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[Ŧ1,2,4], & Margaret L. Schlichting[Ŧ6]

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: [Ŧ]Reflects equal contributions

Version Note: May 25, 2023 – Submitted for publication

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). Both children and adults formed differentiated representations in hippocampus—such that related events were stored as *less* similar than unrelated events. Conversely, only adults formed integrated representations 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 generalization, respectively. Yet, and despite much theoretical interest in this topic[7-9], how children represent such related experiences to enable successful memory reinstatement and memory-guided behavior remains unknown. Here, we test the hypothesis that children's memory representations may differ from adults' given prolonged development of hippocampal-prefrontal systems[9,10-11]. Empirical insight into how memory representation differs across development has important implications for understanding not only what is remembered, but also a myriad of cognitive and academic behaviors that rely on memory representations.

Compared with adults, children have more difficulty both disambiguating between[12-13] and integrating across[14-17] related memories. Such behavioral differences may reflect functional shifts within the neural substrates that support memory formation. Nevertheless, children do form and retrieve memories for related events, with the ability to do so predicting the likelihood of integrating across memories[18], and in turn, academic success[1919]. Thus, it is critical that we understand how children represent related memories in ways that enable successful retrieval and memory-guided behavior later. Notably, given hippocampal development continues across childhood[9-10,20], one possibility is that children's memory representations may be biased in favor of one representational scheme. Yet, which scheme remains an open question, with current developmental theories alternately suggesting that children rely more on either generalization[9] or storage of distinct memory traces[7-8] to represent related events. However, these ideas are founded predominantly on behavioral evidence, which cannot elucidate the nature of hippocampal representations. Here, we measure such representations directly, asking whether the same memory behavior arises from different representations in children and adults.

While it has been shown that children and adults alike engage hippocampus to support retrieval of overlapping memories on average[21], such findings cannot tell us how related experiences are coded 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[22-25], few have quantified representations of related events as they are experienced[26], let alone how related events are stored and retrieved. However, one recent study offers a suggestion 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 across repeated exposures, such reinstatement diminished in younger participants, suggesting that differentiation—in which *differences* between similar experiences are emphasized—may be a more prevalent representational scheme earlier in development. While this study implies that differentiation supports memory-based decisions over integration at younger ages, it did not measure the representational schemes directly, and therefore cannot address how overlapping events are organized in the developing brain. Here, we directly measure the organizational schemes of successfully retrieved hippocampal memories in children and adults, informing the fundamental question of how memory representation differs over developmental time.

We also interrogate how hippocampal representation impacts neocortical reinstatement. Reinstatement of distributed neocortical traces enables high-fidelity recollection of the individual features of past events[28], which are bound together by hippocampus[29]. If hippocampal differentiation and/or integration provides a means of effectively organizing related events with respect to one another, it thus follows that such neural organization should have downstream consequences for neocortical reinstatement of the individual features associated with those events[29-30], particularly in higher-order perceptual regions such as ventral temporal (VTC) and parietal cortex[31-32]. Here, we ask whether the fidelity of neocortical memory representations influences later mnemonic decision making in children as it does in adults[30], and moreover, whether the degree of reinstatement relates to hippocampal representations. Such an approach could reveal whether children and adults alike depend on high-fidelity neocortical reinstatement to guide behavior, as well as whether they do so through similar or different memory codes.

Here, children and adults learned to associate different objects with the same person or place. To ask how those memories were represented in hippocampus, 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. Our design also allowed us to use neural pattern analyses to measure neocortical reinstatement of associative memory features (i.e., a person/place). Our goal was to test whether, when children and adults *successfully* recall overlapping memories—and their associative features—they do so based on similar or different types of memory representation. To this end, we focused on middle childhood (7-10 years), as the ability to retrieve associative features[33-35] and overlapping[12-13] events is available by this age. Although such behavioral evidence suggests that the developing brain can represent related experience in ways that enable recollection of specific details later, here we ask whether such

behavioral retrieval success in children and adults is underpinned by differences in the organization and/or fidelity of the supporting hippocampal and neocortical memory traces.

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 (Fig. 1a). The same person or place (six each) was paired with three different objects, for a total of 36 overlapping pairs. We aimed to produce high-level performance in both age groups, which allowed us to ask about the neural organizational schemes that enabled successful learning and retrieval at different ages. To this end, before MRI scanning (Fig. 1b), participants completed between two-five learning/retrieval repetitions until a 90% performance criterion was reached (most, N=48, had two; Fig. 1c; Supplementary Data: Age-related learning differences).

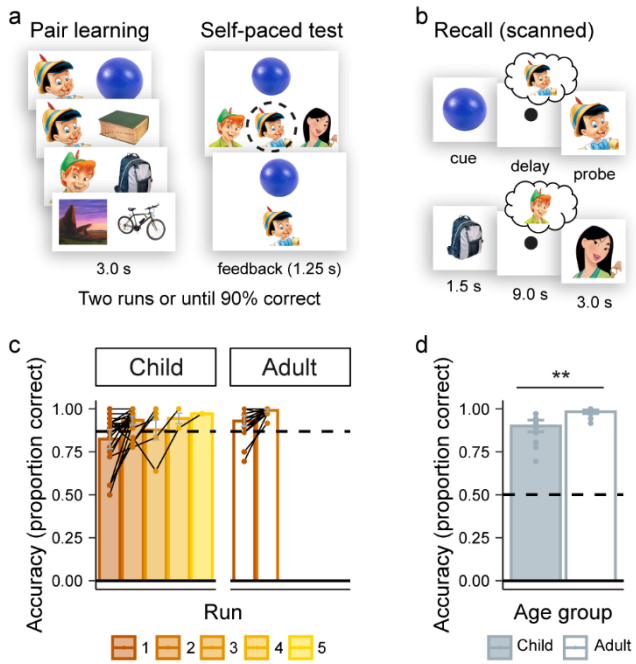


Fig. 1. Procedure schematics on the top depict the initial learning/retrieval and scanned recall tasks while figures on the bottom show corresponding behavioral performance on those tasks. (a) Initial learning/retrieval. On each run, participants studied the same 36 person-object and place-object pairs 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. The paths connecting dots across runs (solid black) depict learning trajectories for each participant. The dashed black line reflects the 90% learning criterion. (d) Scanned recall accuracy. The proportion of correct match/mismatch decisions was averaged across runs and participants, separately for each age group. The dashed black line reflects chance-level performance. Asterisks indicate a significant group difference at $p < .001$. For panels c and d, dots reflect individual participant means and error bars reflect 95% confidence intervals.

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 (Fig. 1b). Both children and adults retrieved the overlapping memories, as evidenced by high and above-chance performance in both groups (*t*s>25.39, *ps*<.001; Fig. 1d;). Yet children still did not reach adult levels of performance ($t(31.49)=5.62$, $p<.001$; Fig. 1d). Thus, as a further step to isolating the representational schemes that enabled successful memory, neural analyses were restricted to trials in which retrieval was successful during final learning and scanned recall (Supplementary Table 1 for trial counts).

Children and adults differentiate related memories. To assess hippocampal organization of related memories, we used representational similarity analysis (RSA)[36] 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 (Fig. 2a). In such cases, the related objects might evoke patterns that are significantly 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 main effect between related and unrelated objects ($F(1,50)=4.54$, $p=.04$, $p\eta^2=.08$) that was qualified by an interaction with age ($F(1,50)=3.93$, $p=.05$, $p\eta^2=.07$), such that children ($F(1,26)=6.16$, $p=.02$, $p\eta^2=.19$) but not adults ($F(1,24)=.02$, $p=.89$, $p\eta^2=.001$) exhibited hippocampal representations for related objects that were *less* similar than those for unrelated objects (Fig. 2b). Notably, this differentiation pattern was absent in visual regions that selectively process 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: Hippocampal differentiation controlling for univariate activation). The approach thus enabled us to uncover subtle differences in the organization of hippocampal memory traces not reflected in neocortex nor via mean hippocampal activation.

