Article:

Ngo, Chi T, Horner, Aidan James orcid.org/0000-0003-0882-9756, Newcombe, Nora S et al. (1 more author) (Accepted: 2019) Development of holistic episodic recollection. Psychological Science. ISSN 1467-9280 (In Press)

Development of Holistic Episodic Recollection

Chi T. Ngo[1], Aidan J. Horner[2,3], Nora S. Newcombe[1], & Ingrid R. Olson[1]

[1] Department of Psychology, Temple University, USA

[2] Department of Psychology, University of York, UK

[3] York Biomedical Research Institute, University of York, UK

Correspondence concerning this article should be addressed to Chi T. Ngo, Department of Psychology, Temple University, 1701 N. 13th Street, Philadelphia, PA 19122. Email: chi.ngo@temple.edu

Abstract

Episodic memory binds together the diverse elements of an event into a coherent representation. This coherence allows for the reconstruction of different aspects of an experience when triggered by a cue related to a past event—a process of pattern completion. Such holistic recollection is evident in young adults, as shown by dependency in the retrieval success for various associations from the same event (Horner & Burgess, 2013, 2014). In addition, episodic memory shows clear quantitative increases during early childhood. However, the ontogeny of holistic recollection is uncharted. Using dependency analyses, we found that 4-year-olds (n=32), 6-year-olds (n=30) and young adults (n=31) all retrieve complex events in a holistic manner, i.e., retrieval accuracy for one aspect of an event predicted accuracy for other aspects of the same event. However, the degree of holistic retrieval increased from age 4 to adulthood. Thus, extended refinement of multi-way binding may be one aspect of episodic memory development.

Keywords: holistic recollection, memory development, pattern completion, episodic memory.

Episodic memory is the capacity to remember the unique combinations of people, objects, and places that make up the specific events of our past (Tulving, 2002). Episodic memory is thought to be stored as an integrated representation enabling holistic retrieval, such that remembering one constituent of an event can elicit the retrieval of other elements from the same event. For instance, if retrieving a place successfully reminds us of the person we met there, it would also likely evoke our memory of what objects we encountered in that event. Longstanding computational models of memory posit that the hippocampus supports this neural computation, termed pattern completion, which reinstates a complete event representation in the presence of a partial cue via reactivation (Marr, 1971; McClelland, McNaughton, & O'Reilly, 1995; Norman & O'Reilly, 2003). According to these models, an exposure to part of a past experience leads to recurrent connectivity in hippocampal CA3 subfield retrieving the conjunctive representation of the entire event (Guzowski, Knierim, & Moser, 2004; Rolls, 2016).

Several distinct conceptualizations of pattern completion and paradigms for assessing it exist in the literature (e.g., Horner & Burgess, 2013; Vieweg et al., 2015; Yassa & Stark, 2011). One paradigm relies on the conceptualization that pattern completion allows for the recovery of the entire event based on a partial cue so that all elements within an event can be retrieved successfully (or not). Using multi-element event paradigms, Horner and Burgess (2013, 2014) demonstrated that young adults indeed retrieve events in a holistic fashion. In these tasks, participants first learn a series of unique events, each of which is comprised of a scene, a person, and an object. Subsequently, participants perform a cued recognition task on every possible cue-test pair of each event. Successful retrieval of one association (e.g., retrieving a person when cued with a scene) was statistically related to the retrieval success of other associations (e.g., retrieving an object when cued with a scene) from the same event. Further, during such cued retrieval, the hippocampus showed signatures of neural reinstatement of another within-event element that was irrelevant to the given test trial (Horner et al., 2015; reviewed in Horner & Doeller 2017). Corroborating these findings, a recent study showed that retrieval not only improved long-term memory for the specific information tested, but also non-targeted information that shared the same spatial context (Jonker, Dimsdale-Zucker, Ritchey, Clarke, & Ranganath, 2018). These results demonstrate that memories are reorganized into integrated events, and provide compelling evidence that the retrieval of multiple elements of an event is mutually contingent. This process may, at least in part, rely on hippocampal pattern completion.

