults reinstate target items. a** Representational similarity analysis approach. Representational similarity analysis was implemented at the level of the full category-selective functional regions of interest (fROIS; from Figure 3a) as well as at the individual voxel level via separate searchlight analyses (see Supplementary Table 3 & Supplementary Figures 5 & 6). During item perception/the pre-exposure task, participants indicated whether an arrow superimposed on each stimulus (face/place) was pointing left or right. For each scanned recall trial, we correlated fMRI activation patterns from the delay period with viewing of the same item during a pre-exposure item perception localizer (same-item; solid arrow), which was compared to viewing of different items from the same visual category (different-item baseline; dashed arrow). For each participant, we computed a single item reinstatement index by averaging the mean same item – different item similarity difference across recall trials. **b** Pattern similarity results within VTC (left) and parietal cortex (right) fROIs. A mean reinstatement index reliably above 0 indicates significant item evidence, reflecting greater same-item similarity as compared to the different-item baseline. Box plots depict the median (middle line), 25th and 75th percentiles (box), and the largest individual values no greater than the 5th and 95th percentiles (whiskers). Dots reflect individual participant means; n = 52 per plot (27 children; 25 adults), with dots extending beyond the whiskers reflecting outliers, defined as values that were 1.5 times greater than the interquartile range (IQR). Symbols above each box reflect a significant within-group effect at a one-tailed threshold of $p < .05$ (*), or that is at trend levels (~). The letters “ns” above and between bars indicate a nonsignificant within-group or between-group effect, respectively.

We also wanted to confirm that modest evidence of item reinstatement in the fROIs was not an artifact of the long delay, particularly in children. To address this possibility, we performed additional analyses in which we separately considered reinstatement during the first and second halves of the delay period in both age groups (Methods). There were two possible ways in which reinstatement might differ by half. The first possibility was a decrease in reinstatement across the delay, suggesting that the delay may be too long for children to hold the reinstated memory in mind. The second possibility was that we would instead see an increase in reinstatement, which would support the idea that it took children time to reinstate the memory (as we anticipated[50]; Methods). Our results of this supplementary analysis were generally more consistent with the latter of the two possibilities, showing that both children and adults exhibited evidence of reinstatement across the full 9-second delay, with category (and some item) reinstatement evidence being stronger in the second half of the delay immediately prior to the decision probe (Supplementary Figure 7).

High-fidelity neocortical reinstatement is linked to mnemonic decision making. One important clue as to whether neocortical reinstatement directly guides memory behavior comes from asking how it relates to decision behavior on a trial-by-trial basis, in which participants were asked to judge whether the retrieved memory matched the decision probe (Figure 1b). To test this, we conducted mixed effects and drift diffusion models to examine whether variability in item (Figure 4b, same-item similarity) and category (Figure 3c) reinstatement during the delay predicted the probability of making a correct and/or faster response within a participant on a given trial. Because our mixed effects model focused on the reinstatement signatures that associated with correct versus incorrect performance, here we did not restrict our analyses to correct trials. We found that more strongly reinstated memories were associated with more accurate responses (VTC category reinstatement: χ^2_1=4.12, *p*=.04, Figure 5a, left; trend for parietal cortex item reinstatement: χ^2_1=3.52, *p*=.06, Figure 5a, right), which were significant predictors of performance despite the fact that our model additionally accounted for the main effect of age (χ^2_1= 29.08, $p < .001$; model comparison: AIC_{base}=848.16, $AIC_{MainEffect}$=817.74, χ^2_1=32.42, *p*<.001; a more complex model that allowed reinstatement to interact with age did not significantly improve model fit, χ^2_1=0, *p*=1.00 [see Methods]), as well as the potential influence of univariate activation, including cue-period hippocampal BOLD activation (χ^2_1=3.47, *p*=.06) and delay-period VTC and parietal BOLD activation (χ^2_1<0.24, *ps*>.63). Follow-up analyses of each age group separately showed that this effect was found in children (VTC: χ^2_1=3.62, *p*=.06; Parietal: χ^2_1=3.42, *p*=.06) but not adults (ps>.56). Likewise, in adults (who reached ceiling levels of accuracy) but not children, faster responses were associated with more strongly reinstated

memories while controlling for hippocampal BOLD activation during the cue which was shown to be a significant predictor of accuracy above (Figure 5b, right; parietal item). Furthermore, in line with the observation that reinstatement increased over the delay, reinstatement at the end (but not beginning) of the delay was behaviorally significant, predicting response accuracy in children and speed in adults (Supplementary Figure 8).

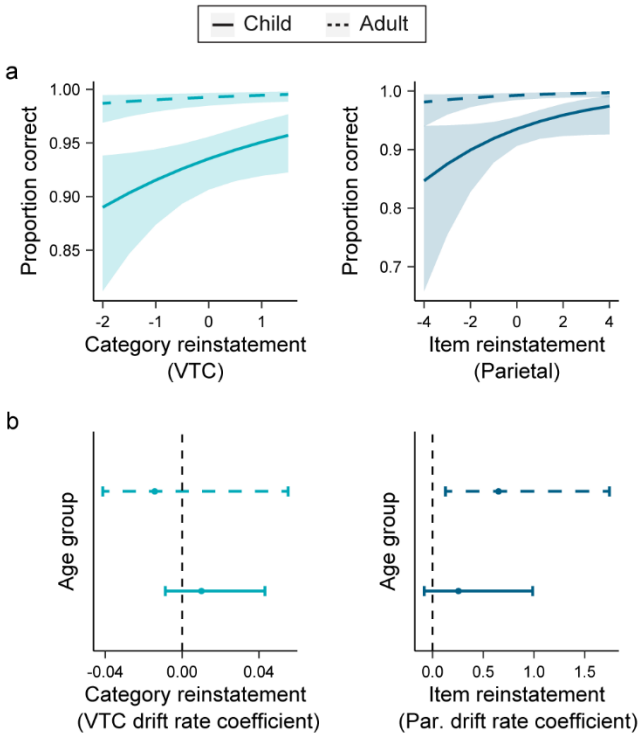

Figure 5. **Enhanced neocortical reinstatement in children and adults promotes superior memory behavior.** We tested the within-participant association between raw same-category and same-item reinstatement during the delay period with the probability of making either a correct **a** or speeded **b** match/mismatch behavioral response. **a** Results of mixed effects model examining decision accuracy. Here, significant predictors are depicted in separate plots for visualization purposes only, as all predictors (see Method) were included in a single statistical model. All predictors were scaled and centered within participants to remove participant-specific effects. For both plots, n = 52 (27 children; 25 adults); age group was modeled as a main effect. We found that greater category-level reinstatement in VTC (left) and a trend for greater item-level reinstatement in parietal cortex (right) each predicted successful performance. Lines (solid and dashed) reflect model predictions, while shaded areas reflect 95% confidence interval around the model predictions. **b** Results of drift diffusion models (DDM) examining decision speed. Here, the VTC category and parietal cortex item reinstatement indices were entered into a single DDM model of response speed, modeled as a linear effect on the drift rate parameter that reflects the rate of evidence accumulation (with two separate models run, one each for children and adults). Whereas higher drift rate coefficients reflect faster evidence accumulation and response time; lower values reflect slower evidence accumulation and response times. The interaction of drift rate and reinstatement in VTC (left) and parietal cortex (right) are depicted as forest plots. The lines represent 95% confidence intervals of the 2,000 posterior parameter estimates (one-tailed), with circles representing the mean of the posterior parameter distributions. Greater item reinstatement in parietal cortex predicted faster response times, but only in adults, which was reflected by drift rate coefficient confidence intervals not encompassing and greater than 0. No significant effects were observed for category reinstatement in VTC, in either adults or children.

Different representational schemes have benefits for reinstatement at different ages. We found that children and adults deploy high-fidelity neocortical representations (Figures 3 & 4) in service of mnemonic decisions (Figure 5). However, whereas children rely on hippocampal differentiation to organize related memories (Figure 2b), adults relied on integration, including in mPFC (Figure 2d) as anticipated. In a final set of analyses, we thus tested whether the degree of hippocampal differentiation and mPFC integration was associated with neocortical reinstatement in children and adults, respectively. In addition to our *a priori* hypothesis regarding mPFC integration, we further explored the relationship between neocortical reinstatement and other searchlight regions that revealed significant integration, either in adults (parietal cortex, posterior cingulate) or children (caudate). When quantifying delay-period reinstatement, we focused on only those neocortical fROIs that, in both age groups and across the full delay, either showed (1) reliable reinstatement (i.e., category reinstatement in VTC and parietal cortex; Figure 3c) or for which (2) trial-wise variability in reinstatement was linked to behavior (i.e., item reinstatement in parietal; Figure 5) at significant or trend levels.

Memory organization in hippocampus (anatomical) and mPFC (searchlight cluster) were significantly associated with category-level reinstatement in VTC for children and adults, respectively. Specifically, hippocampal differentiation was associated with reinstatement in children ($t(57.15)= -2.29$, $p=.03$) but not adults ($t(43.59)=0.43$, $p=.67$; relationship differed significantly between the two groups [Interaction: $\chi^2_1=3.78$, $p=.05$; Figure 6a]). Conversely, mPFC integration was associated with reinstatement in adults ($t(43.31)=2.36$, $p=.02$) but not children ($t(57.84)=-1.32$, $p=.19$; Interaction: $\chi^2_1=6.69$, $p=.01$; Figure 6b). In addition, mPFC integration was associated with item-level reinstatement in parietal cortex ($\chi^2_1=7.77$, $p=0.005$; best-fitting models were those that did not include age groups [i.e., base model; see Methods]). There were no significant associations between hippocampal organization and either item-level reinstatement in parietal cortex ($\chi^2_1=2.39$, $p=0.12$; base model) or category reinstatement in parietal cortex (hippocampus: $\chi^2_1=2.02$, $p=0.16$; mPFC: $\chi^2_1=0.93$, $p=0.33$; both base models).

In contrast to mPFC which revealed a significant association between memory integration and VTC category reinstatement among adults, there was no such association between cue organization and VTC category reinstatement in other regions that showed significant memory integration (Supplementary Figures 2 & 3), including in parietal cortex (Interaction: $\chi^2_1=3.09$, $p=.08$, Adult slope: $p=0.12$, Child slope: $p=0.35$), posterior cingulate (Interaction: $\chi^2_1=5.06$, $p=.02$, Adult slope: $p=0.07$, Child slope: $p=0.19$), or caudate (Base model: $\chi^2_1=0.37$, $p=.54$). Together, the findings suggest that different hippocampal and mPFC representational schemes

are uniquely and differentially beneficial for memory organization and reinstatement at different ages.

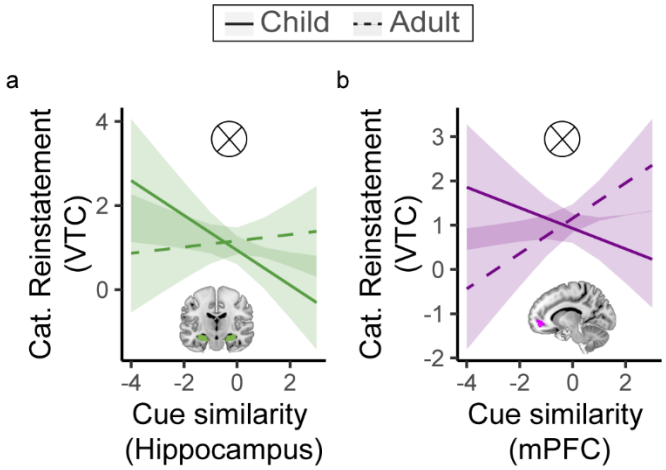

Figure 6. **The hippocampal and mPFC representations that governed how related memories were organized tracked the degree of ensuing neocortical reinstatement of the associated face and scene elements at the within-person level.** In both panels, for each participant and retrieval trial separately, we computed the averaged similarity of a retrieval cue object with related cue objects that shared the same face or scene associate (Figure 2a; related similarity). Thus, whereas lower cue similarity values reflect greater dissimilarity of related memories, higher values reflect greater similarity. We tested the association between that cue similarity value with reinstatement during the corresponding delay period for that same retrieval trial (Figure 3b; Target reinstatement example trial). **a** In children (solid green line) but not adults (dashed green line) there was a negative trial-wise association between cue similarity in anatomically-defined bilateral hippocampus (Figure 2b) and category reinstatement in VTC. In other words, lower similarity in hippocampus was associated with higher-fidelity reinstatement of the target category. **b** In adults (dashed purple line) but not children (solid purple line) there was a positive trial-wise association between cue similarity in the mPFC region defined from the searchlight (Figure 2d) and category reinstatement in VTC. That is, greater similarity in mPFC was associated with higher-fidelity reinstatement of the target category. For both panels (each comprising a different model), age group was modeled as an interaction and both interaction terms were significant at a threshold of $p \leq .05$ (marked here with a tensor product symbol). For each model, predictors (x-axis) were scaled and centered within participants to remove participant-specific effects. Lines (solid and dashed) reflect model predictions, while shaded areas reflect 95% interval around the model predictions.

Discussion

Prominent developmental theories have long proposed that age-related improvements in learning and cognition are rooted in the emergence of an increasingly complex representational system[51]. Yet despite both its theoretical importance[7-8] and educational relevance[14,40], the question of how children—for whom the neural substrates supporting memory are still developing[8-12,27,41-44]—represent related events to enable later memory retrieval remains unanswered. By leveraging fMRI with representational analysis methods, we show that children code related memories via hippocampal differentiation, wherein events that share overlapping features are represented as *less* similar to one another compared to unrelated events. In contrast, adults deploy an alternate representational scheme in which memories for related experiences are integrated, including within mPFC, wherein related events are more similar to one another than they are to unrelated events. Critically, these representational schemes were associated with memory reinstatement in children and adults, respectively: That is, hippocampal differentiation in children and mPFC integration in adults each tracked higher-fidelity neocortical reinstatement, which in turn was linked to superior memory behavior in both groups. Together, these findings indicate that children may be biased toward disambiguating overlapping experiences, while adults build complex representations that code common features across them. Such a shift in memory organization may confer certain behavioral advantages during each period in development. While children's memory representations may enable them to minimize interference among highly similar events[52], adults may have the advantage of relatively greater mnemonic flexibility[19,21-22].