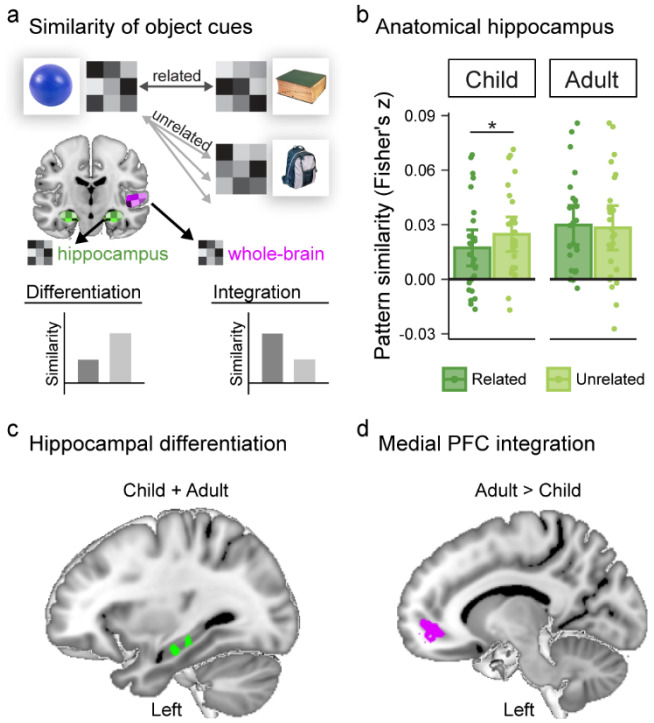


Fig. 2. Children and adults alike differentiate related object cues, while adults also integrate. (a) Depiction of representational similarity analyses. We compared the similarity of fMRI patterns for related as compared with unrelated objects across trials and participants to assess evidence for differentiation (related < unrelated) and integration (related > unrelated). (b) Neural patterns extracted from bilateral anatomical hippocampus. Color-coded bars reflect pairwise similarity values averaged across object cues that were Related or Unrelated, separately for participants within each age group. Error bars reflect 95% confidence intervals around the group mean. Dots reflect individual participant means. Asterisk reflects a significant condition effect at a threshold of $p < .05$. (c) Hippocampal searchlight regions showing a significant main effect of differentiation when examined across the full sample (Child + Adult) displayed on the 1-mm MNI template brain. Clusters are significant after small-volume correction for multiple comparisons within anatomical hippocampus. See Supplementary Table 2 for additional regions that showed main effects and developmental differences in differentiation across the whole brain. (d) Whole-brain searchlight region showing a significant effect of integration in medial PFC which varied by age group, with adults showing significantly greater integration than children. Cluster is significant after correction for multiple comparisons across the whole brain (Supplementary Figure 2 for plot of Related versus Unrelated values). See Supplementary Table 2 for additional regions that showed main effects or age differences in integration across the whole brain.

DEVELOPMENTAL DIFFERENCES IN MEMORY REPRESENTATION

The findings within our *a priori* anatomical hippocampal region of interest (ROI) establish developmental differences in differentiation of 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]. This possibility is particularly likely for adults who, rather than recruit the entire hippocampus when retrieving related memories, rely on focal subregions[21]. In separate searchlight analyses, we therefore looked for voxels exhibiting differentiation or integration of related object cues in hippocampus and neocortex (Fig. 2a). We searched for voxels that showed developmental differences in a representational scheme (two-sample t-test: adult>child; child>adult), as well as voxels that showed a scheme irrespective of age (one-sample t-test: adult+child > 0).

Within hippocampus, clusters in left anterior (Fig. 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 body (Fig. 2c: x, y, z = -29.7, -27.2, -12.8) showed significant differentiation across the whole group (adult+child > 0). Developmental differences in hippocampal differentiation were not observed. However, it is noteworthy that when interrogated in the two age groups separately, these same hippocampal clusters emerged in the child group (at significant or trend levels; anterior: $t(26)=-2.27$, $p=.03$; body: $t(26)=-1.99$, $p=.06$), but not the adult group (ps>.36). This pattern hints that while differentiation can aid memory organization in children and remains available in adults, this scheme may be less consistently invoked by adults. In line with this suggestion, RSA searchlights across the whole brain revealed a number of additional clusters showing significant effects of differentiation, with several clusters showing enhanced differentiation in children but none showing the opposite (Supplementary Table 2).

Integration is more pronounced in adults. Evidence of less pronounced differentiation among adults aligns with our hypothesis that, in contrast to children, adults may make flexible use of multiple representational schemes when organizing related events[5-6]. As such, we next asked whether integration was more pronounced in the adult hippocampus or elsewhere in the brain. We found no evidence of hippocampal integration in adults nor children. However, the whole-brain searchlight revealed that adults showed greater integration in a number of regions (Supplementary Table 2), including medial prefrontal cortex (mPFC) (Fig. 2d; two-sample t-test: adult>child; x, y, z = -15.30, 44.80, -4.43). Within the regions showing integration in adults, children showed either no reliable representational scheme (mPFC) or differentiation (Supplementary Figure 2 for within-group effects), suggesting that the neural substrates implicated in integrated representation in adulthood are not necessarily recruited for the same 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 hippocampal differentiation and adults additionally showing mPFC integration, 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 (f)ROIs within anatomical VTC and parietal cortex (Fig. 3a; Methods), guided by prior work[32]. To quantify the degree to which the target category was reinstated during the delay period, we trained a multi-voxel pattern analysis (MVPA) classifier[37] to identify patterns of activity associated with face and scene processing within each fROI (Supplementary data: Classifier validation accuracy; Methods). We then applied the trained classifier to patterns of fMRI activation measured during the 9-second delay interval (Fig. 3b), estimating the amount of face and scene activation evidenced (Methods). Both age groups showed categorical reinstatement in VTC (adult: $t(24)=15.94$, $p<.001$; child: $t(26)=11.24$, $p<.001$)—consistent with its established role in category coding[32]—while adults additionally showed reinstatement in parietal cortex ($t(24)=5.82$, $p<.001$; Fig. 3c).

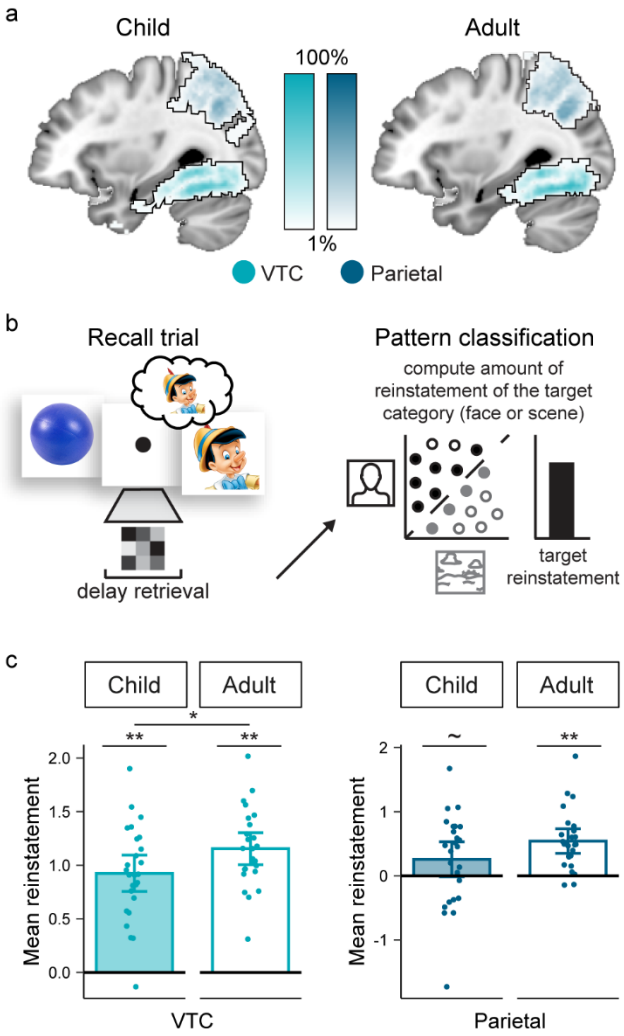


Fig. 3. Children and adults reinstate the target category. (a) Voxels included within the VTC and parietal fROIs, separately for each age group. For each participant, we tested for reinstatement in the voxels that they engaged most during face and scene perception in a separate localizer task, within VTC and parietal cortex separately. Maps show the percentage of participants who had that same voxel in their fROI is 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, 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 solid line) indicates significant target category evidence. Dots reflect individual participant means. Error bars reflect 95% confidence intervals around the mean.