Although some episodic-like memory capacities emerge early in development (reviewed in Bauer, Larkina, & Deocampo, 2010), several aspects of episodic memory are still far from mature in preschool-aged children (reviewed in Olson & Newcombe, 2014). Evidence from research on episodic memory development employing laboratory-based paradigms suggests children's relational memory improves robustly between the ages of 4 and 6, as shown in various tasks using inter-item (e.g., bear-book), item-context (e.g., bear-library; Lloyd, Doydum, & Newcombe, 2009; Sluzenski, Newcombe, & Kovacs, 2006), or item-item-context associations (e.g., bear-book-red house; Ngo, Newcombe, & Olson, 2017; Ngo, Lin, Newcombe, & Olson, 2019; Yim, Dennis, & Sloutsky, 2013). When asked to remember unique item-context pairs (e.g., bear-library), 6-year-olds outperformed 4-year-olds in relational memory (e.g., Lloyd et al., 2009; Sluzenski et al., 2006).

A limitation of past research is the focus on relational memory of individual pairs of items, rather than the integration of the multiple associations that constitute a complex event. Holistic retrieval via pattern completion is more than relational binding. It adds the idea of contingency of retrieval success, based on encoding events as an interwoven network of elements. Given that the hippocampus undergoes protracted development (e.g., Keresztes et al., 2017; Lee et al., 2014), it is likely that pattern completion sub-serving holistic recollection also follows prolonged maturation between early years of life and young adulthood.

The current research specifically focuses on coherence of within-event retrieval as opposed to accuracy of pairwise associative memory. In Experiment 1, we assessed the coherence of within-event retrieval in 4-year-olds, 6-year-olds, and young adults by adapting the multi-element event task for children. The verbal experimental materials were changed to pictorial and child-friendly stimuli (e.g., a cartoon of Shrek). Participants first learned multi-element events, each of which contained a scene, a person, and an object. In a cued recognition test, each element in turn served as cue or as the retrieval item. Pattern completion was indexed by retrieval dependency—the degree to which the accuracy for within-event test trials is mutually contingent (all accurate or inaccurate). Evidence for retrieval dependency was seen in all three age groups, with greater dependency in the adults relative to 4-year-olds. In Experiment 2, we further probed the relationship between pairwise associative recognition and dependency by testing a separate group of 6-year-olds with a different task protocol that resulted in lower pairwise associative recognition memory performance. This experiment revealed that retrieval dependency among the 6-year-olds remained comparable to levels seen in Experiment 1, even when accuracy was dampened.

Experiment 1

Methods

Participants and Power Analysis

In order to ensure that we would have sufficient power (0.80) to detect an age* joint proportion of retrieval indices interaction, we conducted an a priori power analysis of a repeated 3(age) x 2 (proportion of joint retrieval indices: data, independent model) mixed ANOVA using G*power v3.1. The power analysis determined a total sample size of 81 (27 participants/age cohort) with sufficient power (0.80) to detect a medium effect size (0.25). The correlation coefficient between the repeated measures were set at 0 as this value was unknown. We slightly oversampled because we had anticipated that some younger participants might fail to complete the task.

A total of 32 4-year-old (15F, M_{month}= 52.06; SD= 3.37) and 30 6-year-old (17F, M_{month} =76.37; SD= 2.16) children from the Philadelphia suburban areas participated in the study at the Temple Ambler Infant and Child Laboratory and schools in the suburban areas of Philadelphia. All children were free of neurological damage and had no history of developmental disorders as reported by a parent. Six additional children (4 4-year-olds and 2 6-year-old) were tested but were not included in data analyses due to experimenter error (n=2) and incomplete procedure (n=4). In addition to the final sample size reported above, two 6-year-old children performed at 100% accuracy, producing ceiling values on all dependent variables, and therefore were excluded from the analyses. All children received a small toy for their participation. The adult sample consisted of 31 undergraduate students (18F, M_{age}= 20.65; SD = 3.23, range=18-31) from Temple University who participated for partial course credit. All participants gave informed consent and reported to have normal or corrected-to-normal vision. This experiment was completed in accordance with, and approved by, the Institutional Review Board committee at Temple University.