Here, we demonstrate hippocampal differentiation of well-learned associations in children aged 7-10 years. This finding joins one prior study showing hippocampal differentiation in children of a similar age[34], bolstering the conclusion that this neural coding scheme is available by middle to late childhood. One notable extension of the present work is that here we measure not just hippocampal representations but also memory behavior, which enabled us to identify the neural codes available to be leveraged to make decisions. An interesting difference between our work and the previous study is that while the prior study reported similar levels of differentiation in children and adults[34], here we found reliably greater differentiation in children. One possible explanation for this apparent discrepancy might lie in our approach of ensuring participants formed strong associative memories (>90% accuracy): As the prior study showed children exhibited hippocampal differentiation only for previously experienced stimuli (i.e., familiar movies)[34], it may be that the multiple learning repetitions here encouraged robust differentiation in children. Such differentiation may have been further promoted by the

interleaving of our learning exposure with test[3]. By contrast in adults, aspects of our paradigm may have strongly pushed neither integration nor differentiation (discussed below), leading individual differences[23-25]—perhaps more prevalent among adults for whom multiple coding schemes are available[5-6]—to prevail and yield an apparent mix of hippocampal codes in this group. Importantly and in contrast to prior work, here we also measure representation at the endpoint of (rather than during[27]) learning, once memories for related events have stabilized. In doing so, our results reveal hippocampal differentiation—an active coding strategy that might help resolve interference among related memories[1-3]—as the unique means through which the still-developing brain achieves neocortical reinstatement and ultimately, successful behavior.

Beyond hippocampus, children in our study showed differentiation broadly across many neocortical sites. Such a tendency towards differentiation—or in some regions, an entire lack of systematic organization, perhaps akin to pattern separation, wherein non-overlapping traces arise through automatic orthogonalization[53]—contrasted with the representations we observed in adults, which were largely integrated. One interesting exception to this overwhelming bias in children toward differentiation was in caudate, which showed integration. Such an observation of integration underscores that children are not fundamentally incapable of this coding scheme, but rather that it may simply be less prevalent in their brains, primarily appearing in regions that mature earlier than neocortex[54]. Moreover, its isolation to caudate could indicate a role for action-outcome learning during the feedback phase of our test trials in children[54]. Specifically, providing corrective feedback after each behavioral choice (action) may have drawn children's attention to the face/scene "outcome" (which appeared in the center of the screen) that was shared across trials. Integration in caudate aside, our finding of overwhelming neocortical differentiation might suggest that children's neural biases can be characterized as shifting from differentiated (or pattern separated) to integrated over developmental time.

Such a shift toward integration is consistent with recent behavioral studies indicating pronounced developmental differences in reasoning behaviors that rely on extracting novel relations among memories. Although children are able to combine related memories to derive new factual knowledge[13] or make inferential reasoning judgments[12]—behaviors that may seem to imply and indeed (in adults[5,23-25]) benefit from neural integration—children may be accomplishing these same tasks using fundamentally different types of memories relative to adults[14,18,28]. For example, in contrast to adults, children often do not connect related facts or events during learning either spontaneously[18,28] or even when instructed to do so[27,55]. By contrast, similar instructions provided immediately before an inference test do facilitate performance, suggesting an instruction facilitates interrogation of separate memories during

flexible decision making[26]. Such behavioral data coupled with the differentiated neocortical organization observed here therefore points to the possibility that performance on flexible reasoning tasks may be achieved on the basis of different underlying neural memory representations at different ages[20]. Specifically, whereas adults may access an integrated representation directly to facilitate behavior, children may instead retrieve individual (differentiated or separated) representations and later recombine them when explicitly prompted to make an inference judgment[7,26-28]. As we did not directly assess participants' inferential reasoning or knowledge extraction behaviors, how such memory flexibility may be supported by different representational schemes across development remains to be seen. Moreover, whether such an organizing principle is specific to memory or rather would be observed across other domains (e.g., perception[31,56]) awaits further study.

In contrast to children, adults showed evidence for neural memory integration in several neocortical regions including mPFC. Importantly, such neural integration—wherein objects that shared the same associate, but were not directly experienced together, were represented more similarly in the brain than were unrelated objects—emerged despite no explicit demand to link related associations in memory. Indeed here, participants were instructed that they would need only to remember individual pairs, and in fact there was no behavioral requirement that participants make the novel connections (as in e.g., inferential reasoning) at all. While evidence of neural integration may seem surprising in this context, such a finding is consistent with both computational models of inference[57] and extant empirical data from adults[5-6]. For example, when trained to perform a task very similar to the one employed here, neural network models learn indirect connections between objects that share features and outcomes without reinforcement[58]. In other words, integration is an emergent property of these models and parallels neural organization in human mPFC. Likewise, in prior adult work, integration has been observed during/after overlapping pair learning but before any explicit demand to make links among memories in a behavioral inference task (or even awareness that such a test is forthcoming)[5-6,23-25]. Interpreting the present data in the context of this prior literature, we thus conclude that a mature mPFC not only represents directly experienced associations, but further encodes the inferred latent structure of a task or an individual's environment[59]. Certain other aspects of our study, such as the use of well-known stimuli, may have further encouraged mPFC integration as participants could incorporate the new associations into existing memory networks[60].

Adults did not show significant evidence of either integration or differentiation in hippocampus. While this null effect may on its surface may appear to conflict with prior adult studies[5-6], we suggest that it does not; rather, these findings can be reconciled by considering

the nuanced set of individual differences and situational factors[61] that promote integration or differentiation. With respect to individual differences, even when holding the learning scenario constant, people appear to vary in their representational tendencies[23-25]. Such differences may be related to many factors including brain structure[6,9,12], strategy use[62], working memory[14], and even culture[63]. Significant variability across participants would preclude observing reliable integration or differentiation across our group of adults on average. An additional and perhaps related possibility relates to the fact that the mature hippocampus is flexible and can form either differentiated or integrated representations, sometimes in parallel across different subregions[5], depending on the specifics of the learning situation. For example, one study[6] showed (anterior) hippocampal integration when initial associations are learned to a high level of accuracy before overlapping pairs are introduced, yet (both anterior and posterior) differentiation when overlapping associations were instead studied in an interleaved order. Notably, in that study parts of mPFC exhibited integration regardless of learning schedule, underscoring that it may complement more variable hippocampal codes while itself being comparatively less sensitive to task features. Thus, one possibility is that the interleaved training and more complex associative structure (i.e., three overlapping pairs as opposed to two) biased the adult hippocampus but not mPFC away from integration, while at the same time providing insufficient exposures to yield strong and consistent hippocampal differentiation[1-2].

Two other neocortical regions—parietal cortex and posterior cingulate—showed memory organization schemes that differed significantly across age group, with children showing differentiation and adults integration. In other words, the same voxels in these regions served diametrically opposite representational functions across development. Such findings generally align with past work. One possibility suggested by recent adult MRI data[58,64-65] is that both parietal cortex and posterior cingulate play privileged roles in forming and utilizing relational learning structures. For example, posterior cingulate in adults has been implicated in forming knowledge structures that reflect spatiotemporal regularities[64-65] while parietal cortex enables “shortcuts” through representational space to support rapid memory-guided decisions[58]. At the same time, others have speculated that parietal cortex is an episodic memory buffer that accumulates mnemonic evidence for decision-making[66-67]. In the present study, parietal cortex may index accumulation of the differentiated representations in children (from hippocampus) and the integrated representations in adults (from mPFC), prior to reflecting accumulation of the associated face/scene during the delay. However, note that here, in contrast to hippocampus and mPFC, parietal organization during the cue did not predict reinstatement of the associated face or scene during the ensuing delay, which might be expected under such an explanation.

Interestingly, very few studies have interrogated parietal or posterior cingulate development, and those that did tended to focus on the development of metamemory processes[42,68]. The present findings thus build on this work, suggesting that children and adults may rely on similar neocortical sites to activate a relevant memory space, the nature of which fundamentally differs as a function of age-related capacities in hippocampus and mPFC.

Our neural pattern analysis approach allowed us to detect memory reinstatement in both children and adults, which was moreover linked to behavior in both groups. Specifically, children and adults alike showed evidence of category- and item-level reinstatement of associative memory elements in neocortex. Notably, whereas adults showed broad evidence of item reinstatement in the ROI as a whole across the entire delay, item-level reinstatement in children was observed at significant or trend-levels (1) across the whole VTC ROI, but only during the second half of the delay (Supplementary Figure 7; trend) or (2) in a number of searchlight clusters, including hippocampus, across the entire delay (Supplementary Figures 5 & 6). Together these findings suggest that it is possible to quantify the internal contents of specific memories in children, provided the methods are sensitive to its relatively more circumscribed physical location and more delayed temporal profile. While prior work has shown that children and adults form similar neocortical representations of specific items at encoding[31], here we extend this finding to show comparable representations at memory retrieval. Moreover, the degree to which the associated faces and scenes were reinstated in neocortex, particularly during the second half of the delay window, had relevance for memory-based decision making. More specifically, greater category-level VTC and item-level parietal reinstatement predicted successful decisions in children, whereas greater item-level parietal reinstatement predicted faster decision speed in adults. As discussed above, these findings are generally consistent with the view that parietal cortex plays a privileged role in accumulating mnemonic evidence in service of decision-making[67].

Theories of memory development have long speculated that memory in childhood depends on the integrity of the hippocampal-neocortical system[10], and the strength with which distributed memory traces are bound together, in particular[69]. Therefore, having established the behavioral relevance of neocortical reinstatement, our final question was whether the hippocampal and mPFC representational schemes employed to organize overlapping memories tracked the fidelity with which those associated memory elements were reinstated in neocortex. Medial PFC representations related to item-level reinstatement in both groups, but this signature did not show developmental differences. Interestingly however, we did observe significant age differences in the organizational schemes predicting category-level reinstatement: Specifically,

trial-by-trial hippocampal differentiation among children (but not adults) and mPFC integration among adults (but not children) each tracked with category reinstatement in VTC. These associations were unique to hippocampus and mPFC, as they were not observed in other regions showing significant integration in either children or adults (i.e., parietal cortex, posterior cingulate, and caudate). By testing the relationship between representations in hippocampus and mPFC and neocortical reinstatement, the current study further establishes a role for hippocampus and mPFC in reinstating neocortical traces in development—a correspondence that has previously only been shown in adults[39,70]. Collectively, our results reveal that children and adults both engage in neocortical memory reinstatement, but through different representational pathways, underpinned by hippocampal and mPFC neural organization.

Taken together, the present research provides key insight into the development of neural representations associated with learning, representing, and retrieving overlapping events from memory. We show that memory for related events was associated with different underlying representational schemes at different ages. Specifically, children exhibited differentiation in the hippocampal system, which was linked to higher-fidelity neocortical reinstatement of event details, which was in turn associated with superior memory behavior. In contrast to children, hippocampal differentiation was unrelated to neocortical reinstatement in adults; instead, in adults mPFC integration benefitted neocortical reinstatement, which in turn tracked faster response speeds.. Our findings align well with current perspectives of neural development which suggest that differences in the maturational pace of the hippocampus and prefrontal cortex—and the neural representations they support—contributes to the shift in balance between memory specificity and flexibility[8]. Notably, the findings further suggest that children and adults form and retrieve memories using different types of representations. Importantly, given that the ability to organize, retrieve, and extend memories for related information is critical to knowledge development and predicts long-term academic success in both children and adults[18], the present results suggest that reducing overlap among related information may aid learning and memory in educationally relevant ways.

Methods

Participants

Eighty volunteers (*range* = 7.16-30.09 years) recruited from the greater Austin, TX area participated in the initial behavioral screening session. Of those, 17 individuals were excluded before the MRI scanning session due to voluntary withdrawal owing to discomfort in the mock MRI scanner (*N* = 2 children), scheduling conflicts (*N* = 2 children and 1 adult), neuropsychological assessments in the clinical range (*N* = 2 children and 4 adults), MRI contraindication(s) (*N* = 2 children and 1 adult), left-handedness (*N* = 1 child), or failure to identify the requisite number of familiar stimuli necessary for completion of the learning tasks (*N* = 2 children). All participants met our inclusion threshold for IQ and other screening criteria (see below for details regarding neuropsychological, cognitive, and basic participant assessments**)**. As such, 63 of the 80 participants also returned for the subsequent MRI scanning session.

The current report focuses on MRI data collected from 52 individuals: 25 adults aged 19-30 years (*M* = 22.32; *SD* = 3.29; 14 self-reported as female and 11 as male) and 27 children aged 7-10 years (*M* = 9.10, *SD* = 1.09; 15 females and 12 males, based on parental report). Eleven additional children participated in the MRI session but were excluded due to: (1) low behavioral performance (<90% final learning after five runs [*N* = 1] or failing to perform significantly above chance [50%] on the retrieval task [*N* = 3]), (2) excessive motion during the functional MRI scans (*N* = 1; see MRI image processing below), or (3) incomplete data (*N* = 6). Incomplete data was operationalized as failure to provide all three runs of the recall task or at least two out of three runs of the category localizer. Because a primary aim was to measure representation of overlapping memories—through comparing similarity of the three object cues that shared the same retrieval target, each of which was presented in a different run of the recall task—loss of a single recall run precluded full assessment of the similarity among all related items from a corresponding set. Moreover, to ensure that the trained MVPA classifier could accurately detect when participants were viewing faces and scenes in the localizer task (see below), at least two runs were needed, one of which was used to train and one of which was used to test the trained algorithm. There were no issues with data loss for the item localizer task, thus precluding any need for exclusion.