We next examined the degree to which the target item was reinstated during the delay. Here, 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; Fig. 4a), as compared to viewing other items from the same visual category (different-item similarity; Methods - general linear model [GLM] approach[38-39]). Item-level reinstatement in VTC was significant in adults ($t(24)=2.32$, $p=.02$; 1-tailed) and a trend in children ($t(26)=1.52$, $p=.07$; 1-tailed), but not significant at either age in parietal cortex (ts<1.13; ps>.14; 1-tailed). Given only modest item reinstatement effects in the category-selective fROIs, for both children and adults, we additionally conducted item-level reinstatement searchlight analyses across the brain, which were small-volume corrected within the VTC and parietal fROIs as well as within hippocampus given prior adult work[30]. For both groups, we found significant item reinstatement in the fROIs and elsewhere (Supplementary Table 3), though developmental differences were observed in hippocampus, with the locus of item reinstatement shifting from posterior to anterior between childhood and adulthood (Supplementary Figure 3).

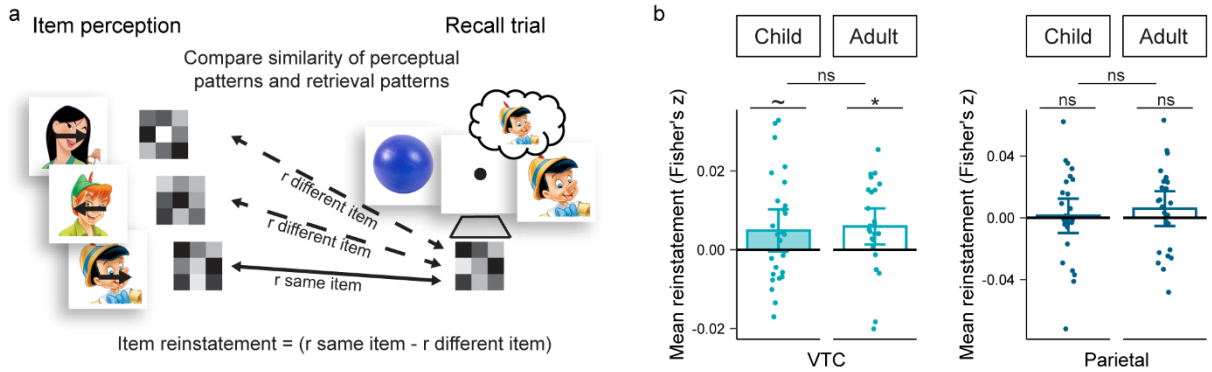


Fig. 4. Children and adults reinstate target items. (a) Representational similarity analysis approach. Representational similarity analysis was implemented at the level of the full category-selective fROIS as well as at the individual voxel level via separate searchlight analyses (see Supplementary Table 3 & Supplementary Figure 3). 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 category-selective VTC (left) and parietal cortex (right) fROIs from Figure 3a. A mean reinstatement index reliably above 0 indicates significant item evidence, reflecting greater same-item similarity as compared to the different-item baseline. Dots reflect individual participant means. Error bars reflect 95% confidence intervals around the mean.

High-fidelity memory reinstatement facilitates mnemonic decision making. One important clue as to whether memory reinstatement directly guides memory behavior comes

from asking how it relates to decision behavior on a trial-by-trial basis. That is, if neural reinstatement reflects the fidelity of retrieved memories, then it follows that the degree of neocortical reinstatement during the delay will relate to subsequent decision behavior, in which participants were asked to judge whether the retrieved memory matched the decision probe (Fig. 1b). To test this, we conducted mixed effects and drift diffusion models to examine whether variability in item (Fig. 4b, same-item similarity) and category (Fig. 3c) reinstatement predicted the probability of making a correct and/or faster response within a participant on a given trial (Methods). Indeed, we found that more strongly reinstated memories were associated with more accurate responses (VTC category reinstatement: χ^2_1=4.12, *p*=.04, Fig. 5a; trend for parietal cortex item reinstatement: χ^2_1=3.52, *p*=.06, Fig. 5a), even when accounting for the main effect of age (χ^2_1= 29.08, *p* < .001; model comparison: AIC_{base}=848.16, AIC_{age}=817.74, χ^2_1=32.42, *p*<.001) and nuisance regressors, including cue-period univariate hippocampal activation as well as delay-period VTC and parietal univariate activation. Follow-up analyses of each age group separately hinted that this effect may be specific to children (VTC: χ^2_1=3.62, *p*=.06; Parietal: χ^2_1=3.42, *p*=.06) not adults (ps>.56). Likewise, in adults (who reached ceiling levels of accuracy), faster responses were associated with more strongly reinstated memories while controlling for hippocampal BOLD activation during the cue period (Fig. 5b; parietal item).

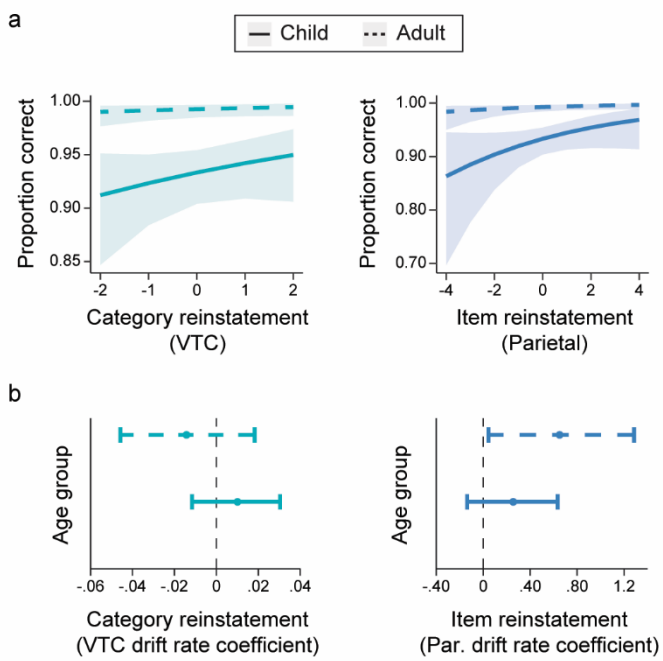


Fig. 5. Enhanced neocortical reinstatement in children and adults promotes superior memory behavior. In both panels, for each participant and retrieval trial separately, we tested the association among raw same-category and same-item reinstatement during the delay period with the probability of making a correct (a) or speeded (b) match/mismatch behavioral response. (a) Mixed effects models examining decision accuracy. In children, 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. For both plots, the predictors were scaled and centered within participants to remove participant-specific effects. (b) Drift diffusion models (DDM) examining decision speed. Here, the VTC category and parietal cortex item reinstatement indices were entered into separate DDM models of response speed, modeled as a linear effect on the drift rate parameter that reflects the rate of evidence accumulation. Thus, 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 plotted as forest plots. The bars represent 95% confidence intervals of the 2,000 posterior parameter estimates, with circles representing the mean of the posterior parameter distributions. We found that greater item reinstatement in parietal cortex predicted faster response times, but only in adults, which was reflected by drift rate coefficient confidence intervals 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 (Figs. 3&4) in service of mnemonic decisions (Fig. 5). However, whereas children seem to rely on hippocampal differentiation to organize related memories (Fig. 2b), adults additionally show mPFC integration (Fig. 2d). In a final set of analyses, we thus tested whether the degree of hippocampal and mPFC cue differentiation and integration, respectively, was associated with neocortical reinstatement in children and adults separately (see Methods). We restricted our