Memory Task

Materials.

We sampled 24 cartoon images of distinct scenes (12 indoor, e.g., aquarium, 12 outdoor scenes, e.g., playground), 24 cartoon images of common objects (e.g., watch), and 24 images of cartoon characters from non-overlapping movies/books (12 males, e.g., Pinocchio, and 12 females, e.g., Alice) from Google image search engine. From this pool of selected images, we then constructed 24 "events", each comprised of a scene (e.g., an aquarium) a person (e.g., Alice), and an object (e.g., wallet). The event assignment of the elements was randomized, with the exception that items with pre-experimental associations (e.g., books and library) were not assigned to the same event. Every possible cue-test combination of each event was tested, resulting in 6 test trials ([1] cue: scene—test: person; [2] cue: scene—test: object; [3] cue: person—test: scene; [4] cue: person—test: object; [5] cue: object—test: scene; [6] cue: object—test: person) per event and totaling 144 test trials. All experimental stimuli are publically available at https://osf.io/arphg/.

Procedure.

Children. All participants were tested individually. The task procedure administered to children, or the child task procedure, entailed two encoding—test blocks, which occurred immediately after one another. Each block consisted of 12 encoding and 72 test trials, all presented on a 13' laptop screen. Prior to encoding, participants were told that they would see many different stories and that they should pay close attention to all of the different elements including the scene, person, and object altogether in each story. Then, participants viewed a series of events (12s/each; 0.5 ITI). A short voice-recorded narrative accompanied each event (e.g., "Alice went to the aquarium, but she dropped her wallet there; the wallet was lost in the aquarium") (see Figure 1A). Each narrative consisted of three sentences, with each sentence highlighting one pairwise association within the event. The order of the pairwise associations within each narrative was not fixed or counterbalanced across the events. The implementation of the narrative was to engage children in the task and to increase the likelihood that children would pay attention to all the elements in an event. Prior to encoding, we provided one example (playground, Elastic girl, hat) in order to acquaint the participants with the encoding task.

Immediately after the encoding phase of each block, participants performed a self-paced four-alternative forced choice task. We tested participants on every possible cue-retrieval combination (e.g., cue: scene; test: person) of each studied event, resulting in 6 test trials per event, which totaled 72 test trials per block. On each trial, a cue and four options were presented simultaneously on the screen (see Figure 2A). Among 4 options, one was a target—the correct item because it belonged to the same event as the cue. The