Our target sample size in each age group (*N* = 25) was guided by prior, related research in adults (Mack & Preston, 2016), which showed that this number was sufficient to detect reliable item- and category-level reinstatement in hippocampus and neocortex, as well as hippocampal-neocortical correspondences. Our combined child and adult sample size also aligns well with past developmental work employing multivariate methods similar to those used

here (Callaghan et al., 2021). An approximately equal number of children were recruited of each age within the 7-10 age range and careful efforts were made to achieve approximately equal numbers of males and females within each age. The final sample consisted of native English speakers who identified as 2% American Indian or Alaskan (*N* = 1 child), 15% Asian (*N* = 2 children and 6 adults), 4% Black or African American (*N* = 1 child and 1 adult), 73% White (*N* = 21 children and 17 adults), and 6% mixed racial descent (*N* = 2 children and 1 adult). Twenty-one percent of the sample was Hispanic (*N* = 6 children and 5 adults). All participants (child and adult) and parents/guardians (hereafter referred to as "parents") provided verbal assent at the start of the behavioral screening session as well as written informed consent at the start of the MRI scanning session (if they took part in that session). Participants received monetary compensation for their participation at each session, at a rate of $10/hour for the behavioral screening session and $25/hour for the MRI scanning session. There was also an opportunity to obtain up to a $10 bonus during the MRI scanning session, based on the number of MRI scanning runs completed. All procedures were approved by the University of Texas at Austin Institutional Review Board.

Materials and stimuli

Neuropsychological, cognitive, and basic participant assessments. Participants (or their parents) completed a series of measures to confirm that participants met screening criteria for the MRI study, which assessed: (1) basic participant information (self- or parent-reported), including right handedness, no color blindness, and whether English was a native language, (2) general intelligence in the normal range or higher (self-completed for children and adults), as indexed by the vocabulary and matrix reasoning subtests of the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II; Wechsler, 2011; inclusion threshold: <2 standard deviations below the mean), (3) psychiatric symptoms below the clinical range (self- or parental-reported), as indexed by the Symptom Checklist (SCL)-90-Revised (Derogatis, 1977) in adults (inclusion threshold: standardized global severity score <= 62) and the Child Behavior Checklist (CBCL; Achenbach, 1991) in children (inclusion threshold: standardized global severity score <= 63). Participants or their parents also provided information on socioeconomic status (adults and parents), life stress (adults, children, and their parents), and pubertal stage for children (via parental report). These measures were collected for exploratory purposes as part of a separate study (Coughlin, Ben-Asher, Roome, Varga, Moreau, Schneider, & Preston, 2022) and are thus not considered in the current report.

Stimuli. The goal of the present research was to measure patterns of neural activity engaged when separate yet related associative memories were *successfully* learned and

retrieved. We reasoned that successful retrieval of associative memories may vary as a function of stimulus familiarity. If so, neural evidence of reinstatement may be particularly robust for well-known stimuli. Thus, to equate the potential influence of familiarity across trials and age groups, a customized stimulus set consisting of highly familiar images was created for each participant (see Stimulus familiarity rating task below). Each stimulus set consisted of color images of six faces (three females; three males), six scenes, and 36 objects. All stimuli were sized to 400x400 pixels. Face and scene stimuli were images of characters and places, respectively, obtained from child-friendly films and television shows. Objects were photographs of common, everyday items, not obtained from films or television shows. Each face and scene was randomly paired with three different objects, for a total of 36 pairs comprising 12 sets of related items (e.g., Set 1: $face_1$-$object_1$, $face_1$-$object_2$, $face_1$-$object_3$; Set 2: $face_2$-$object_4$, $face_2$-$object_5$, $face_2$-$object_6$; etc.). This structure allowed us to ask both how individual objects that were associated with the same face or scene (e.g., $object_1$ versus $object_2$) were represented with respect to one another, as well as how they could each give rise to reinstatement of a consistent neural pattern representing a particular face or scene (e.g., $face_1$). Indeed, our pair learning structure was adapted from other tasks, with the intention of giving participants experiences that could (and have) yielded either integrated or differentiated memories[5-6], so that we could ask whether there are developmental differences in the types of codes that emerge.

Procedure

The experiment consisted of two sessions: (1) an initial behavioral screening session where participant interest in and eligibility for the MRI scanning session were assessed and (2) an MRI scanning session where all the main behavioral and imaging data were acquired. The delay between sessions ranged from 1 to 168 days (*Mean* = 34.17; *SD* = 29.69; *Median* = 27.0). Notably, the delay never exceeded six months, thus ensuring that standardized assessments used to determine eligibility at the initial screening session remained valid indicators of neuropsychological and cognitive function over the course of one's participation in the study. Moreover, task instructions were always reviewed at the start of the second session, removing any memory burden, and by extension, the need to match delay intervals among participants.

Behavioral screening (Session 1). To gauge interest and comfort with the MRI protocol that would be implemented at Session 2, participants (and their parents, if minors) were guided through the procedures to anticipate during the MRI session, including the imaging and behavioral task protocols. Eligibility for the MRI session was also assessed through standardized and experiment-specific tasks. The screening session had four phases: mock

scanner practice, standardized assessments, stimulus familiarity rating task, and behavioral practice tasks.

Mock scanner practice. Following informed assent and consent/permission regarding the screening and scanning procedures, participants were exposed to the mock MRI scanner where they practiced lying supine in the bore for 1-2 minutes while audio recordings of scanner noises were played aloud.

Standardized assessments. If participants were comfortable in the mock scanner, they and/or their parents completed a series of measures to confirm that participants met screening criteria for the MRI (see Neuropsychological, cognitive, and basic participant assessments above).

Stimulus familiarity rating task. To control for the potential influence of familiarity on learning and reinstatement across trials, individuals, and age groups, a customized stimulus set was generated for each participant. Participants were shown a maximum of 326 images (67 faces, 99 scenes, and 160 objects), presented randomly one at a time on a computer screen. Participants were instructed to indicate *how familiar* each image was, by choosing one of three options: (1) not at all, (2) a little bit, or (3) very. For images judged as “very” familiar, participants were then asked to name or describe the image (name common objects; name characters; describe scenes), which allowed us to control for potential differences in verbally expressible prior knowledge. Participants had an unlimited amount of time to make a response and made their familiarity ratings and verbal responses aloud. The experimenter entered all responses into the computer, scoring the verbal responses online as either correct (1) or incorrect (0). Only images that were judged as highly familiar and then correctly labeled were eligible for inclusion in the final stimulus set. In addition, this and all other experimental tasks were designed and administered using custom scripts in MATLAB R2017a (MathWorks, Natick, MA), which enabled automatic termination of the task when the number of images needed for the primary behavioral tasks was obtained (i.e., eight faces [two for practice], eight scenes [two for practice], and 40 objects [four for practice]). Therefore, the majority of participants did not provide familiarity ratings for all 326 stimuli.

Although stimulus sets necessarily differed across participants, several additional steps were taken to further match the types of stimulus content included across stimulus sets. First, only three object images from a semantic category (e.g., foods, athletic equipment, musical instruments, etc.) could be included in the same stimulus set. For example, although multiple musical instruments were included in the full set of 160 objects, once three instruments were identified, all other instruments were manually skipped by the experimenter. This design criteria

served to ensure that all stimulus sets spanned a variety of semantic categories. Second, and like the object inclusion criteria, we ensured that each of the faces and scenes were drawn from distinctive contexts. That is, only one face or scene image from a particular film or television show could be included in the same stimulus set. Thus, if a participant correctly identified “Peter Pan”, all other faces (e.g., Tinker Bell) and scenes (e.g., Captain Hook’s pirate ship) from the same film context were removed from the set of possible images. In addition to matching stimulus variety across participants, these stimulus criteria also served to increase the distinctiveness of face, scene, and object stimuli, which was intended to reduce the impact of cross-association similarity according to prior knowledge of the stimuli themselves.

Behavioral practice tasks. Finally, participants practiced short versions of the four experimental tasks that would be performed during the MRI session (see below for details). Participants repeated each task until a performance criterion of 90% or above was reached, which was typically achieved on the first round. Although the practice tasks included participant-selected, well-known stimuli, they did not overlap with those used in the main MRI experiment. The purpose of the practice tasks at Session 1 was to ensure that participants could perform the tasks.

MRI scanning (Session 2). Participants/parents provided informed consent, and were then reminded of the task instructions and repeated the same behavioral practice tasks as those completed at Session 1. Importantly, because these same practice tasks were repeated before MRI scanning at Session 2, participants were not expected to remember the task instructions over the delay. Following the practice tasks, participants then completed the following four experimental phases: pre-exposure item localizer (arrow detection task; in the MRI), pair learning (initial learning/retrieval task; outside of the MRI), memory recall (match/mismatch decision task; in the MRI), and category localizer (repeat detection task; in the MRI).

Pre-exposure item localizer (arrow detection task). The fMRI data acquired during the pre-exposure phase was used to estimate patterns of neural activity associated with perception of the face and scene items that would later be retrieved during the scanned memory recall task (see below), thus enabling quantification of the extent to which these patterns were reinstated in memory. As such, participants were exposed to the six face and scene images that comprised the to-be-learned object-face and object-scene associative pairs, presented through a rapid event-related design. Critically, the pre-exposure phase occurred prior to pair learning and scanned recall, thus allowing for clean estimation of the neural patterns associated with each item, prior to forming an association with the object cues.

On each trial, a face or scene stimulus was presented for 1.5s with a black fixation dot superimposed on the center of the image. After a random delay (250-750 ms) from stimulus onset, the fixation dot changed to a left- or right-facing arrow. The arrow was presented for the remainder of the 1.5s stimulus presentation and for an additional 1s after stimulus offset. Participants were instructed to respond as quickly as possible to the arrow direction by pressing one of two buttons on an MRI-compatible button box. The concurrent arrow task was intended to ensure attention to the stimuli. Following the offset of the arrow, a black fixation dot was presented alone for 2.0, 3.5, or 5.0s, with each inter-trial interval (ITI) randomly jittered but occurring approximately equally often for an average ITI of 3.5s across the total run. Each face and scene was presented three times in a pseudo-random order, such that at least two intervening images were presented in between repetitions of the same face or scene image. The number and order of left and right arrows was randomized. Participants completed three consecutive runs, each lasting 4.8 minutes, amounting to a total pre-exposure phase of about 15-20 minutes when accounting for brief breaks and check-ins in between each run.

Task performance was defined as the proportion of arrows for which the arrow direction was correctly detected, averaged across runs. Performance was high in both children (*M = 0.98; SD = 0.03*; Range = 0.89-1.00) and adults (*M = 0.99; SD = 0.01*; Range = 0.97-1.00), indicating that participants of all ages attended to the images during the pre-exposure item localizer scans.

Pair learning (initial learning/retrieval task). The present study sought to characterize the neural representations that supported *successful* learning and retrieval of individual yet related events. As such, before the scanned memory recall task, participants first completed the pair learning task (Figure 1a) which was intended to encourage the formation of strong memories, as evidenced by successful retrieval of the target faces and scenes associated with each object on a three-alternative forced-choice memory test (Figure 1c). Participants performed several consecutive learning-retrieval blocks to a 90% accuracy criterion, which were conducted outside of the MRI scanner.

On each trial of the learning block, a face-object or scene-object pair was presented for 3.5s followed by a 0.5s fixation cross (Figure 1a; Pair learning). The face and scene images appeared on the left side of the screen while the object images appeared on the right side of the screen. Each of the 36 pairs was shown once within the learning block, in a random order. Participants were instructed to view the pairs and to create a story relating the two images together.

Following the learning portion, participants were tested on their memory for the 36 pairs. On each trial, an object cue appeared above three face or scene options, one of which was the

correct associate (Figure 1a; Self-paced test). Participants were instructed to choose the face or scene image that was originally paired with the object by pressing one of three buttons on the computer keyboard (self-paced). Distracter options were pseudo-randomly assigned, with the constraint that distracters were always other items from the learning set drawn from the same visual category as the target image. That is, if the correct answer was a face, both distracters were also faces such that the three choices were always either all faces or all scenes. This design criteria served to ensure that participants were able to retrieve the specific association, rather than the gist of the category. Following each choice, regardless of accuracy, a feedback screen appeared for 1.25s, which featured the object together with the correct face or scene. The feedback screen and subsequent test trial were separated by a 0.5s fixation screen. The pairs were tested in a random order and the correct answers were randomly distributed across the first, second, and third position options. Participants performed at least two runs of learning-retrieval blocks, with different randomized trial orders across each of the runs. If the test accuracy was greater than or equal to 90% on the second run, participants continued on to the scanned memory recall task. Otherwise, participants repeated the learning and retrieval blocks until the 90% criterion was reached (maximum number of repetitions = 5; see Figure 1c and Supplementary Data).

Memory recall (match/mismatch decision task). The fMRI data acquired during the memory recall task was used to test the primary questions regarding how related memories are represented in hippocampus, as well as the degree to which reinstatement of associated memory elements is evidenced in hippocampus and neocortex. As depicted in Figure 1b, on each trial, the object cue was presented for 1.5s followed by a 9s delay interval in which only a fixation dot appeared on the screen. Presenting the objects in isolation enabled us to use this initial cue period of the trial to assess how memories for the related pairs were represented (i.e., different objects that were paired with the same face or scene stimulus, relative to a different face or scene stimulus; Figure 2a).