consideration of the association between cue organization and delay reinstatement to those fROIs that either showed reliable reinstatement in both age groups separately (i.e., category reinstatement in VTC; Fig. 3c) or to those for which trial-wise variability in reinstatement was linked to behavioral success in both age groups (i.e., item reinstatement in parietal; Fig. 5). This choice ensures that the reinstatement regions are matched in terms of behavioral relevance and reliable recruitment across age, and thus reduces the likelihood of spurious age differences in the association between representational organization and reinstatement. In both groups, we found associations with category-level reinstatement in VTC (but not item-level in parietal). Specifically, greater hippocampal cue differentiation in children (χ^2_1=6.51, *p*=.01; Fig. 6a) and mPFC integration in adults (χ^2_1=4.30, *p*=.04; Fig. 6b) were each related to category-level VTC reinstatement, consistent with the interpretation that different representational schemes are differentially beneficial for memory organization, reinstatement, and behavior at different ages.

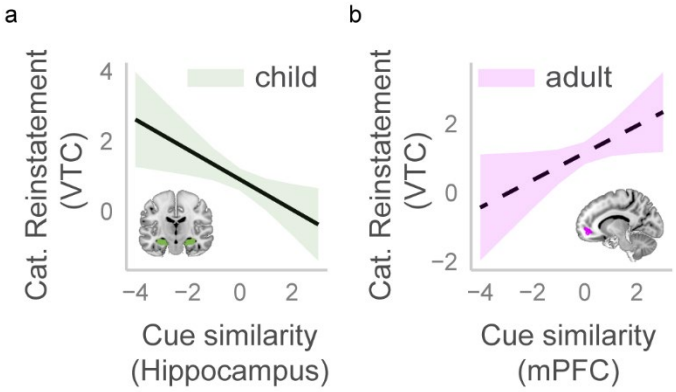


Fig. 6. The hippocampal and mPFC representations that governed how related memories were organized tracked the degree of ensuing neocortical reinstatement of the associated face and scene elements at the within-person level. In both panels, for each participant and retrieval trial separately, we computed the averaged similarity of a retrieval cue object with related cue objects that shared the same face or scene associate (Fig. 2a; related similarity). Thus, whereas lower cue similarity values reflect greater dissimilarity of related memories; higher values reflect greater similarity. We tested the association among that cue similarity value with reinstatement during the corresponding delay period for that same retrieval trial (Fig. 3b; Target reinstatement). (a) In children, there was a negative trial-wise association between cue similarity in bilateral hippocampus 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, there was a positive trial-wise association between cue similarity in mPFC and category reinstatement in VTC. That is, greater similarity in mPFC was associated with higher-fidelity reinstatement of the target category. For both panels, the cue similarity regions were selected for showing a systematic neural organizational scheme, namely (a) a main effect of differentiation in children (but not adults) or (b) significant integration in adults (but no reliable coding scheme in children). For both panels, the predictors were scaled and centered within participants to remove participant-specific effects.

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—and in particular, one that enables representation of related experiences in terms of both their commonalities and differences[40]. Yet despite both its theoretical importance[7-9] and educational relevance[19], the question of how children, for whom the neural substrates supporting memory are still developing[10-11,20], represent related events in ways that enable later memory retrieval remains unanswered. By leveraging fMRI with representational analysis methods, we show that when learning related events, adults make use of multiple representational schemes that simultaneously emphasize differences between events in hippocampus and commonalities across them in mPFC. In contrast, we show that children deploy a single representational scheme—hippocampal differentiation—wherein events that shared overlapping features were represented as *less* similar to one another compared to unrelated events. Critically, different representational schemes had benefits for reinstatement of associative event features at different ages, which in turn, benefitted mnemonic decision making. Hippocampal differentiation tracked higher-fidelity neocortical reinstatement during recall in children while mPFC integration tracked higher-fidelity neocortical reinstatement in adults. These findings indicate that children may be biased toward disambiguating overlapping experiences, an organizational scheme that may be especially important earlier in life as a way of minimizing interference between highly similar events[41]. While such a representational scheme improves children's recall of individual events (or associations as measured here), it may be limited relative to adults' capacity to represent both event differences and commonalities. Evidence for mPFC integration in adults may promote building complex representations that flexibly extend beyond direct experience[42].

Our primary objective in designing the present study was to isolate how the developing brain represents overlapping memories in ways that enable successful retrieval later. Although children often perform less accurately than adults when both disambiguating between[13] and integrating across[14-17] overlapping events, such developmental differences need not reflect a fundamental inability. Indeed, children as young as 6 years can successfully recall and connect related experiences[18]. From approximately 7 to 10 years, further developments are observed, with single-shot learning of individual associations[43] and discrimination of highly similar memories[13] becoming adult-like. Notably, in addition to the emergence of more adult-like

associative memory capacities from 7-10 years, this age period also marks the offset of accelerated childhood forgetting[44].

Collectively, these behavioral findings hint at important changes in the neurobiological system that supports representation of related events, and highlights middle childhood (7-10 years) as a candidate window in which this memory system has achieved at least some adult-like functionality. Consistent with this idea, children with medical disorders that damage the hippocampus have impaired associative memory, with such impairments being more pronounced if the insult occurs before age 7[45]. This finding points to a potential sensitive window for development of the hippocampus before age 7, which converges with observations of catastrophic retroactive interference for related memories prior to this age[41,46], as resistance to such interference has been linked to hippocampal function in adults[47]. Yet neither the neurobiological data, nor changes in memory behavior, can tell us how children successfully represent and recall related experiences, or whether they do so on the basis of similar or different types of memory representation as evidenced in adults. By quantifying representational similarity of related memories as they were recalled from memory, we thus offer a direct window into the representational competencies available within the early hippocampal memory system.

We found that hippocampal differentiation, but not integration, was associated with successful learning and retrieval of related memories in children. We further show that there were pronounced developmental differences in integration, such that adults showed significantly greater integration in mPFC, as well as within parietal cortex and posterior cingulate. These findings suggest that organization of overlapping memories may be supported by differentiated representation in hippocampus early in life, with the capacity to link related events in service of forming integrated knowledge structures becoming more refined as development of prefrontal cortex and its connections with hippocampus and/or neocortical regions progresses through adulthood[10-11,13]. That adults also showed hippocampal differentiation further supports this interpretation. That is, if differentiated representation constitutes the primary means through which the still-developing brain stores related experiences, we would expect this neural scheme to remain available in the more developed brain, after other representational schemes subsequently become available.

Notably, the general lack of evidence for neural integration in children (found only in caudate; Supplementary Table 2) is consistent with a growing body of behavioral work[14-17], and cannot be explained by age-related differences in memory success, as overall performance was high and we moreover only considered neural representations for correct trials here. This developmental difference is also unlikely to be due to differential experience, as has been

suggested to drive other age-related representational differences reported in the literature[26], as in addition to controlling for stimulus familiarity, children had the same amount of exposures to the pairs compared to adults. One limitation, however, is that matching learning exposures between groups, when most adults learned the pairs in a single exposure, may have enabled adults to employ an overt integration strategy on subsequent exposures. As such, whether young learners may form integrated representations when information is over-learned or under a different set of learning conditions remains an open question.