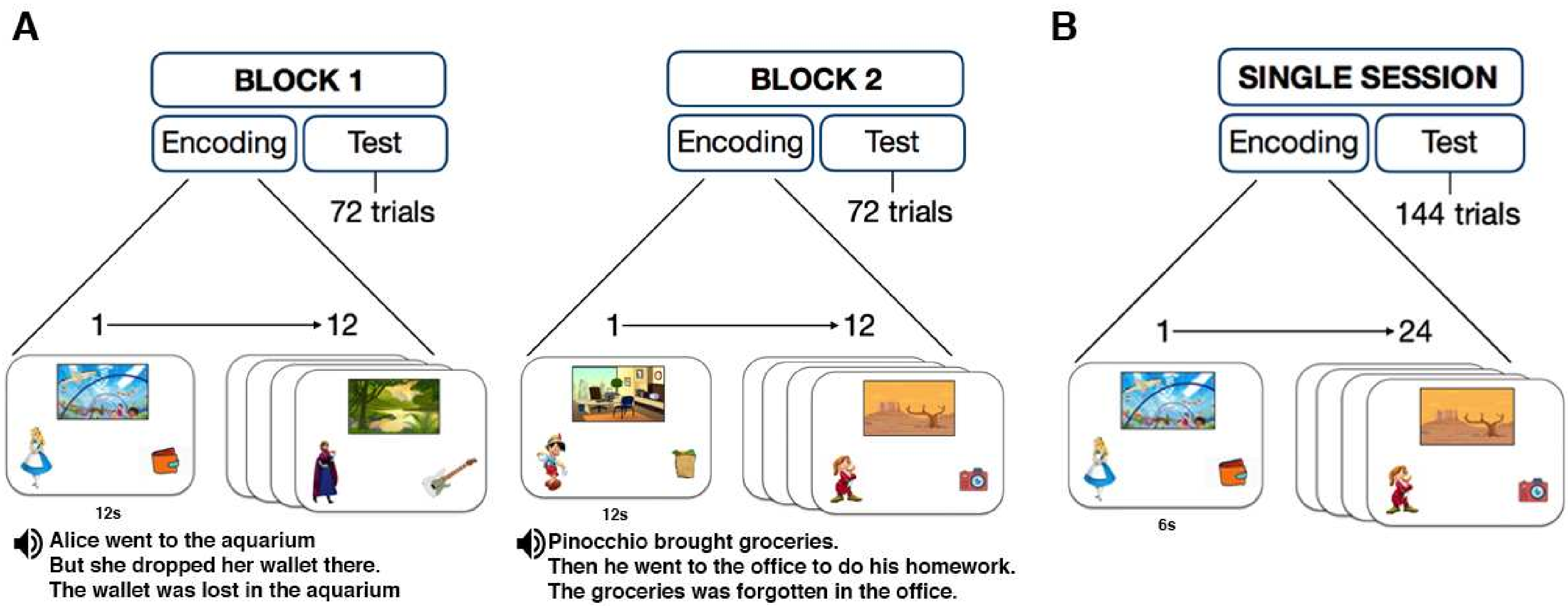

Figure 1. A schematic depiction of the child (A) and adult (B) multi-element event task procedures. In the child task procedure, participants viewed 24 events presented in two encoding-test sessions, each comprised of 12 events. Each event lasted 12s and was accompanied by a voice-recorded narrative. In the adult task procedure, participants studied 24 6-second events together and without the recorded narrative.