During the delay period, participants were instructed to retrieve and hold the associated face or scene in memory in preparation for an upcoming memory decision. Because the delay period was separated from the object cue and only consisted of a fixation dot, this period of the trial enabled us to measure reinstatement of the associated face or scene. Related to this point, our reason for implementing a 9-second delay in this task was to ensure that children had sufficient time to reinstate a high-fidelity memory representation. Past developmental electrophysiological work has shown pronounced age-related differences in the time course of neural processing during associative memory retrieval (e.g., Bauer, Stevens, Jackson, & San

Souci, 2012; Rollins & Riggins, 2017). Moreover, children are less likely to show electrophysiological signatures of memory retrieval in recording windows as those typically observed in adults, but do show similar signatures late in recording windows (e.g., Hajcak & Dennis, 2009). As such, here we opted for a relatively long delay, so that even if children were slower to reinstate associative memories compared with adults, we would still be able to capture such neural reinstatement in both groups.

To ensure that participants successfully retrieved the associative memories in the MRI scanner, the recall task further included a delayed match to memory decision (Figure 1b). That is, following the delay, a memory probe was presented that was either the correct associate (i.e., had been paired with the object previously; a match trial) or an incorrect associate (i.e., had been paired with a different object previously; a mismatch trial). The memory probe appeared for 3.0s. Participants were instructed to judge the probe as a match or a mismatch by pressing one of two buttons on an MRI-compatible button box. As was the case during initial learning/retrieval, mismatch probes were always drawn from the same visual category as the target associate, necessitating retrieval of the specific item associate rather than gist category information. Finally, to maximize the efficiency of the slow event-related design and the ensuing neural pattern estimates (see analyses), null fixation events with durations of 4.5, 6.0, and 7.5s were randomly sampled from a uniform distribution and intermixed between each trial.

Notably, during piloting, we observed that children often confused the start and end of each trial. That is, without explicitly prompting for a decision during the probe presentation, children sometimes confused the memory cues (i.e., objects) and probes (i.e., faces/scenes), leading to match/mismatch responses to the object cues rather than the face/scene decision probes. We took careful steps to reduce this confusion, as identification of the cue and probe periods of the task were necessary to ensure that participants retrieved the target associates. First and foremost, the fixation cross that preceded the object cue was presented in green, which was described as a signal that a new trial and memory cue was about to “start”. Second, 1s before the 3s probe presentation elapsed, a red fixation dot was superimposed on the center of the face/scene probe image, signaling that the trial was about to “stop”. Participants were instructed to log a response on the button box when they saw the red dot, if they had not done so already for that memory probe.

Participants completed three runs of the scanned recall task, each consisting of 12 trials. Across the 12 trials, each of the six faces and scenes was a retrieval target once. Moreover, across the three runs, each of the 36 unique object-face/scene pairs was tested once. As such, each of the three related objects that shared the same face or scene associate was tested in a

different run. Trial order within a run was pseudo-randomized, such that trials from the same stimulus category and condition (e.g., face mismatch trials) never appeared back-to-back. Additionally, target pairs were pseudo-randomly assigned to the match/mismatch condition, with the constraints that (1) within a run, there were three trials of each of the four stimulus category/condition combinations (e.g., face mismatch) and (2) across runs, the same retrieval target (e.g., Pinocchio) was assigned to each match/mismatch condition approximately equally often. Each run lasted 4.10 minutes for a total scanned retrieval phase of about 15 minutes when accounting for brief breaks and check-ins in between runs.

Task performance was defined as the proportion of match/mismatches that were accurately detected across all runs. Above-chance performance was determined for each individual participant using a binomial test of significance (i.e., minimum number = 25/36 trials correct). All participants included in imaging analyses performed significantly above chance on this memory task, indicating that participants successfully retrieved the associated memories during fMRI scanning (see Participants section above for details on the number of participants excluded for failing to perform above chance). Missing responses during this match/mismatch decision task were rare, yet still occurred (*M*=0.75 trials across participants, *SD*=1.63). To equate the number of trials between analyses that focused on the relationship between neural reinstatement and behavioral performance, both in terms of (1) accuracy and (2) reaction time (see Figure 5 and Supplementary Figure 8), we opted to exclude missing responses. That is, although missing trials could have been counted as inaccurate responses, because there is no corresponding reaction time for such trials, this would have led to differences in the trial counts among complementary analyses. Thus, for consistency, all reports of match/mismatch accuracy excluded trials for which there was no response (including group means; see Figure 1c). Importantly, this did not impact the overall pattern of results nor participant exclusions.

Category localizer (Repeat detection task). At the end of the session, participants performed a blocked, 1-back repeat detection task that included faces, scenes, objects, and scrambled objects. The fMRI data acquired during this task were used to identify functional regions of interest (fROIs) that were sensitive to face and scene processing (see below; Region of interest definition) and to train our multivoxel pattern analysis (MVPA) classifier to decode viewing of different stimulus types, thereby enabling estimation of category-level reinstatement during scanned memory recall. Stimuli were drawn from the same set as those used in the Stimulus familiarity rating task (see above) but did not overlap with the images selected for the main experiment.

Within each run of the repeat detection task, participants viewed a total of 96 images, 24 from each stimulus category (e.g., faces). On each trial, a stimulus was presented for 1.5s followed by a 0.5s black fixation dot. Participants were instructed to indicate when a stimulus was identical to (i.e., a repeat of) the immediately preceding image. Stimuli were blocked by category, with six items (including 1 repeat trial) presented per block. Thus, each block lasted a total of 12s. There were four blocks of each stimulus type per run. Stimulus blocks appeared in a pseudo-random order, with the constraint that a set consisting of one of each of the four categories (face, scene, object, scrambled object) occurred before repeating a new set of four category blocks. To ensure that the same stimulus block never appeared consecutively between sets, and to offer participants opportunities to rest periodically within the run, baseline blocks—in which participants viewed a black fixation dot for 12s—were inserted between each set of four stimulus blocks. Participants completed up to three runs of the repeat detection task, with unique stimuli presented in each run. For some participants, individual localizer runs were not completed due to time constraints. Specifically, 12% of participants (*N* = 6 children) contributed only two of the three possible localizer runs. Each run lasted approximately 4.4 minutes for a total category localizer task of 9-14 minutes, depending on the number of runs completed. Task performance was defined as the proportion of repeats that were correctly detected, averaged across runs. Performance was high in children (*M = 0.94; SD = 0.07*; Range = 0.69-1.00) and adults (*M = 0.98; SD = 0.03*; Range = 0.90-1.00), mirroring previous work (Schlichting et al., 2021). High behavioral performance, together with evidence of above-chance MVPA classification accuracy in all participants which did not differ between age groups (Supplementary Data), indicates that participants of all ages attended to the images.

MR image acquisition

Whole-brain imaging data were acquired on a 3.0T Siemens Skyra system. A high-resolution T1-weighted MPRAGE structural volume (TR = 1.9 s, TE = 2.43 ms, flip angle = 9 degrees, FOV = 256 mm, matrix = 256 x 256, voxel dimensions = 1 mm isotropic) was acquired for co-registration and parcellation. Functional data were acquired using a T2*-weighted multiband accelerated EPI pulse sequence, oriented approximately 20 degrees off the AC-PC axis (TR = 1.5 s, TE = 30 ms, flip angle = 71 degrees, FOV = 220 mm, matrix = 110 x 110, slice thickness = 2 mm isotropic voxels, multiband acceleration factor = 3, GRAPPA factor = 2). At least three fieldmaps were acquired (TR = 647 ms, TE = 5.00/7.46 ms, flip angle = 5 degrees, matrix = 64 x 64, voxel dimensions = 1.7 x 1.7 x 2.0 mm) to correct for magnetic field distortions, which were collected before the first run of each functional MRI phase (i.e., item pre-exposure,

memory recall, and category localizer) as well as any time the participant exited the scanner for a break in between functional runs within a single functional MRI phase.

MR image preprocessing

Data were preprocessed and analyzed using FSL 5.0.9 (FMRIB's Software Library, http://www.fmrib.ox.ac.uk/fsl) and Advanced Normalization Tools 2.1.0 (ANTs) (Avants et al., 2011) using a custom preprocessing pipeline (as in Morton, Zippi, Noh, & Preston, 2021). The T1 structural scan for each participant was corrected for bias field using N4BiasFieldCorrection. Freesurfer 6.0.0 was used to automatically segment cortical and subcortical areas based on the processed structural image (Desikan et al., 2006).

Functional scans were corrected for motion through alignment to the center volume using MCFLIRT, with spline interpolation. These realignment parameters were used to compute framewise displacement (FD) and DVARS for each volume. Individual volumes with both an FD and DVARS exceeding a threshold of 0.5 mm were identified. In addition to identifying these high-motion, 'bad' volumes, we also marked the volumes immediately before and two volumes immediately after each 'bad' volume. If more than 1/3 of the volumes were identified as 'bad', the scan was excluded from further analyses. As outlined above (see Participants), one child participant was excluded from analyses because more than 1/3 of the timepoints from a single run of the memory recall task exceeded our FD and DVARS threshold. All other participants contributed all three runs of the pre-exposure and memory recall tasks, and at least two runs of the category localizer task (see details above regarding time constraints for collection of the category localizer in some participants).

Functional scans were unwarped using an adapted version of epi_reg from FSL, which uses boundary-based registration (Greve & Fischl, 2009). This initial boundary-based registration was then followed by ANTs which applied intensity-based registration to refine any misalignments between functional scans and the T1 structural image. Nonbrain tissue was removed from the unwarped functional scans by projecting a brain mask derived from FreeSurfer into each participant's native functional space. Average brain-extracted unwarped functional scans were registered to a single reference scan collected in the middle of the scanning session (i.e., the second memory recall run) using ANTs. After calculating all transformations, motion correction, unwarping, and registration to the reference functional scan were conducted using B-spline interpolation, in two steps—a nonlinear step followed by a linear step—to minimize interpolation. The bias field for the average image was estimated for each scan using N4 bias correction implemented in ANTs and removed by dividing the timeseries by a single estimated bias field image. Functional time series were high-pass filtered (126 s

FWHM) and smoothed at 4 mm using FSL's SUSAN tool. As a result of this registration process, all functional and structural data were coregistered in each participant's native functional space, defined as the participant's middle functional run. All within-subject analyses were conducted in this native functional space, whereas group-level analyses were conducted in template MNI space.

Region of interest (ROI) definition

Anatomical ROIs. Our analyses focus on related hippocampal memory representation and neocortical reinstatement in children and adults. Thus, analyses were conducted in three a priori anatomical ROIs (aROIs) that have been reliably linked to related memory representation and/or reinstatement in past research with adults, including hippocampus (Mack & Preston, 2016; Molitor et al., 2021; Schlichting et al., 2015), ventral temporal cortex (VTC; Gordon, Rissman, Kiani, & Wagner, 2014; Trelle et al., 2020), and parietal cortex (Lee & Kuhl, 2016). The hippocampal mask was delineated manually on the 1mm MNI152 template based on a cytoarchitectonic atlas (Öngür et al., 2003; Price & Drevets, 2010). The bilateral hippocampal aROI was reverse normalized into each participant's native functional space using ANTs. Neocortical ROIs were also bilateral, but defined in participants' native functional space based on the automated Freesurfer parcellation. For the parietal mask, the angular gyrus (AG), supramarginal gyrus (SMG), superior parietal lobule (SPL), and intraparietal sulcus (IPS) were summed (Lee and Kuhl, 2016; for example). For the VTC mask, inferior temporal cortex, fusiform gyrus, and parahippocampal gyrus were summed (Trelle et al., 2020; for example).

Functional ROIs. An extensive body of data shows that both the spatial extent and degree of category-selective activation increases between childhood and adulthood (Golarai, Liberman, & Grill-Spector, 2007; Meissner et al., 2019; Peelen et al., 2007; Scherf, Behrmann, Humphreys, & Luna, 2007; 2011). Thus, to mitigate the possibility that age-related differences in either the spatial extent and/or the degree of category-selective cortical activation could influence our ability to measure age-related differences in the *representational fidelity* of face- and scene-selective reinstatement, we further restricted analysis of cortical reinstatement to a smaller set of 500 voxels that were most activated during face and scene perception during the separate category localizer task (as in Fandakova et al., 2019). In line with additional data showing that face- (Cohen et al., 2019; Golarai, Liberman, & Grill-Spector, 2015) and scene- (Fandakova et al., 2019; Golarai, Liberman, Yoon, Grill-Spector, 2010) representations are strikingly similar in children and adults when distributed activation patterns are examined within category-selective regions, we did not assume spatial contiguity between the most activated voxels. More specifically, for each participant, we conducted a general linear model (GLM) to

identify the 250 most active, distributed voxels separately for each stimulus type (i.e., 250 face and 250 scene voxels) within the anatomically-defined VTC and parietal cortex ROIs (see below; Estimation of condition-level face and scene activation). This resulted in two functional regions of interest (fROIs), each containing approximately 500 voxels that were tuned to face and scene processing for each participant (Figure 3a). Occasionally, the same individual voxel was identified as among both the most active face and scene voxels, resulting in nominally fewer than 500 total voxels (*VTC Range*: Adult = 344-448 and Child = 320-457; *Parietal Range*: Adult = 304-431 and Child = 288-410). Together, our approach served to ensure that developmental differences in the spatial extent, spatial location, and/or activation profile of category-selective voxels during perception could not account for any developmental differences observed in reinstatement within these regions.