On the surface, our finding that hippocampal differentiation enabled related memory in children may appear to conflict with recent perspectives which have argued that memory specificity for related events continues to develop through middle childhood[9]. On the contrary, the finding that 76% of adults reached the retrieval criterion within a single learning exposure, in contrast to only 37% of children, lends support for the view that memory for specific, related events is undergoing further development. Likewise, evidence of hippocampal differentiation here—an active coding strategy often attributed to resolving interference among related memories that may be simultaneously reactivated during related encoding[1-3]—may seem incompatible with other work which has shown that children under 10 fail to reinstate overlapping memories[27]. Importantly, in the present research, neural representations were measured only following multiple learning and retrieval attempts. In contrast, the neural reinstatement in prior research was assessed during initial learning, wherein retrieval of related memories was not overtly encouraged. Because children show differences in their use of organizational schemes during learning[48], we speculate that interleaving learning exposures with explicit retrieval tests here may have served to strengthen memory formation, and consequently, differentiated representation and/or later reinstatement. Though the current design necessitated bringing all individuals to high levels of specific memory performance, we hypothesize that variability in the capacities to retrieve and extend related memories may be associated with the rate at which the hippocampal memory system—and differentiated neural organization, in particular—becomes functionally mature across individuals. The present results thus provide fundamental insights into memory representation in children that could have important implications for our understanding of how to promote learning of and memory for related information, key predictors of educational success across the lifespan[19].

By testing the relationship between representations in hippocampus and neocortex, the current study further establishes a role for hippocampus in organizing neocortical traces in development—a correspondence that has only been shown in adults[49,30]. Theories of memory development have long speculated that memory in childhood depends on the integrity of the

hippocampal-neocortical system[10,20], and the strength with which distributed memory traces are bound together, in particular[47]. In the present study, both children and adults showed evidence of category- and item-level reinstatement of associative memory elements in neocortex, demonstrating the feasibility of quantifying retrieval of the internal contents of specific memories in children. While prior work has shown that children and adults form similar neocortical representations of specific items at encoding[23], here we extend this finding to show comparable representations once such experiences are retrieved—a critical step toward characterizing the fidelity with which information is represented in memory. Moreover, the degree to which the associated faces and scenes were reinstated in neocortex 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.

Having established the behavioral relevance of hippocampal, mPFC, and neocortical representations, the 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, and by extension, memory behavior. Indeed, we found that trial-by-trial hippocampal differentiation among children and mPFC integration among adults each tracked with category reinstatement in VTC. Taken together, our results reveal that children and adults achieve the same profile of mnemonic performance and neocortical memory reinstatement, but through different representational pathways, such that the hippocampal and prefrontal representations that govern how memories are organized—and their functional significance for guiding behavior—change with age.

By identifying the specific parts of hippocampus associated with organization of related memories (Fig. 2b) from those involved in reinstatement of specific memory features (Supplementary Fig. 3), our results also help explain the mixed profile of hippocampal BOLD activation reported in past research. Cross-sectional univariate data suggests children first rely on either the full hippocampus[21] and/or posterior regions[50] when retrieving event-specific details associated with overlapping events, before eventually recruiting anterior hippocampus as shown in adulthood[21,30]. Here, we show that whereas children recruit the whole hippocampus when differentiating related events (Fig. 2b), they recruit only posterior hippocampus to reinstate the specific items associated with those events (Supplementary Fig. 3 - left). These results in children—coupled with evidence of both differentiation (Fig. 2c) and item reinstatement in anterior hippocampus in adults (Supplementary Fig. 3 – right)—thus converges with the view that memory processes are differently distributed in the immature hippocampus[50] with

refinements across the long-axis through adulthood supporting more complex representational and mnemonic functions[4-6]. Here, we build on this proposal, showing that the early hippocampus prioritizes memory specificity, through disambiguating similar events via differentiation while simultaneously forming item-specific traces in posterior regions.

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 benefitted from different representational schemes at different ages. Specifically, children benefitted from differentiation in the hippocampal system, such that this learning scheme was associated with higher-fidelity reinstatement of associated details in neocortex, which itself was associated with superior memory behavior. Although adults also showed evidence of hippocampal differentiation—albeit in a more focal region than observed in children—it did not benefit their neocortical reinstatement (and therefore memory performance). Instead, we found a beneficial relationship between formation of integrated memory traces in prefrontal cortex and neocortical reinstatement. Together, these findings align well with current perspectives of neural development which suggest that differences in the maturational pace of the hippocampus and/or prefrontal cortex—and the neural representations they support—contributes to the shift in balance between memory specificity and generalization[9]. Notably, the findings further suggest that children and adults make mnemonic decisions using different types of memory 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[19], the present results suggest that learning experiences that reduce the overlap among related information may aid learning and memory in educationally relevant ways.

Online 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 statistically below chance [50%] performance 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 triad. 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. 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 triads (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., $face_1$). All stimuli were sized to 400x400 pixels.

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 discarded 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. Importantly, the purpose of the practice tasks at Session 1 was to ensure that participants could perform the tasks. 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.

MRI scanning (Session 2). Following informed consent and completion of the same behavioral practice tasks as those completed at Session 1, participants completed the following four experimental phases in the MRI environment: pre-exposure item localizer (arrow detection task), pair learning (initial learning/retrieval task), memory recall (match/mismatch decision task), and category localizer (repeat detection task).

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 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 (Fig. 1a). Participants performed several consecutive learning-retrieval blocks to a 90% accuracy criterion, which were conducted outside of the fMRI 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 (Fig. 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 (Fig. 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 (see Fig. 1c and Supplementary data: Age-related learning differences).

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; Fig. 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.

To ensure that participants successfully retrieved the associative memories in the MRI scanner, the recall task further included a delayed match to memory decision (Fig. 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 due to below-chance performance).

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: Classifier validation accuracy), 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 (Fig. 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. Two parallel sets of models were performed, one focused on the object cues and one focused on the delay period. For both 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 (Fig. 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 model to quantify the degree of ensuing neocortical item (Fig. 4a) and category (Fig. 3b) reinstatement following the object cue. To this end, we extracted neural patterns for each of the individual 9-second delay periods—during which only a fixation dot was presented—each modeled as a 9s regressor and convolved with the canonical double gamma HRF. 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 36 retrieval-related delay patterns 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 (Fig. 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 (Fig. 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 (Fig. 3a) using PyMVPA (Hanke et al., 2009) and custom Python scripts.

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-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: Classifier validation accuracy).

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 (Fig. 3b). To this end, the same classifier as above was used, except rather than the leave-one-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 9-second 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 retrieval pattern, 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 (one number for face evidence; one number for scene evidence; summing to 1) 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) and to each other (independent samples t-test). 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, here and for all group-level neural comparisons (see item reinstatement and overlapping memory representation), analyses were restricted to trials in which retrieval was successful at both final learning and scanned memory recall.

Participant-level 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 Fig. 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, restricted to correct retrieval and final pair learning trials (Supplementary Table 1 for trial counts).

To assess the effect of age and relatedness on hippocampal memory representations (Fig. 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 with Bonferroni-adjusted p-values to control for multiple comparisons.

Participant-level 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 (Fig. 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 Fig. 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), yielding an estimate of whether item reinstatement was reliably evidenced on successful trials in each fROI (Fig. 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), separately for each fROI (Fig. 4b).

Participant-level 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 (though still constrained to final correct learning trials), and were directly entered into the within-subject mixed effect and drift diffusion models (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 match/mismatch 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.

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 tapplied 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.

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 2).

Item reinstatement. A searchlight analysis extension of the main item reinstatement analysis (Fig. 4b) was conducted to identify localized voxels that exhibited significant item-level reinstatement at the whole-brain level (Supplementary Table 3), which was then small-volume corrected within the fROIs (VTC, parietal; Supplementary Table 3) as well as within the anatomical hippocampal mask (Supplementary Figure 3). With respect to the fROIs, this complementary approach allowed us to ask whether item reinstatement—which was only reliably observed in adults in VTC (Fig. 4b)—could be detected within relatively more selective regions nested within those engaged by stimulus category as a whole (i.e., the full fROIs; Fig. 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 (Fig. 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

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). 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.