General linear models (GLMs)

Estimation of condition-level face and scene activation (category localizer). To define fROIs for each participant and neocortical region, the category localizer data were interrogated for regions that showed preferential engagement during encoding of faces and scenes. For each run of the category localizer (i.e., repeat detection task), data were modeled using a GLM implemented in FEAT Version 6.00. Stimulus categories (faces, scenes, objects, scrambled objects, and fixations) were modeled as five individual regressors. That is, for each visual category, the four 12.0s-blocks (each containing six stimuli from the corresponding category) were combined into a single regressor and convolved with the canonical double-gamma hemodynamic response function (HRF). For all task regressors, temporal derivatives were included and temporal filtering was applied. Moreover, motion parameters calculated through MCFLIRT (see above) and their temporal derivatives were added as additional confound regressors. Because our goal was to identify voxels that showed increased engagement during face and scene events, we separately estimated the contrasts of face above implicit baseline as well as scene above implicit baseline, which yielded two whole-brain statistical images quantifying voxel-wise activation (one for faces and one for scenes), separately for each run and participant.

After modeling functional data within each run, the resulting statistical images were averaged across localizer runs for each participant using fixed effects (higher level analysis in FEAT). This yielded two final z-stat images per participant, with the statistical images reflecting how much each voxel was activated by face and scene processing for a given participant, across all runs. Because functional data were already co-registered across runs, no additional registration step was necessary. Moreover, as the fROIs were used to conduct additional

analyses only in each participant's native functional space, no averaging across participants or spatial normalization to a group template was necessary. The resulting face and scene whole-brain maps were then each masked with the VTC and parietal aROIs separately. As discussed above, within each anatomically-masked statistical image, the face and scene fROIs were defined as the 250 voxels showing the greatest amount of activation (i.e., those with the highest voxel-wise z-stats). We then combined the face- and scene-selective voxels within each individual region, yielding two fROIs (one VTC; one parietal cortex) for each participant, each containing approximately 500 voxels that were activated during face and scene processing.

Estimation of trial-level neural patterns evoked during the memory recall task. Three parallel sets of models were performed, one focused on the object cues and two focused on the delay period (one that modeled the full 9-s delay period as a single regressor of interest; the other that modeled the 9-second delay period as two separate regressors of interest: one corresponding to the first 4.5 s and the other to the second 4.5 s of the delay). For all models, separate event-specific univariate GLMs were conducted for each run, using the least squares single (LSS) method (Mumford et al., 2012; Xue et al., 2010). Within each run, the individual events were modeled at the trial-level for the phase of interest (i.e., cue or delay), with the other phases modeled at either the (1) trial level—if the phase occurred before the phase of interest, and thus was also considered an event of interest as it could influence trial-wise estimates of the primary event of interest (i.e., object cues that preceded the delay phase in the delay model) or the (2) condition level—if the phase occurred after the phase of interest, and thus could not be expected to influence trial-wise estimates of the event of interest (i.e., delay phase the followed the object in the cue model; match/mismatch probe in both the cue and delay models). For all trial-level GLMs (object cue, delay, pre-exposure [described below]), the GLMs included additional confound regressors to account for motion: six head motion parameters, their temporal derivatives, framewise displacement, and DVARS (Power et al., 2012). We also accounted for high-motion 'bad' volumes identified during pre-processing (see above; MR image processing), by including additional motion regressors for time points during which head motion exceeded both the FD and DVARS thresholds. All models were conducted on the spatially pre-processed fMRI data and temporal filtering was further applied to each event during modeling.

We used the cue model to interrogate the degree of similarity among patterns of fMRI activation evoked by related versus unrelated memories in hippocampus (Figure 2a), defined as object cues that were overlapping versus nonoverlapping, respectively (see above; Stimuli). To this end, we extracted neural patterns for each of the individual object cue presentations, each modeled as a single 1.5s regressor and convolved with the canonical double gamma HRF. As

noted above, nuisance regressors of no interest that accounted for later trial variance were also modeled, including the delay (by category: face and scene targets separately) and probe (category x probe type: face match, face mismatch, scene match, scene mismatch). The resulting beta images consisted of voxelwise parameter estimates for each of the 36 object cue activation patterns (12 per run; three associated with each face and scene stimulus) for each participant. These voxelwise estimates were used as inputs to the representational similarity analysis (RSA) assessing the similarity of related object cues relative to unrelated object cues from the same stimulus category (see below; Overlapping memory representation).

We used the delay models to quantify the degree of ensuing neocortical item (Figure 4a) and category (Figure 3b) reinstatement following the object cue, conducting two separate models that either (1) collapsed across the full 9s-delay period or (2) split the delay period into halves (4.5s). To this end, we extracted neural patterns for each of the individual delay periods—during which only a fixation dot was presented—each modeled as either a single 9s regressor (Figures 3 & 4) or as two 4.5s regressors (Supplementary Figure 7) and convolved with the canonical double gamma HRF. For both models, additional regressors that accounted for the preceding trial-wise variance were also modeled, namely, the object cues, as were regressors that accounted for later trial variance during the probe period (category x probe type: face match, face mismatch, scene match, scene mismatch). This approach thus resulted in one voxelwise beta image for each of the retrieval-related delay patterns (36 for the main 9s model; 72 for the supplementary split-half model) for each participant which were used as inputs to the MVPA and RSA reinstatement analyses (see below; Category reinstatement and Item reinstatement, respectively).

Estimation of item-specific neural patterns evoked during perception (pre-exposure item localizer). Finally, we wanted to quantify the degree of item-specific reinstatement during the delay period of the memory recall task—that is, neural evidence for reinstatement of a particular face or scene, despite simultaneous presentation of only a fixation dot during the delay period (Figure 4a). For this purpose, we estimated neural patterns associated with the perception of each of the six faces and scenes from the pre-exposure item localizer phase (i.e., arrow detection task), prior to any face-object or scene-object associative learning taking place. Separate LSS GLMs were conducted for each of the scanned pre-exposure runs, with each of the six faces and scenes modeled as a single event. That is, each of the three 1.5s stimulus presentations of an individual face/scene stimulus from a run were collapsed into a single regressor and convolved with the canonical double gamma HRF. The motion nuisance regressors mirrored those included for the object cue and delay period models.

This approach thus resulted in one voxelwise beta image for each of the six face and scene images for each run and participant (as in Mack & Preston, 2016), which were used as inputs to the RSA analyses (see below; Item reinstatement).

Multivariate pattern classification (MVPA) analysis

The goal of the classification analysis was to quantify the degree to which category-level information (face vs. scene evidence) was reinstated in the VTC and parietal fROIs during the memory recall task (Figure 3b). Pattern classification analyses were conducted on detrended and z-scored data at the within-participant level, in each participant's native functional space separately within each participant's functionally defined VTC and parietal fROIS (Figure 3a) using PyMVPA (Hanke et al., 2009) and custom Python scripts.

Notably, an RSA approach could also have been used to quantify category-level reinstatement (see Representational Similarity Analysis below; Supplementary Figure 4). Nevertheless, we opted to use an MVPA-based decoding measure over an RSA approach in our main analyses (Figures 3, 5, & 6) because our primary goal was to assess whether reinstatement at the item and/or category levels predicted behavior. To that end, we included both item and category reinstatement scores in the same statistical models (see below: Trial-level mixed effects models). Because the item and category measures calculated from the RSA approach were derived from the same pre-exposure scans—and were thus necessarily non-independent—the MVPA-based decoding measure derived from an independently scanned category localizer task allowed us to avoid potential issues with multicollinearity. Indeed, when both item- and category-level reinstatement were defined using RSA, we found the scores to be highly correlated with one another on a trial-by-trial basis (VTC: $\chi^2(1)=807.84$, $p<.001$; Parietal: $\chi^2(1)=181.57$, $p<.001$). Using the MVPA-based category reinstatement score allowed us to mitigate this issue.

Category decoding during perception. We first ensured that the MVPA classifier could accurately detect when participants were viewing faces and scenes. To this end, a sparse multinomial logistic regression (SMLR) classifier (lambda = 0.1, the package default) was trained to discriminate neural activation patterns associated with viewing faces and scenes in the category localizer task. The classification training was conducted on the preprocessed, spatially smoothed functional timeseries data. Each of the individual measurement volumes corresponding to a face or scene block (eight volumes per block: 12s of measurements per block/1.5s per measurement volume) was labeled according to the visually presented stimulus category. The fMRI data was shifted by 4.5s to account for the hemodynamic lag and these volume labels were used as inputs to the classifier algorithm for training.

Classification performance was first evaluated with a leave-one-run-out cross-validation approach, such that unlabeled fMRI patterns from one individual run (i.e., one “fold”) were labeled as a face or scene, according to the trained classifier for that participant. In other words, for each unlabeled neural activation pattern corresponding to an individual face or scene volume, the trained classifier labeled the volume as either a face or scene pattern—a prediction that was based on the correspondence between the unknown neural activation pattern and the classifier’s trained algorithm for face and scene activation patterns from the other runs. This validation approach was repeated for each run separately, resulting in two or three cross-validation folds, depending on the number of runs available. Classifier performance was then assessed by averaging the accuracy across the predictions for all cross-validation folds, yielding a single accuracy score for each participant which was compared to chance (50% for two categories) via one-sample t-tests (see Supplementary Data).

Category reinstatement. After validating that the classifier could accurately decode patterns of face and scene activation, we then applied the classifier to the memory recall task to quantify the degree to which face and scene patterns were reinstated in memory during the delay period (Figure 3b). To this end, the same classifier as above was used, except rather than the leave-one-run-out approach, here the classifier was trained on *all* runs of the localizer phase for each fROI. To predict the degree of face and scene information evidenced during the delay period of each scanned recall trial, the trained classifier was applied to the trial-level delay patterns modeled through the GLM approach (see above; estimation of trial-level neural patterns evoked during the memory recall task). More specifically, for each 9s or 4.5s retrieval pattern (corresponding to the full delay model or split-half delay model, respectively; see above), the pattern-classification algorithm estimated the relative amount of face and scene activation evidenced out of a probability of 1, yielding two continuous probabilities for each trial (two numbers for each full 9s delay trial, two numbers for the first 4.5s of the split half, and two numbers for the second 4.5s of the half). Of the two continuous probabilities for each trial, one number corresponded to face evidence; one number to scene evidence; both summing to 1 and each ranging from 0 (no evidence for a category) to 1 (perfect evidence for a category). Chance-level evidence for a category corresponded to a raw value of 0.5, reflecting equivalent evidence for both categories. To correct for non-normality, the classifier output was transformed to logits (log odds = log[x/(1-x)]) (as in Richter, Chanales, & Kuhl, 2016; Trelle et al., 2020). A logit value of 0 corresponded to a raw, chance-level reinstatement value of 0.5. Larger positive values indicated greater evidence for the target category relative to the nontarget category;

whereas, larger negative logit values indicated greater evidence for the nontarget relative to the target category.

Group-level analyses. Reliable reinstatement of the target category was evaluated within and between age groups by comparing across-participant mean target reinstatement probability values against chance levels (one-sample t-tests; chance = 0; one-tailed) and to each other (independent samples t-test; two-tailed). Importantly, we were interested in what *successful* memory reinstatement looks like in the developing brain and whether the neural underpinnings of such reinstatement change with age. As such, with the exception of two analyses in which we used participant-level item reinstatement to predict decision accuracy (Figure 5a) and response time (Figure 5b) on the match/mismatch task, here and for all other neural analyses, analyses were restricted to trials in which performance was successful at both unscanned final learning and scanned memory recall (Supplementary Table 1 for trial counts). As explained in more detail below (Statistical analyses: Performance-related trial restrictions), the performance restriction mitigated the possibility that age-related differences in overall performance could have accounted for age-related differences in the neural organizational schemes and/or neural reinstatement measures of interest.

Deriving trial-specific measures for within-participant analyses. In addition to group-level analyses, raw trial-wise probability estimates for the target retrieval category (i.e., face evidence when the target associate was a face) were also entered into within-participant behavioral mixed effect and drift diffusion models, enabling quantification of the association between target reinstatement and match/mismatch decision accuracy and speed, respectively (see below; Statistical analyses – Trial-level mixed effects models and Trial-level drift diffusion models, respectively).

Representational similarity analysis

Overlapping memory representation. To characterize how related object cues were represented in hippocampus relative to unrelated object cues, a representational similarity analysis approach (RSA) was implemented (see Figure 2a; Kriegeskorte et al., 2008). The RSA was implemented in PyMVPA (Hanke et al., 2009) within anatomically defined bilateral hippocampus in each participant's native functional space. Within this aROI, similarity matrices were generated by calculating the pairwise Pearson's correlation coefficients for the 36 statistics images corresponding to the object cues in the memory recall task (see above; Estimation of trial-level neural patterns evoked during the memory recall task), which were transformed to Fisher's *z*.

Group-level analyses. Differences in pattern similarity between objects that shared the same retrieval target (i.e., related objects) and objects that shared different retrieval targets from the same category (i.e., unrelated objects) were computed by averaging the similarity values corresponding to each relatedness condition, using custom scripts, operating in MATLAB 2017 (MathWorks, Natick MA). Importantly, only across-run similarity values were included to mitigate the influence of temporal autocorrelation on pattern similarity estimates (Mumford et al., 2012). The participant-level related and unrelated mean pattern similarity estimates were then entered into group analyses, by averaging across participants within a group.

To assess the effect of age and relatedness on hippocampal memory representations (Figure 2b), a 2 (child, adult) x 2 (related, unrelated) repeated measures analysis of variance (ANOVA) was conducted. Here, and for all significant age group x condition interactions, we next interrogated the effect of condition within each age group through separate repeated measures ANOVAs and/or paired sample t-tests.

Deriving trial-specific measures for within-participant analyses. In addition to group-level pattern similarity estimates which collapsed across all correct trials and across participants in each age group, trial-level similarity estimates were also computed for each individual participant. As for the group x condition-level analyses, for each object cue, the pattern similarity between that object and the two across-run objects that shared the same retrieval target was computed and averaged. This trial-level raw “related” neural estimate was entered into mixed effects models interrogating the association between within-participant variability in hippocampal representation and ensuing neocortical reinstatement (see below; Trial-level mixed effects models).

Notably, unlike the group-level analyses in which we tested whether differentiation and/or integration was reliably observed for related objects compared to a baseline unrelated object measurement, here we chose to focus on raw “related” neural estimates (i.e., similarity of object cues that shared the same face/scene associate) rather than baseline-subtracted estimates (i.e., related items minus unrelated items). The latter would have necessitated the use of a different baseline for each trial. Because varying baselines could differentially influence—and overwhelm—pattern similarity estimates for individual object cues/trials, in all mixed effects models we report raw related values reflecting the correlation between neural patterns objects that shared an overlapping retrieval target without reference to unrelated items.

Item reinstatement. To characterize the degree to which *item*-specific information was reinstated during memory recall (Figure 4a), RSA (Kriegeskorte et al., 2008) was again used to

compute how much more similar retrieval patterns were to perception of the same item (*same-item*) compared to perception of other items from the same category (*different-item*).

Group-level analyses. The RSA was implemented in PyMVPA (Hanke et al., 2009) within functionally defined VTC and parietal cortex in each participant's native functional space (see above; Functional ROIs). Within each fROI (see Figure 3a), we computed the similarity between the face and scene perception patterns derived from the pre-exposure item localizer and the retrieval patterns derived from the delay portion of the scanned memory recall task. Similarity estimates were generated by calculating pairwise perception/retrieval Pearson's correlation coefficients (see above; estimation of trial-level neural patterns evoked during perception; estimation of trial-level neural patterns evoked during the recall task).

Differences in pattern similarity between same-item similarity values (e.g., r of retrieval $_{face1}$ & perception $_{face1}$) and different-item similarity values from the same visual category (e.g., r of retrieval $_{face1}$ & perception $_{face2}$) were computed by transforming correlation values to Fisher's z scores and averaging the similarity values corresponding to the same- and different-item conditions, using custom MATLAB scripts. Because individual similarity values were always computed between scanned retrieval delay patterns and pre-exposure perceptual patterns, which necessarily occurred across runs, temporal autocorrelation was not a concern (Mumford et al., 2012). A mean item reinstatement index was derived for each participant and each fROI by subtracting the different-item similarity values (always restricted to the same category) from the same-item similarity values, restricted to correct retrieval and final pair learning trials. These participant-level estimates were averaged across participants within each group and compared to chance levels (i.e., 0; one-sample t-tests; one-tailed), yielding an estimate of whether item reinstatement was reliably evidenced on successful trials in each fROI (Figure 4b). To examine age-related differences in item reinstatement, the group-level item reinstatement indices were compared between children and adults (independent samples t-test; two-tailed), separately for each fROI (Figure 4b).

Deriving trial-specific measures for within-participant analyses. In addition to group-level pattern similarity estimates which collapsed across all correct trials and across participants within an age group, trial-level similarity estimates were also computed at the within-person level. Whereas the group-level analyses focused on whether children and adults showed reliable delay period reinstatement of the pre-exposure phase item (perceived) representations, here, the participant-level analyses examined whether trial-by-trial variability in such item reinstatement corresponded with match/mismatch behavior, as well as with similarity among overlapping object cues. As such, for each delay period, the pattern similarity between that

retrieval pattern and the three same-item perceptual patterns from the pre-exposure phase was computed and averaged. Trial-level “same-item” raw reinstatement indices were derived for each participant and fROI, for both correct and incorrect match/mismatch trials, and were directly entered into the within-subject mixed effect and drift diffusion models (though for the DDM, only match trials were included; see below). Such models tested whether perception/retrieval similarity for the same item—reflecting the degree to which the item was reinstated—corresponded with the likelihood and/or speed with which participants made the subsequent decision.

As in the trial-wise overlapping memory analyses described above, here we chose to focus on raw same-item similarity (i.e., similarity of the item with itself) rather than baseline-subtracted item similarity (i.e., item with itself minus item with same-category items). Thus, in all mixed effects models we report raw item reinstatement values reflecting the correlation between neural retrieval and perception of the same item without reference to other items from the same category.

Category reinstatement (RSA approach). In addition to estimating item reinstatement, the same RSA approach can be used to quantify category-level reinstatement. Although our trial-level models focused on the MVPA-based decoding measure to avoid potential issues with multicollinearity between our category (category localizer/MVPA) and item (item pre-exposure/RSA) measures (see above), we further feature group-level category RSA findings for comparison (Supplementary Figure 4). More specifically, we repeated the same group-level analyses as for item reinstatement (see above), with perception patterns derived from the same pre-exposure item localizer and retrieval patterns derived from the same delay portion of the scanned memory recall task. However, instead of item reinstatement scores, here we calculated category reinstatement scores, defined as differences in pattern similarity between same-category similarity values (e.g., two faces; r of retrieval $_{face1}$ & perception $_{face2}$) and different-category similarity values (e.g., a face & a scene; r of retrieval $_{face1}$ & perception $_{scene1}$).

Searchlight analyses.

In addition to the RSA analyses above which examined overlapping memory representation across the entire anatomical hippocampus and item reinstatement in functionally-defined categorical ROIs, two searchlight extensions were conducted to identify more localized voxels that exhibited (1) significant differentiation and/or integration of overlapping object cues across the entire brain, and in hippocampus in particular and (2) significant item-level reinstatement across the entire brain, and within category-selective fROIs and hippocampus in particular. As above, for both searchlight analyses, only similarity values corresponding to

correct final learning and match/mismatch trials were included in the similarity matrices, enabling identification of regions selectively recruited during *successful* overlapping memory representation and/or reinstatement. Moreover, all searchlights were run across the whole-brain grey matter masks within each participant's native functional space. Searchlight maps were then masked and statistically evaluated within each ROI separately. We detail the general searchlight approach in the overview below.

Overview. All searchlight analyses implemented the same pattern similarity approaches that were performed across the full ROIs above, using PyMVPA. Yet rather than calculating one similarity index for each ROI—that collapsed across all voxels in an ROI for a given participant—here, the similarity contrasts (i.e., overlapping representation or item reinstatement) were calculated at the level of individual searchlight spheres (radius = 3 voxels). A random permutation approach was used to evaluate the statistical significance of the observed index within each searchlight sphere, whereby the actual observed contrast—(1) overlapping representation: related minus unrelated object cues (for integration; or vice versa for differentiation) or (2) item reinstatement: same minus different item—was compared to a null distribution of indices calculated from randomly permuting the values that contributed to the observed contrast. For example, with respect to the item reinstatement searchlight, we randomly shuffled assignment of the same-item and different-item similarity values, resulting in a random item reinstatement index (same- minus different-item similarity). For each searchlight analysis, the permutation process was conducted 1,000 times within each searchlight sphere, yielding a null distribution of contrast indices (see below for more details about the different searchlight analyses). A z-score reflecting how much greater the actual observed index was compared to the mean of the null distribution values was computed, resulting in a whole-brain z-map for each searchlight analysis.

The above searchlight approach was conducted at the individual participant level, within each participant's native functional space. To then evaluate whether neighboring voxels showed statistically significant item reinstatement or related memory representation across individuals, we applied a cluster-based random permutation approach. To align individual voxels across individuals, we first normalized the participant z-maps to MNI group space using nonlinear SyN transformations in ANTs. Using Randomise in FSL (Winkler et al., 2014), we then assessed the mean similarity value across participants, wherein an actual observed value was calculated by averaging the observed z-stat index—reflecting related memory representation or item reinstatement—across participants for each individual voxel. To evaluate the statistical significance of the observed, group-level index, each across-participant voxelwise z-stat was

compared to a null distribution of 5,000 permuted samples via Randomise. The random permutation approach was conducted in one of two ways, depending on the type of group contrast. When evaluating whether a particular group—children, adults, or the full sample—exhibited reliable related memory representation or item reinstatement, a one-sample t-test approach was performed. In contrast, when evaluating group differences—children versus adults, or vice versa—an unpaired sample t-test approach was performed.

With respect to the one-sample approach, on each of the 5,000 permutations, the sign of participants' z-maps were randomly inverted within Randomise. That is, a subset of the participant z-maps were multiplied by -1 and a new, permuted across-participant index was calculated. As discussed in detail below, because of how the searchlight contrasts were set up, reliable reinstatement and related memory representation were always reflected by positive similarity values. For example, the observed item reinstatement contrast subtracted different-item similarity from same-item similarity, such that greater same-item similarity relative to different-item similarity would necessarily entail a positive item reinstatement value. As such, the 5,000 random inversions were used to compute a voxelwise probability map reflecting the proportion of null distribution sample values greater than or equal to the observed group-level index (i.e., *p* values). Put concretely, a voxel value of $p = .01$ would indicate that only 50 out of the 5,000 permuted estimates exceeded the actual observed group-level index.

With regard to the two-sample approach, rather than changing the sign of participants' z-maps, each voxelwise permuted sample was generated by randomly shuffling assignment of participants to the adult or child group within Randomise. Here, the 5,000 random permutations were used to compute a voxelwise probability map reflecting the proportion of null distribution sample values greater than or equal to the observed group difference, run separately for the adult > children contrast and children > adult contrast within Randomise. In this case, a voxel value of $p = .01$ would indicate that only 50 out of the 5,000 permuted estimates exceeded the actual observed group difference.

Finally, following computation of the voxel-wise statistical maps via Randomise, we then sought to identify significant clusters of spatially contiguous voxels (Woo et al., 2014), thereby controlling for the family-wise error rate across voxels. Here, we applied FSL's cluster function to the statistical output from Randomise, whereby neighboring voxels for which the p-value exceeded our $p < .01$ threshold (across participants) were assigned to a cluster and the voxel-size of each cluster was summed. We then used AFNI's 3dFWHMx method and the spatial AutoCorrelation function to conduct simulations to ask how often we would expect a cluster of a given voxel size based on chance alone. First, the smoothness of the original data on which the

native searchlights were run was estimated, as was how correlated a given voxel was with its neighbors (i.e., -acf method). Second, these acf estimates were fed into the AFNI 3dClustSim function (Cox, 1996), which models the minimum number of contiguous voxels needed for a cluster to be deemed statistically significant at a voxel threshold of $p < .01$ and a cluster threshold of $p < .05$. In particular, 3dClustSim generated 10,000 artificial simulations (the default) with the same dimensions and smoothness as the searchlight dataset, but composed of noise derived from the residuals of the GLMs used to model the searchlight data. The maximum cluster size was obtained for every random permutation. For the observed data, only clusters the size of which exceeded the 95th percentile value of the of the random cluster sizes were considered to have met the threshold for statistical significance (i.e., $p < .05$; two-sided, second-nearest neighbor). Third, and finally, we used FSL to mask the clusters (from FSL cluster; step one above) for which the cluster size reached this family-wise error threshold.

Searchlight for overlapping memory representation. A searchlight extension of the overlapping memory representation analysis conducted in anatomical hippocampus was conducted to identify localized voxels that exhibited significant differentiation and/or integration of related object cues, both at the whole-brain level and then small-volume corrected within hippocampus. To test the hypothesis that differentiated or integrated representations may be present in smaller and/or different regions of the brain, we conducted two searchlight analyses that looked for differentiation and integration of related memory representations, operationalized as either (1) lower or (2) higher pattern similarity for object cues associated with an overlapping stimulus (e.g., $face_1$) relative to object cues associated with different stimuli from the same visual category (e.g., $face_1$ versus $face_2$), respectively. In particular, similarity matrices were again generated by calculating the pairwise Pearson's correlation values between the object cue patterns for related versus unrelated items (constrained to the same stimulus category), transformed to Fisher's *z*. In the integration searchlight, the observed and randomly permuted differentiation indices were calculated by subtracting the unrelated similarity values from the related similarity values—in essence, searching for regions that showed related > unrelated within each searchlight sphere. In the differentiation searchlight, we performed the inverse contrast, such that the observed and randomly permuted integration indices were calculated by subtracting the related similarity values from the unrelated similarity values—thereby reflecting regions that showed unrelated > related pattern similarity within each sphere.

To evaluate the presence of age-invariant representational patterns, we tested for significant differentiation and integration regions collapsed across children and adults via a one-sample permutation t-tests (minimum full-sample cluster sizes: whole-brain = 85 voxels,

hippocampus = 9 voxels). We then assessed age-specific representational patterns in one of two ways. Specifically, we tested for regions that showed greater integration in adults relative to children as well as in children relative to adults, via two-sample unpaired permutation t-tests (same cluster thresholds as for one-sample t-tests collapsed across children and adults above). Importantly, an interaction of integration with age group is ambiguous with respect to whether significant integration and/or differentiation exists within each age group. As such, for regions that survived cluster correction, we separately tested for integration or differentiation within each age group, and report age interactions only when an effect was reliable within at least one group (i.e., enhanced integration was driven by significant integration in one group and/or by significant differentiation in the other group; see Supplementary Figure 3).

Searchlight for item reinstatement. A searchlight analysis extension of the main item reinstatement analysis (Figure 4b) was conducted to identify localized voxels that exhibited significant item-level reinstatement at the whole-brain level (Supplementary Table 3; Supplementary Figure 5), which was then small-volume corrected within the fROIs (VTC, parietal; Supplementary Table 3; Supplementary Figure 5) as well as within the anatomical hippocampal mask (Supplementary Figure 6). With respect to the fROIs, this complementary approach allowed us to ask whether item reinstatement—which was only reliably observed in adults in VTC (Figure 4b)—could be detected within relatively more selective regions nested within those engaged by stimulus category as a whole (i.e., the full fROIs; Figure 3a). Within these whole-brain searchlights, similarity matrices were again generated by calculating the pairwise Pearson's correlation values between the scanned retrieval delay patterns and pre-exposure patterns of viewing that same item versus different items (Figure 4a; constrained to the same stimulus category), transformed to Fisher's *z*. The observed and randomly permuted item reinstatement indices were calculated by subtracting the different-item similarity values from the same-item similarity values within each searchlight sphere.