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. To test this question, we quantified model fits for a model including age group, as compared to a base model that did not include age as a predictor. Because age group explained significant variance in behavioral performance (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 (log 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 nuisance 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. 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 (Fig. 2b) and integration in adults (Fig. 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 learning and retrieval trials. Given the continuous nature of the dependent measure, here we employed four separate linear mixed effects model (lmer function) assessing the trial-wise association between hippocampal (children)/mPFC

(adult) cue similarity (i.e., the raw pattern similarity for related cue items) and neocortical reinstatement, separately for parietal cortex item reinstatement (same-item similarity) and VTC category reinstatement (log transformed classifier evidence for target category). As above, individual model predictors were scaled and centered across all trials within a participant prior to analyses. Model comparisons were not conducted, as the child and adult models were conducted separately.

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/mismatch 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. Guided by the significant reinstatement effects on decision accuracy (Fig. 5a), here we focused on trial-by-trial evidence of category-level reinstatement within the VTC fROI (Fig. 3b; target reinstatement) as well as item-level reinstatement within the parietal cortex fROI (Fig. 4a; raw same item estimates). We modeled the linear effect between VTC category reinstatement and parietal item reinstatement on the drift rate parameter. We additionally controlled for mean hippocampal BOLD response on each trial. The models 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 20,000 MCMC samples, after discarding the first 10,000 samples for burn-in, resulting in a distribution of regression coefficients.

To evaluate statistical significance among the neural predictors and the drift rate parameter, we computed the proportion regression coefficients that were less than 0, 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/mismatch 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.

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. (2021). We know more than we ever learned: Processes involved in accumulation of world knowledge. *Child development perspectives* 15.4: 220-227.
8. 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.
9. 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.
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. 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.
12. Ngo, C. T., Newcombe, N. S., & Olson, I. R. (2018). The ontogeny of relational memory and pattern separation. *Developmental science*, *21*(2), e12556.
13. Rollins, L., & Cloude, E. B. (2018). Development of mnemonic discrimination during childhood. *Learning & Memory*, *25*(6), 294-297.
14. 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.
15. 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.
16. 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.
17. Wilson, J. T., & Bauer, P. J. (2021). Prompt-facilitated learning: The development of unprompted memory integration and subsequent self-derivation. *Memory & Cognition*, *49*, 1473-1487.

18. 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.
19. 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.
20. Riggins, T., Canada, K. L., & Botdorf, M. (2020). Empirical evidence supporting neural contributions to episodic memory development in early childhood: implications for childhood amnesia. *Child development perspectives*, *14*(1), 41-48.
21. 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.
22. 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.
23. 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.
24. 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.
25. 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.
26. 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.
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. Dickerson, B. C., & Eichenbaum, H. (2010). The episodic memory system: neurocircuitry and disorders. *Neuropsychopharmacology*, *35*(1), 86-104.
29. Marr, D., 1971. Simplememory: a theory for archicortex. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 262, 23–81.
30. 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.
31. 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.
32. 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.
33. Bauer, P. J., Burch, M. M., Scholin, S. E., & Güler, O. E. (2007). Using cue words to investigate the distribution of autobiographical memories in childhood. *Psychological Science*, *18*(10), 910-916.

34. Ghetti, S., & Angelini, L. (2008). The development of recollection and familiarity in childhood and adolescence: Evidence from the dual-process signal detection model. *Child development*, *79*(2), 339-358.
35. Riggins, T. (2014). Longitudinal investigation of source memory reveals different developmental trajectories for item memory and binding. *Developmental Psychology*, *50*(2), 449.
36. Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 4.
37. 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.
38. 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.
39. 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.
40. Karmiloff-Smith, A. (1986). From meta-processes to conscious access: Evidence from children's metalinguistic and repair data. *Cognition*, *23*(2), 95-147.
41. Darby, K. P., & Sloutsky, V. M. (2015). The cost of learning: interference effects in memory development. *Journal of Experimental Psychology: General*, *144*(2), 410.
42. Preston, A. R., & Eichenbaum, H. (2013). Interplay of hippocampus and prefrontal cortex in memory. *Current biology*, *23*(17), R764-R773.
43. Ghetti, S., DeMaster, D. M., Yonelinas, A. P., & Bunge, S. A. (2010). Developmental differences in medial temporal lobe function during memory encoding. *Journal of Neuroscience*, *30*(28), 9548-9556.
44. Bauer, P. J., & Larkina, M. (2014). The onset of childhood amnesia in childhood: A prospective investigation of the course and determinants of forgetting of early-life events. *Memory*, *22*(8), 907-924.
45. Ghetti, S., Lee, J. K., Sims, C. E., DeMaster, D. M., & Glaser, N. S. (2010). Diabetic ketoacidosis and memory dysfunction in children with type 1 diabetes. *The Journal of pediatrics*, *156*(1), 109-114.
46. 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.
47. Kuhl, B. A., Shah, A. T., DuBrow, S., & Wagner, A. D. (2010). Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nature neuroscience*, *13*(4), 501-506.
48. 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.
49. 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.
50. DeMaster, D. M., & Ghetti, S. (2013). Developmental differences in hippocampal and cortical contributions to episodic retrieval. *Cortex*, *49*(6), 1482-1493.

Supplementary Information

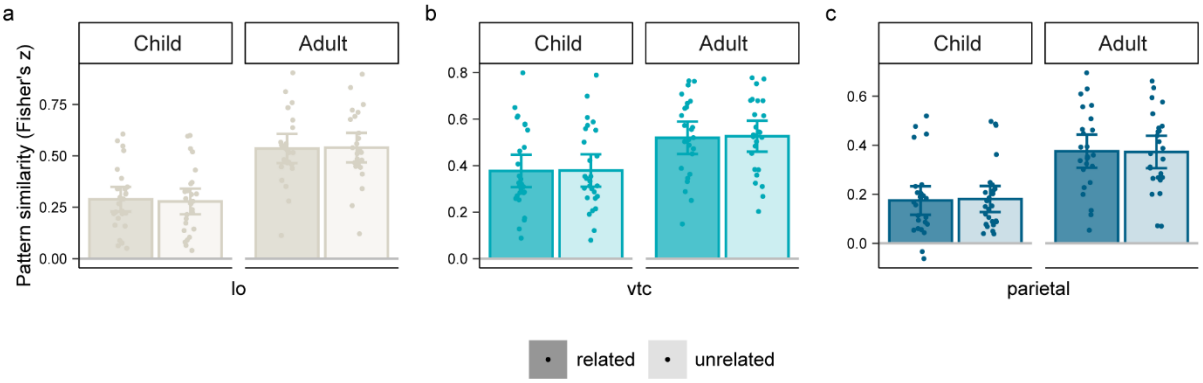
Age-related learning differences. Relative to adults, children required more learning exposures to reach the 90% retrieval criterion (Figure 1C; Adult: *M* number exposures = 1.24, *SD* = 0.44; Child: *M* number exposures = 1.85, *SD* = 0.95; $t(50) = 2.95$, $p = .005$). More specifically, 76% of adults ($N = 19$) reached the 90% accuracy criterion on the first learning block with the remaining adults ($N = 6$) reaching the criterion by the second learning block. In contrast, 37% of children ($N = 10$) reached the accuracy criterion on the first learning block, 52% of children ($N = 14$) on the second learning block, and 11% of children ($N = 3$) required more than two learning blocks (3 runs*: N* = 1*, Age* = 9.08*;* 4 runs: *N* = 1*; Age* = 7.83*;* 5 runs*: N* = 1*; Age* = 7.78*).* Children achieved a similar level of final learning performance to adults, albeit across more repetitions (Children: *Accuracy [Proportion correct]* = 0.96, *SD* = 0.03; Adults: *Accuracy [Proportion correct]* = 0.99, *SD* = 0.02). Nevertheless, performance on the final learning run was still statistically higher in adults, $t(45.05) = 4.24$, $p < .001$. We controlled for this group difference in all further analyses by restricting neural analyses to trials on which the corresponding associative learning pair was successfully retrieved on the final learning run.