To parallel the main fROI analyses above, we tested for significant item reinstatement regions within the child and adult groups separately. To that end, for each age group, a one-sample permutation t-test was performed, yielding cluster-corrected regions that reflected significant item-level reinstatement in children and adults. As discussed above, the group-level voxel-wise p-maps were masked with the within-group VTC and parietal fROIS, as well as within anatomical hippocampus. Because the cluster size that corresponded to a familywise alpha of 0.05 necessarily varied as a function of the smoothness and autocorrelation of the voxels that comprised the particular ROIs (see above), cluster thresholds were defined separately for each

age group and each ROI (minimum cluster sizes: whole-brain - Adult: 79, Child: 95; VTC – Adult: 20, Child: 23; Parietal - Adult: 32, Child: 29; hippocampus - Adult: 10, Child: 10).

Statistical analyses

Performance-related trial restrictions. Where possible, for all statistical analyses that evaluated neural measures of memory organization and/or reinstatement (mean-level group analyses, searchlights, and trial-level models that focused on the relationship between memory organization and reinstatement [Figure 6]), we have limited our analysis to "correct" trials to reduce the possibility that age-related differences in overall performance contaminate our neural results. Specifically, for the aforementioned analyses, we measure both overlapping neural memory organization and item/category reinstatement considering only trials for which the participant both (1) learned the pair initially, which we defined as having remembered it correctly on the final retrieval task from the pair learning phase and (2) correctly responded during the match/mismatch decision of the scanned memory recall task. The logic is that if they either did not learn the pair initially, or did not correctly identify the probe as a match or mismatch, there is a higher likelihood they were reinstating the wrong item (or nothing at all).

The performance-related restrictions were applied to all statistical analyses with the exception two in which we used participant-level category and item reinstatement to predict decision accuracy and speed on the match/mismatch task (Figure 5). In the case of accuracy (Figure 5a), because we wanted to predict accuracy on the match/mismatch task at the trial level, we needed to include both correct and incorrect trials as outcomes to run our logistic regression. Thus, here, the mixed effects analysis was limited according to neither performance on the unscanned final learning task, nor the actual scanned recall task. This approach of including all trials was necessary to ensure participants (who performed quite well given the criterion approach, and who showed a high trial-by-trial correspondence between final learning task and scanned task performance) had a sufficient number of incorrect trials.

In the case of response speed (Figure 5b), there were no trial restrictions based on behavioral accuracy, as the DDM approach requires inclusion of both correct and incorrect trials. However, we did restrict our analysis to only *match* trials to equate the difficulty of the decision-making process, which would be expected to vary across match and mismatch trial types, especially on mismatch trials which by their very nature require non-mnemonic processes (i.e., because participants cannot rely on the matching target memory that was reinstated).

Mean-level group analyses. Group analyses comparing conditional differences between age groups and/or within age groups were conducted via t-tests and ANOVAs in SPSS (Version 27), with Bonferroni-adjusted p-values reported to control for multiple comparisons.

While we provide these within-group paired t-tests, we do not additionally report follow-up comparisons within condition, across age groups (e.g., related condition in children versus adults). Our reason for omitting these comparisons is because absolute multivariate neural values are not particularly meaningful on their own, as many aspects of data quality and/or preprocessing can impact the overall correlation values one obtains and should instead be considered relative to some meaningful baseline to be interpretable. Therefore, we opt to focus on the difference between related and unrelated, and whether this difference differs within and between age groups. In cases where the assumption of sphericity was violated in ANOVAs, the degrees of freedom were adjusted using Greenhouse-Geisser corrections. Likewise, for independent samples t-tests, when the Levene's Test indicated unequal variances between age groups, the degrees of freedom were adjusted accordingly.

Notably, for category and item reinstatement (but not other effects), we evaluated the statistical significance within each group via a one-tailed t-test (e.g., Figures 3c & 4b). The rationale for using one-tailed t-tests here and not for other analyses (e.g., memory organization; Figure 2b) is that only for the reinstatement analysis was our prediction strongly directional. In other words, we were only interested in effects wherein reinstatement is significantly *greater than* (not merely different from) zero, and therefore 1-tailed tests are appropriate. By contrast, for memory organization analyses both directions of difference are interpretable (i.e., related < unrelated [less than 0] is taken as evidence of differentiation, whereas related > unrelated [greater than 0] implies integration). Therefore, throughout the paper all other statistical tests are two-tailed unless otherwise noted.

Trial-level mixed effects models. Trial-wise analyses were conducted using mixed effects models using the lme4 package (Bates, Machler, Bolker, & Walker, 2015) in RStudio (2020; version 1.2.5042). We first tested whether trial-wise variability in neocortical category and item reinstatement explained unique variance in successful performance on the scanned memory recall match/mismatch task. As noted above, this analysis did not control for behavioral performance (see Performance-related trial restrictions), as this was the dependent measure of interest. To this end, we quantified model fits for a model including age group, as compared to a base model that did not include age as a predictor of match/mismatch decision accuracy. Because age group explained significant variance in behavioral performance (as a main effect; modeling age as an interaction did not improve model fit above the main effect model, $p = 1.00$; see Results: High-fidelity memory reinstatement facilitates mnemonic decision making), we accounted for the main effect of age in the main model assessing the association between neocortical reinstatement and performance. More specifically, we conducted a logistic mixed

effect model (glmer function) using a binomial linking function (family = binomial) that included both item reinstatement (same-item similarity) and category reinstatement (logit transformed classifier evidence for target category) indices in the VTC and parietal cortex fROIs during the delay interval (i.e., four separate reinstatement predictors), as well as additional regressors that captured variability in univariate hippocampal activation during the preceding cue period, as well as univariate VTC and parietal activation during the delay period. We modeled the participant-specific association as a random slope. Moreover, individual model predictors were scaled and centered across all trials within a participant prior to analyses, thus standardizing the distribution between subjects. For this main effect model, statistical significance of each individual predictor was assessed using a Wald chi-square test (Fox & Weisberg, 2011).

Given evidence of significant hippocampal differentiation in children (Figure 2b) and integration in adults (Figure 2d) during the cue period, we additionally conducted trial-wise, within-participant mixed effects models to test whether the degree to which hippocampal/mPFC representations were differentiated/integrated was associated with neocortical reinstatement during the delay interval, restricted to correct final learning and scanned recall trials (i.e., the same performance-related restrictions applied elsewhere). Given the continuous nature of the dependent measure, here we employed linear mixed effects models (lmer function). More specifically, in models that included both children and adults, we assessed the trial-wise association between cue similarity (i.e., the raw pattern similarity for related cue items), separately for anatomical hippocampus and the mPFC searchlight cluster, and neocortical reinstatement, separately for parietal cortex item reinstatement (same-item similarity) and VTC and Parietal category reinstatement (logit transformed classifier evidence for target category). This resulted in six models: (1) hippocampus/VTC category; (2) hippocampus/parietal category; (3) hippocampus/parietal item; (4) mPFC/VTC category; (5) mPFC/parietal category; and (6) mPFC/parietal item. As above, individual model predictors were scaled and centered across all trials within participant prior to analyses. For models in which there is a significant (or trend-level) interaction term (see below), statistical significance of each individual predictor was assessed by getting separate slopes for each group (i.e., simple_slopes function from reghelper R package).

In both hippocampus and the mPFC, we found significant associations with category-level reinstatement in VTC, which were marginally better fit by models that included an interaction between cue organization and age (Hippocampus: AIC_{base}=10042, $AIC_{interaction}$=10041, χ^2_2=4.75, *p*=.09; mPFC: AIC_{base}=10043, $AIC_{interaction}$=10042, χ^2_3=7.52, *p*=.06) but not by a model that included age as a main effect (Hippocampus: *p*=.33; mPFC: *p*=.33). We

also found an association between mPFC integration and item-level reinstatement in parietal cortex which was best-fit by a base model that did not include age group as a factor (Main Effect and Interaction model fits: ps > 0.28).

To determine the specificity of these effects to hippocampus/mPFC, we additionally conducted exploratory analyses in which we ran the same models for other searchlight regions that showed significant integration in either children or adults (see Supplementary Table 2; Supplementary Figures 2 & 3). Because VTC category reinstatement was the only variable that tracked both hippocampal and mPFC organization (in contrast to parietal item reinstatement which only tracked mPFC organization), these follow-up analyses were restricted to category-level reinstatement in VTC, resulting in three separate models: (1) caudate/VTC category; (2) parietal cortex/VTC category; and (3) posterior cingulate/VTC category. As for all models described above, we assessed model fits for each separate model, first comparing a base model (for which age was not included as a predictor) to a model that included a main effect of age. We then further compared that best-fit model (Base vs. Main Effect) to a model that included an interaction with age. Modeling the interaction improved the model fits in posterior cingulate (p=.05) with a trend in this direction in parietal cortex (p=.13), which were both accompanied by significant interaction terms in the models themselves (see Results). In contrast, for caudate, neither the main effect nor the interaction models improved fit above the base model (ps > .31).

Trial-level drift diffusion models. Following previous work (Mack & Preston, 2016), we used drift diffusion modeling (DDM; Ratcliff, 1978) to assess whether reinstatement during the delay period predicted, for each individual trial, response speed on the subsequent match decision. In particular, we evaluated the association between drift rate v—defined as the rate at which evidence for a memory-based decision accumulates—and response time distributions on the mnemonic decision task. There were no trial restrictions based on behavior applied for the purposes of this analysis, as the DDM approach requires inclusion of both correct and incorrect trials. However, we did restrict our analysis to only *match* trials (i.e., mismatch trials were excluded) to equate the difficulty of the decision/matching process, which would be expected to vary across mismatch trials and therefore be an additional source of noise. Guided by the significant reinstatement effects on decision accuracy (Figure 5a), here we focused on trial-by-trial evidence of category-level reinstatement within the VTC fROI (Figure 3b; target reinstatement) as well as item-level reinstatement within the parietal cortex fROI (Figure 4a; raw same item estimates). We modeled the linear effect between VTC category reinstatement and parietal item reinstatement on the drift rate parameter. Within this model, we additionally

controlled for mean hippocampal BOLD response on each trial, as this was a significant predictor in the mixed effects accuracy models (unlike BOLD VTC and parietal responses; see Results). The models—each containing three predictors—were run separately for children and adults, using the Hierarchical Drift Diffusion Model toolbox (HDDM version 0.5.3; Wiecki et al., 2013) which implements hierarchical Bayesian parameter estimation using Markov-Chain Montel Carlo (MCMC) sampling methods. Specifically, models were estimated with 5,000 MCMC samples, after discarding the first 1,000 samples for burn-in and thinning of 2, resulting in a distribution of 2,000 regression coefficients.

To evaluate statistical significance among the neural predictors and the drift rate parameter, we computed the proportion of regression coefficients that were more or less than 0 depending on the sign of the coefficient median, constituting a p-value for each neural measure. Only predictors for which the corresponding p-values exceeded the 95th percentile (i.e., $p < .05$) were considered statistically significant. The models thus resulted in three regression coefficients reflecting the relationship between each of the neural measures (i.e., parietal item reinstatement, VTC category reinstatement, cue-period hippocampal BOLD activity) and drift rate for the match decisions. In situations where reinstatement is high, this should result in a faster accumulation of evidence (or drift rate) and a faster response time. When item reinstatement is low, evidence accumulates gradually and the RT is slower. Therefore, positive coefficients would provide evidence that greater reinstatement is associated with a higher drift rate, or faster evidence accumulation. Notably, all model fits were assessed beforehand, including checks on model convergence using the Gelman-Rubin R-hat statistic (see Mack & Preston, 2016 for details) as well as ensuring that the neural models provided a better fit (i.e., smaller deviance information criterion [DIC] value) than a base DDM model that did not include effects of reinstatement or hippocampal cue activation.

Acknowledgments

The data reported here were collected with support by funding from NIMH R01 grant 100121 (ARP), NICHD F32 grant 095586 (NLV), NSERC Discovery Grant RGPIN-04933-2018 (MLS), Canada Foundation for Innovation JELF (MLS), and Ontario Research Fund 36876 (MLS). The funders had no role in study design, data collection and analysis, or preparation of the manuscript. The authors extend their gratitude to Katharine Guarino, Christine Coughlin, Kim Nguyen, Lauren Quesada, and Drew (Francis) Hussey for assistance with piloting, data collection and analysis. We thank Tammy Tran for assistance with stimulus development. We also thank Katherine Sherrill and Neal Morton for input on data processing. Finally, we thank the families and participants, as this work would not have been possible without their help.

Author Contributions

M.L.S, M.L.M. and A.R.P. conceptualized the study. M.L.S., M.L.M., A.R.P., N.L.V. and H.E.R. designed the experimental task. N.L.V., H.E.R., L.M., and E.M.H. collected the data. N.L.V., M.L.S., M.L.M., and R.J.M. analyzed the data. N.L.V. drafted the paper, and M.L.S. and A.R.P. revised the manuscript for review and finalization by all other authors.

Competing Interests statement

The authors declare no competing interests.

Data Availability

The source data that support the findings of this study are provided in the Open Science Framework repository at https://osf.io/6ny4g/?view_only=8dc01c7fa3a6441b8267e9340024f712.

Code Availability

The custom code—including all required content and instructions for use—that supports the findings of this study are available in the Open Science Framework repository at https://osf.io/6ny4g/?view_only=8dc01c7fa3a6441b8267e9340024f712.