In addition to comparing learning performance between age groups, we also tested for age-related performance differences within the child group alone. More specifically, we tested for differences between younger (7- & 8-year-olds; $N = 14$) and older (9- & 10-year-olds, $N = 13$) children, which resulted in nonsignificant differences in initial learning (Younger: *Accuracy* = 0.83, *SD* = .10; Older: *Accuracy* = .82, *SD* = .18; $t(18.69) = .19$, $p = .85$), final learning (Younger: *Accuracy* = 0.96, *SD* = .03; Older: *Accuracy* = .97, *SD* = .03; $t(25) = -.56$, $p = .56$), and the number of runs required to reach the 90% learning criterion (Younger: *M number exposures* = 2.07, *SD* = 1.14; Older: *M number exposures* = 1.62, *SD* = .65; $t(25) = 1.26$, $p = .22$).

Hippocampal differentiation controlling for univariate activation. The overall difference in hippocampal pattern similarity between related and unrelated object cues observed in children (Fig. 2b) remained when mean hippocampal univariate activation during the cue period was entered as a covariate into the analysis of variance, $F(1,25) = 4.78$, $p = .04$, partial $\eta^2 = .16$. This finding suggests that the decreased representational similarity observed for related relative unrelated memories cannot be attributed to broad changes in mean hippocampal activation in response to the retrieval cue, but rather are the result of differences in the distributed patterns of activation across hippocampal voxels.

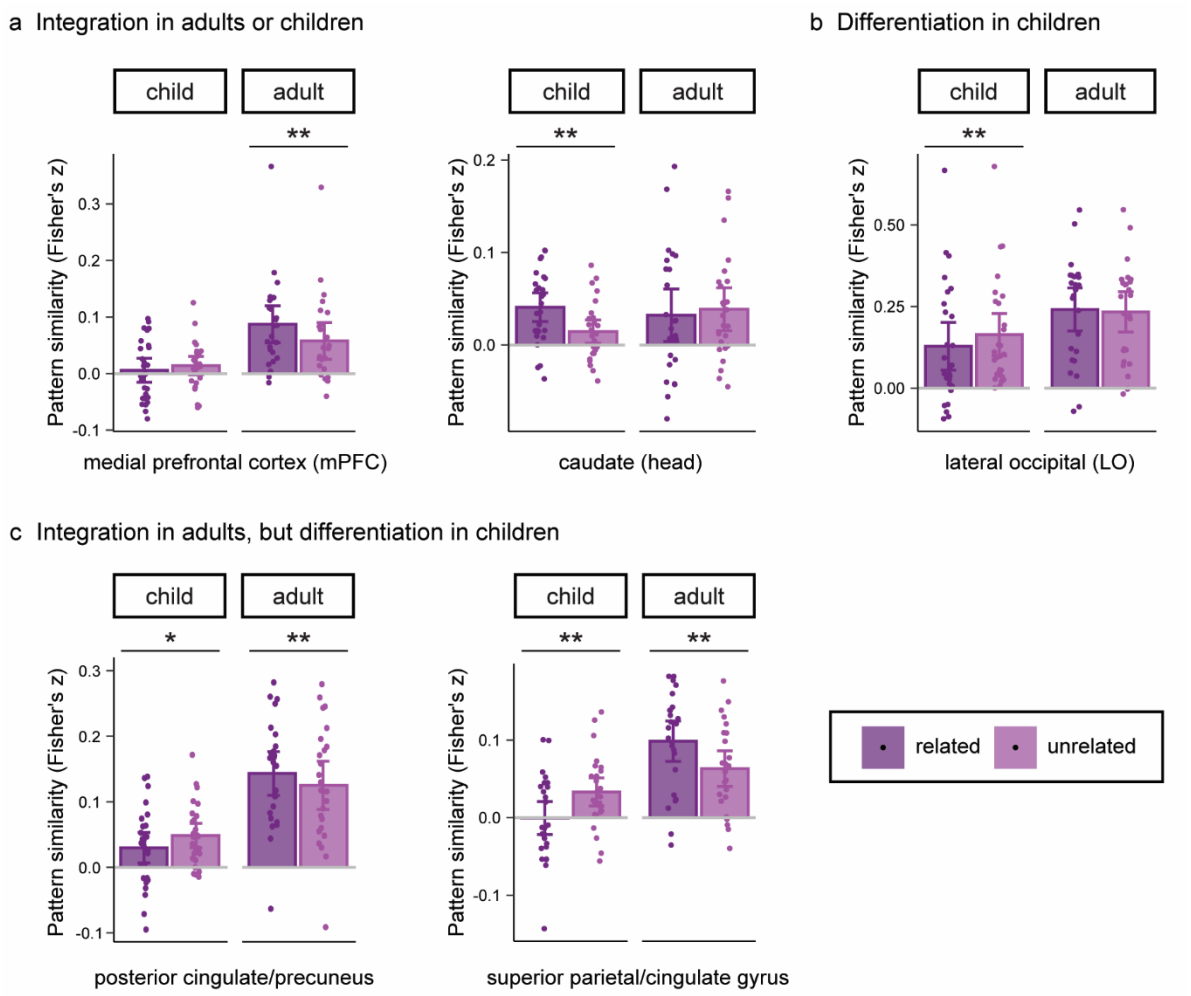
Classifier validation accuracy. To quantify the degree to which the target category was reinstated during the delay period, we used the category localizer to train an multivoxel pattern analysis (MVPA) classifier to identify patterns of activity associated with face and scene processing in the ventral temporal cortex (VTC) and parietal functional regions of interest (fROIs). Ensuring that the MVPA classifier could accurately detect when participants were viewing faces and scenes in the category localizer task was a necessary first step before testing our key question, which was whether face and scene patterns were reinstated during the retrieval task. Cross-validation performance was significantly above chance levels (Chance = 0.5; VTC: $M = 0.97$, $SD = 0.02$, $t(51) = 135.16$, $p < .001$, all participants above chance: ps < 1.50×10^{-15}; parietal: $M = 0.85$, $SD = 0.08$, $t(51) = 32.74$, $p < .001$, all participants significantly above chance: ps < 7.70×10^{-05}) and did not differ as a function of age (VTC: $t(50) = 1.21$, $p = .23$; parietal: $t(50) = 1.38$, $p = .17$). Together, this pattern of results suggests that face and scene patterns were discriminable for each participant and age group in each fROI.

Supplementary Figure 1. *Analyses of the degree of similarity among neural activation patterns for related relative to unrelated object cues revealed no evidence of systematic neural organization in perceptual regions, neither through differentiation nor integration among children or adults.*



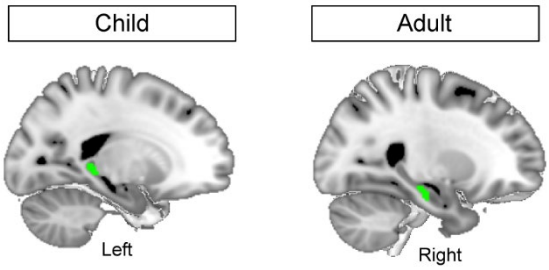
(a) Neural patterns extracted from bilateral lateral occipital (LO) cortex, derived from an automated anatomical mask using Freesurfer segmentation. There was neither a main effect of similarity ($F(1,50) = 1.25$, $p = 0.27$, partial $\eta^2 = .02$) nor an interaction with age ($F(1,50) = 2.27$, $p = 0.14$, partial $\eta^2 = .04$). Given that significant hippocampal differentiation was only observed in the child group, we also examined the effect of similarity in this group alone. There were no significant difference between related and unrelated object cues ($F(1,26) = 2.43$, $p = 0.13$, partial $\eta^2 = .09$), suggesting that the primary hippocampal differentiation effects cannot be attributed to neural similarity of the visual objects themselves, but rather result from mnemonic coding schemes. (b) Neural patterns extracted from functionally-defined ventral temporal cortex (VTC; see main Fig. 3a and supplemental method). There was neither a main effect of similarity ($F(1,50) = 0.46$, $p = 0.50$, partial $\eta^2 = .01$) nor an interaction with age ($F(1,50) = 0.89$, $p = 0.35$, partial $\eta^2 = .02$). There were also no significant difference between related and unrelated object cues when we examined children alone, ($F(1,26) = 0.03$, $p = 0.87$, partial $\eta^2 < .001$). (c) Neural patterns extracted from functionally-defined parietal cortex (see main Fig. 3a and supplemental method). There was neither a main effect of similarity ($F(1,50) = 0.32$, $p = 0.58$, partial $\eta^2 = .01$) nor an interaction with age ($F(1,50) = 0.62$, $p = 0.43$, partial $\eta^2 = .01$). There were also no significant difference between related and unrelated object cues when we examined children alone, ($F(1,26) = 0.69$, $p = 0.41$, partial $\eta^2 = .03$).

Supplementary Figure 2. *Whole-brain RSA searchlight regions (from Supplementary Table 2) showing significant developmental differences in either (a) integration, (b) differentiation, or (c) both integration and differentiation simultaneously.*



We compared the similarity of fMRI patterns for related as compared with unrelated objects across trials in regions for which there was an age-related difference (adult > child or child > adult; Supplementary Table 2) in differentiation (related < unrelated) or integration (related > unrelated). Error bars reflect 95% confidence intervals around the group mean. Dots reflect individual participant means. Asterisks reflect a significant within-age effect at a threshold of * $p < .05$ or ** $p < .001$.

Supplementary Figure 3. *Children and adults reinstate target items in different subregions of hippocampus.*



Results of item reinstatement representational similarity analysis approach (Fig. 4a), which was implemented at the individual voxel level via separate searchlight analyses in children (child > 0) and adults (adult > 0). For each participant and voxel, we computed a single item reinstatement index by averaging the mean same item – different item similarity difference across recall trials. We observed that children reinstated specific memory associates in posterior hippocampus (*x*, *y*, *z* = -18.7, -37.5, -0.67), whereas adults exhibited specific item representations in anterior hippocampus (*x*, *y*, *z* = 28.3, -20.7, -17.2; Fig. 4c). This mirrors the general pattern that has been observed previously in terms of locus of univariate activation shifting from posterior to anterior hippocampus across development during associative retrieval[21]. Significant item reinstatement clusters are displayed on the 1-mm MNI template brain. Clusters were identified on a whole-brain template (see Supplementary Table 3 for additional regions) and are significant after small-volume correction for multiple comparisons within anatomical hippocampus.

Supplementary Table 1. Mean and range of number of correct trials for final learning, scanned recall, and for the overall neural analyses after accounting for learning and scanned recall, separately for each age group.

	Learning/Retrieval		Scanned Recall		Overall	
	Mean (SD)	*Range*	*Mean (SD)*	*Range*	*Mean (SD)*	*Range*
Child	34.67 (1.07)	33-36	30.96 (3.26)	25-35	29.96 (2.90)	25-35
Adult	35.68 (0.69)	33-36	35.32 (0.95)	33-36	35.08 (1.32)	31-36

Note: The minimum number of trials represented in the table (25) corresponds with the behavioral threshold for above-chance (0.5), based on a Binomial Test ($p<.05$; two-tailed).

Supplementary Table 2. Whole-brain RSA cue differentiation/integration searchlight results. Regions reflect clusters of voxels that survived cluster correction. For clusters showing age interactions (Adult > Child or Child > Adult), effects were considered significant if cue differentiation/integration was reliable at the within-group level in the group that showed the enhanced pattern (see Supplementary Figure 2 for within-group plots in each region). Values (X, Y, Z) reflect cluster center of gravity (COG).

	Region	*Hemisphere*	*N vox*	*t*	*X*	*Y*	*Z*
Differentiation	***Main Effect***						
	Cuneus	R	164	1.84	13	-81	39
	Anterior parahippocampal gyrus	L	120	2.37	-29	-5	-41
	Insula cortex	R	118	3.51	34	-7	3
	Subgenual anterior cingulate cortex	L	99	3.54	-6	6	-9
	Adult > Child						
	None	-	-	-	-	-	-
	Child > Adult						
	Posterior cingulate/precuneus[c]	R	568	2.91	8	-43	6
	Superior parietal lobule/cingulate gyrus[c]	R	482	1.54	31	-34	57
	Lateral occipital cortex[b]	R	124	2.77	56	-63	1
Integration	***Main Effect***						
	None	-	-	-	-	-	-
	Adult > Child						
	Posterior cingulate/precuneus[c]	R	568	2.91	8	-43	6
	Superior parietal lobule/cingulate gyrus[c]	R	482	1.54	31	-34	57
	Medial prefrontal cortex (mPFC)[a]	L	188	4.14	-15	45	-4
	Child > Adult						
	Caudate (head)[a]	L	142	3.61	-11	21	3

Note: 2 mm isotropic voxels. MNI coordinates rounded to the nearest mm. Age interaction patterns: [a]integration in one group, null pattern in other group, [b]differentiation in one group, null pattern in other group, [c]opposing neural organizational schemes in each group (Supplementary Figure 2 for corresponding within-group plots).

Supplementary Table 3. Whole-brain RSA item reinstatement searchlight results. Regions reflect clusters of voxels that survived cluster correction within each age group separately. Values (X, Y, Z) reflect cluster center of gravity (COG).

Region	*Hemisphere*	*N vox*	*t*	*X*	*Y*	*Z*
Child						
Occipital pole	L	112	1.71	-11	-100	11
Intraparietal sulcus/inferior parietal lobule†	R	61	3.29	48	-38	44
Fusiform*	R	39	3.28	28	-67	-9
Parahippocampal cortex*	R	23	1.99	34	-39	-12
Adult						
Occipital lobe**	B	1064	2.86	2	-92	4
Dorsomedial prefrontal ctx/ant. cingulate	B	370	2.58	-2	42	24
Posterior cingulate/medial occipital	B	275	2.36	-1	-55	17
Medial prefrontal cortex (mPFC)	B	192	2.4	2	53	-5
Intraparietal sulcus††	R	135	2.96	32	-51	44
Anterior hippocampus/MTL cortex	R	111	1.82	34	-23	-20
Angular gyrus	L	109	2.53	-49	-51	20
Posterior parahippocampal/fusiform gyrus**	R	101	2.07	26	-41	-12
Fusiform gyrus**	L	93	5.28	-23	-68	-8
Inferior frontal gyrus/Insula	R	91	2.15	38	24	-5
Lingual/fusiform gyrus**	L	89	2.57	-23	-48	-10
Superior parietal lobule	L	88	3.85	-47	-37	57
Frontal pole	L	82	2.09	-27	54	11
Lateral occipital*	L	63	2.95	-40	-63	-6
Lateral occipital*	R	49	3.98	40	-81	-16

*Note: 2 mm isotropic voxels. MNI coordinates rounded to the nearest mm. *denotes significant after small volume correction in VTC; **denotes significant in whole-brain and VTC small-volume corrected analyses; †denotes significant after small volume correction in parietal cortex; ††denotes significant in whole-brain and parietal small-volume corrected analyses.*