References

1. Chanales, A. J., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among spatial memories triggers repulsion of hippocampal representations. *Current Biology*, *27*(15), 2307-2317.
2. Favila, S. E., Chanales, A. J., & Kuhl, B. A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nature communications*, *7*(1), 11066.
3. Hulbert, J. C., & Norman, K. A. (2015). Neural differentiation tracks improved recall of competing memories following interleaved study and retrieval practice. *Cerebral Cortex*, *25*(10), 3994-4008.
4. Collin, S. H., Milivojevic, B., & Doeller, C. F. (2015). Memory hierarchies map onto the hippocampal long axis in humans. *Nature neuroscience*, *18*(11), 1562-1564.
5. Molitor, R. J., Sherrill, K. R., Morton, N. W., Miller, A. A., & Preston, A. R. (2021). Memory reactivation during learning simultaneously promotes dentate gyrus/CA2, 3 pattern differentiation and CA1 memory integration. *Journal of Neuroscience*, *41*(4), 726-738.
6. Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature communications*, *6*(1), 8151.
7. Bauer, P. J., & Varga, N. L. (2017). Similarity and deviation in event segmentation and memory integration: Commentary on Richmond, Gold, & Zacks. *Journal of Applied Research in Memory and Cognition*, *6*(2), 124-128.
8. Keresztes, A., Ngo, C. T., Lindenberger, U., Werkle-Bergner, M., & Newcombe, N. S. (2018). Hippocampal maturation drives memory from generalization to specificity. *Trends in Cognitive Sciences*, *22*(8), 676-686.
9. Bauer, P. J., Dugan, J. A., Varga, N. L., & Riggins, T. (2019). Relations between neural structures and children's self-derivation of new knowledge through memory integration. *Developmental cognitive neuroscience*, *36*, 100611.
10. Ghetti, S., & Bunge, S. A. (2012). Neural changes underlying the development of episodic memory during middle childhood. *Developmental cognitive neuroscience*, *2*(4), 381-395.
11. Østby, Y., Tamnes, C. K., Fjell, A. M., Westlye, L. T., Due-Tønnessen, P., & Walhovd, K. B. (2009). Heterogeneity in subcortical brain development: a structural magnetic resonance imaging study of brain maturation from 8 to 30 years. *Journal of neuroscience*, *29*(38), 11772-11782.
12. Schlichting, M. L., Guarino, K. F., Schapiro, A. C., Turk-Browne, N. B., & Preston, A. R. (2017). Hippocampal structure predicts statistical learning and associative inference abilities during development. *Journal of cognitive neuroscience*, *29*(1), 37-51.
13. Bauer, P. J., & San Souci, P. (2010). Going beyond the facts: Young children extend knowledge by integrating episodes. *Journal of experimental child psychology*, *107*(4), 452-465.
14. Varga, N. L., Esposito, A. G., & Bauer, P. J. (2019). Cognitive correlates of memory integration across development: Explaining variability in an educationally relevant phenomenon. *Journal of Experimental Psychology: General*, *148*(4), 739.
15. Ngo, C. T., Newcombe, N. S., & Olson, I. R. (2018). The ontogeny of relational memory and pattern separation. *Developmental science*, *21*(2), e12556.
16. Rollins, L., & Cloude, E. B. (2018). Development of mnemonic discrimination during childhood. *Learning & Memory*, *25*(6), 294-297.

17. Shing, Y. L., Finke, C., Hoffmann, M., Pajkert, A., Heekeren, H. R., & Ploner, C. J. (2019). Integrating across memory episodes: Developmental trends. *Plos one*, *14*(4), e0215848.
18. Bauer, P. J., Cronin-Golomb, L. M., Porter, B. M., Jaganjac, A., & Miller, H. E. (2021). Integration of memory content in adults and children: Developmental differences in task conditions and functional consequences. *Journal of Experimental Psychology: General*, *150*(7), 1259.
19. Schlichting, M. L., & Preston, A. R. (2015). Memory integration: neural mechanisms and implications for behavior. *Current opinion in behavioral sciences*, *1*, 1-8.
20. Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. *Psychological review*, *119*(3), 573.
21. Preston, A. R., & Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in memory. *Current biology*, *23*(17), R764-R773.
22. Varga, N.L., Morton, N. W, & Preston, A. R. (2024). Schema, inference, and memory. In M. J. Kahana & A. D. Wagner (Eds.), Handbook of human memory: Foundations and applications. Oxford University Press.
23. Schlichting, M. L., & Preston, A. R. (2014). Memory reactivation during rest supports upcoming learning of related content. *Proceedings of the National Academy of Sciences*, *111*(44), 15845-15850.
24. Varga, N. L., & Manns, J. R. (2021). Delta-modulated cortical alpha oscillations support new knowledge generation through memory integration. *NeuroImage*, *244*, 118600.
25. Zeithamova, D., Dominick, A. L., & Preston, A. R. (2012). Hippocampal and ventral medial prefrontal activation during retrieval-mediated learning supports novel inference. *Neuron*, *75*(1), 168-179.
26. Bauer, P. J., Varga, N. L., King, J. E., Nolen, A. M., & White, E. A. (2015). Semantic elaboration through integration: Hints both facilitate and inform the process. *Journal of Cognition and Development*, *16*(2), 351-369.
27. Schlichting, M. L., Guarino, K. F., Roome, H. E., & Preston, A. R. (2022). Developmental differences in memory reactivation relate to encoding and inference in the human brain. *Nature Human Behaviour*, *6*(3), 415-428.
28. Wilson, J. T., & Bauer, P. J. (2021). Prompt-facilitated learning: The development of unprompted memory integration and subsequent self-derivation. *Memory & Cognition*, *49*(7), 1473-1487.
29. Sastre III, M., Wendelken, C., Lee, J. K., Bunge, S. A., & Ghetti, S. (2016). Age-and performance-related differences in hippocampal contributions to episodic retrieval. *Developmental cognitive neuroscience*, *19*, 42-50.
30. Callaghan, B., Gasser, C., Silvers, J. A., VanTieghem, M., Choy, T., O'Sullivan, K., ... & Tottenham, N. (2021). Age-related increases in posterior hippocampal granularity are associated with remote detailed episodic memory in development. *Journal of Neuroscience*, *41*(8), 1738-1754.
31. Fandakova, Y., Leckey, S., Driver, C. C., Bunge, S. A., & Ghetti, S. (2019). Neural specificity of scene representations is related to memory performance in childhood. *NeuroImage*, *199*, 105-113.
32. Kazemi, A., Coughlin, C. A., DeMaster, D. M., & Ghetti, S. (2022). Contextual features in the developing hippocampus: A representational similarity analysis. *Hippocampus*, *32*(4), 286-297.

33. Qin, S., Cho, S., Chen, T., Rosenberg-Lee, M., Geary, D. C., & Menon, V. (2014). Hippocampal-neocortical functional reorganization underlies children's cognitive development. *Nature neuroscience*, *17*(9), 1263-1269.
34. Benear, S. L., Horwath, E. A., Cowan, E., Camacho, M. C., Ngo, C. T., Newcombe, N. S., ... & Murty, V. P. (2022). Children show adult-like hippocampal pattern similarity for familiar but not novel events. *Brain Research*, *1791*, 147991.
35. Marr, D., 1971. Simplememory: a theory for archicortex. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 262, 23–81.
36. Ritchey, M., Wing, E. A., LaBar, K. S., & Cabeza, R. (2013). Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. *Cerebral cortex*, *23*(12), 2818-2828.
37. Trelle, A. N., Carr, V. A., Guerin, S. A., Thieu, M. K., Jayakumar, M., Guo, W., ... & Wagner, A. D. (2020). Hippocampal and cortical mechanisms at retrieval explain variability in episodic remembering in older adults. *Elife*, *9*, e55335
38. Kuhl, B. A., & Chun, M. M. (2014). Successful remembering elicits event-specific activity patterns in lateral parietal cortex. *Journal of Neuroscience*, *34*(23), 8051-8060.
39. Mack, M. L., & Preston, A. R. (2016). Decisions about the past are guided by reinstatement of specific memories in the hippocampus and perirhinal cortex. *Neuroimage*, *127*, 144-157.
40. Brod, G. (2021). Generative learning: Which strategies for what age?. *Educational Psychology Review*, *33*(4), 1295-1318.
41. Brod, G., Lindenberger, U., & Shing, Y. L. (2017). Neural activation patterns during retrieval of schema‐related memories: Differences and commonalities between children and adults. *Developmental science*, *20*(6), e12475.
42. Fandakova, Y., Selmeczy, D., Leckey, S., Grimm, K. J., Wendelken, C., Bunge, S. A., & Ghetti, S. (2017). Changes in ventromedial prefrontal and insular cortex support the development of metamemory from childhood into adolescence. *Proceedings of the National Academy of Sciences*, *114*(29), 7582-7587.
43. Ofen, N., Kao, Y. C., Sokol-Hessner, P., Kim, H., Whitfield-Gabrieli, S., & Gabrieli, J. D. (2007). Development of the declarative memory system in the human brain. *Nature neuroscience*, *10*(9), 1198-1205.
44. Simmonds, D. J., Hallquist, M. N., Asato, M., & Luna, B. (2014). Developmental stages and sex differences of white matter and behavioral development through adolescence: a longitudinal diffusion tensor imaging (DTI) study. *Neuroimage*, *92*, 356-368.
45. Riggins, T. (2014). Longitudinal investigation of source memory reveals different developmental trajectories for item memory and binding. *Developmental Psychology*, *50*(2), 449.
46. Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 4.
47. Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in cognitive sciences*, *10*(9), 424-430.
48. Mumford, J. A., Turner, B. O., Ashby, F. G., & Poldrack, R. A. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *Neuroimage*, *59*(3), 2636-2643.
49. Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J. A., & Poldrack, R. A. (2010). Greater neural pattern similarity across repetitions is associated with better memory. *Science*, *330*(6000), 97-101.
50. Hajcak, G., & Dennis, T. A. (2009). Brain potentials during affective picture processing in children. *Biological psychology*, *80*(3), 333-338.

51. Karmiloff-Smith, A. (1986). From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. *Cognition*, *23*(2), 95-147.
52. Darby, K. P., & Sloutsky, V. M. (2015). The cost of learning: interference effects in memory development. *Journal of Experimental Psychology: General*, *144*(2), 410.
53. Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. *Psychological review*, *110*(4), 611.
54. Van den Bos, W., Cohen, M. X., Kahnt, T., & Crone, E. A. (2012). Striatum–medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. *Cerebral cortex*, *22*(6), 1247-1255.
55. Abolghasem, Z., Teng, T. H. T., Nexha, E., Zhu, C., Jean, C. S., Castrillon, M., ... & Schlichting, M. L. Learning strategy differentially impacts memory connections in children and adults. *Developmental Science*, e13371.
56. Skalaban, L. J., Chan, I., Rapuano, K. M., Lin, Q., Conley, M. I., Watts, R. R., ... & Casey, B. J. (2024). Representational Dissimilarity of Faces and Places during a Working Memory Task is Associated with Subsequent Recognition Memory during Development. *Journal of Cognitive Neuroscience*, *36*(3), 415-434.
57. Ritvo, V. J., Nguyen, A., Turk-Browne, N. B., & Norman, K. A. (2023). Differentiation and Integration of Competing Memories: A Neural Network Model. *bioRxiv*.
58. Morton, N. W., Schlichting, M. L., & Preston, A. R. (2020). Representations of common event structure in medial temporal lobe and frontoparietal cortex support efficient inference. *Proceedings of the National Academy of Sciences*, *117*(47), 29338-29345.
59. Chan, S. C., Niv, Y., & Norman, K. A. (2016). A probability distribution over latent causes, in the orbitofrontal cortex. *Journal of Neuroscience*, *36*(30), 7817-7828.
60. Bein, O., Reggev, N., & Maril, A. (2020). Prior knowledge promotes hippocampal separation but cortical assimilation in the left inferior frontal gyrus. *Nature Communications*, *11*(1), 4590.
61. Zeithamova, D., & Preston, A. R. (2017). Temporal proximity promotes integration of overlapping events. *Journal of cognitive neuroscience*, *29*(8), 1311-1323.
62. Richter, F. R., Chanales, A. J. H., & Kuhl, B. A. (2016). Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. NeuroImage, 124, 323–335.
63. P.R. Millar, S.J. Serbun, A. Vadalia, A.H. Gutchess. Cross-cultural differences in memory specificity. Culture and Brain, 1 (2–4) (2013), pp. 138-157, 10.1007/s40167-013-0011-3
64. Baldassano, C., Hasson, U., & Norman, K. A. (2018). Representation of real-world event schemas during narrative perception. *Journal of Neuroscience*, *38*(45), 9689-9699.
65. Pudhiyidath, A., Morton, N. W., Viveros Duran, R., Schapiro, A. C., Momennejad, I., Hinojosa-Rowland, D. M., ... & Preston, A. R. (2022). Representations of temporal community structure in hippocampus and precuneus predict inductive reasoning decisions. *Journal of Cognitive Neuroscience*, *34*(10), 1736-1760.
66. Vilberg, K. L., & Rugg, M. D. (2008). Memory retrieval and the parietal cortex: a review of evidence from a dual-process perspective. *Neuropsychologia*, *46*(7), 1787-1799.
67. Zhang, Z., Yin, C. & Yang, T. Evidence accumulation occurs locally in the parietal cortex. Nat Commun **13**, 4426 (2022). PMC9339004.
68. Tamnes, C. K., Walhovd, K. B., Grydeland, H., Holland, D., Østby, Y., Dale, A. M., & Fjell, A. M. (2013). Longitudinal working memory development is related to structural maturation of frontal and parietal cortices. *Journal of cognitive neuroscience*, *25*(10), 1611-1623.

69. Howe, M. L., & O'Sullivan, J. T. (1997). What children's memories tell us about recalling our childhoods: A review of storage and retrieval processes in the development of long-term retention. *Developmental Review*, *17*(2), 148-204.
70. Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature neuroscience*, *19*(5), 665-667.