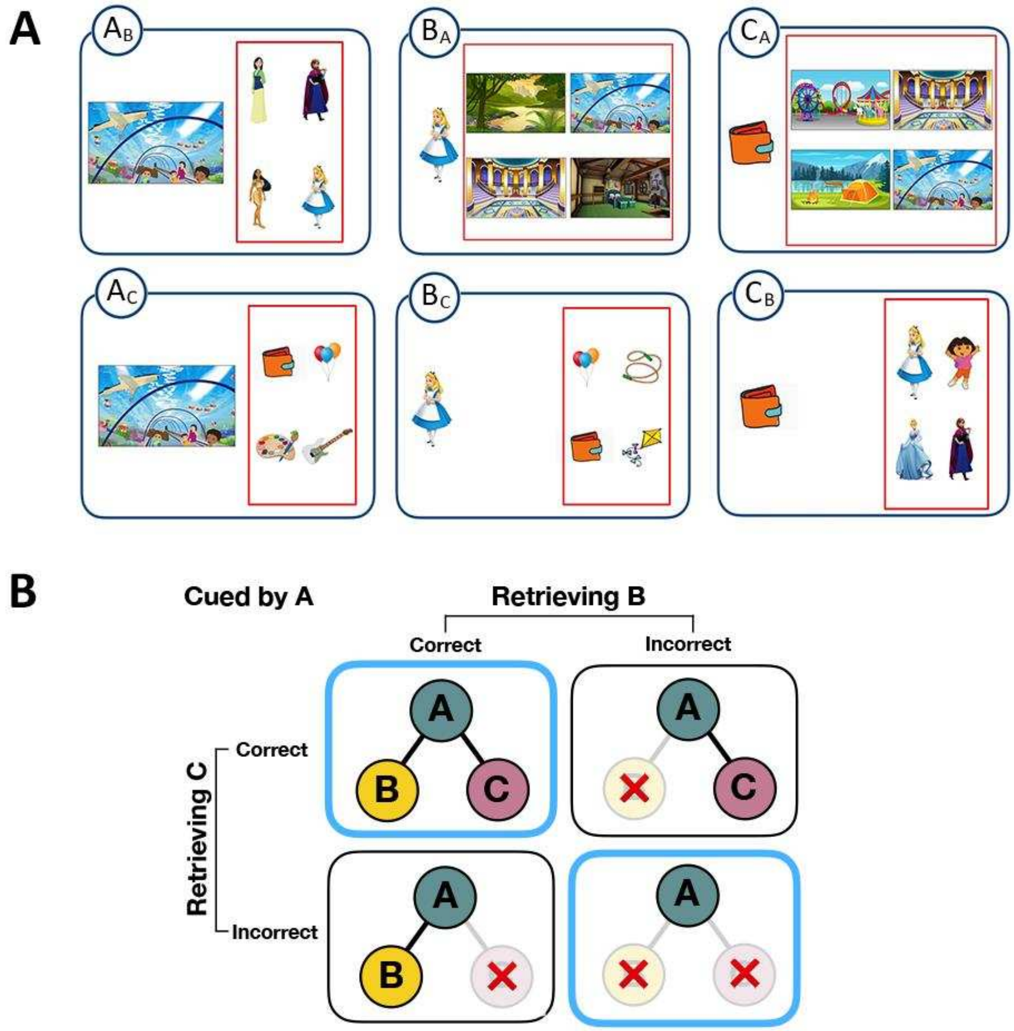

Figure 2. (A) Examples of 6 retrieval types per event in the test phase. Each element of a studied event took turn serving as the cue (item presented on the left side of the screen) and the tested element (one of the four options presented inside the red box). (B) A schematic depiction of how the proportion of joint retrieval for A_BA_C pairs was computed for each participant by concatenating the proportion of events in the blue outlined boxes out of the total number of events.

three lures were same-category elements from different events. The lures always came from the events that contained the same-sex characters, so that participants could not eliminate lures based on general mnemonic heuristics (e.g., remembering that there was a female character who went to the aquarium). Across all 24 events, any given two test trials that had overlapping cue items (e.g., A_B^1 and A_C^1) or tested items (e.g., B_A^1 and C_A^1) only shared 1 foil item (out of 3) with respective to their event membership. For example, for the A_B test trial of event 1, the foils included the B elements from events 2, 3, and 4, whereas for the A_C trial of event 1, the foils included the C elements from events 3, 5, and 7 (one B and one C foil both from event 3). Furthermore, all items served as foils an equal number of times across all 144 test trials. Children were asked to point to one of the four options that belonged to the same story as the cue on the left side of the screen. Positions of the correct answer were counterbalanced across the entire test phase. There were no missing responses as the response time was unrestricted. The memory task took approximately 40 minutes.

Adults. Adults were administered the adult task procedure, which is similar to the child task procedure, but with a few differences. First, the whole procedure was administered in a single encoding (24 events) —test (144 trials) procedure. Second, no narratives were implemented at the encoding phase to avoid potential ceiling performance in young adults. Third, each encoding trial was presented for 6s (see Figure 1B).

Verbal Intelligence

Standardized tests of verbal intelligence were included as a control variable to assess whether verbal intelligence between groups of same-aged children differed (6-year-old children in Experiments 1 and 2).

Children. All children were administered the Kaufman's Brief Intelligence Test, 2nd edition (KBIT-2; Kaufman & Kaufman, 1990) to assess general verbal intelligence. Children were instructed to choose one of the six images simultaneously shown on a page that was the best match for a word or phrase (e.g., what is something that floats and you can ride in — a boat), and to respond with a one-word answer to verbal riddles (e.g., what eats carrots and has long ears? — bunny). The test, with increasing level of difficulty in each section, was terminated when a child provided 4 incorrect responses consecutively. A standard score was calculated for each child based on his/her age.

Adults. Adults were administered the 45-item American National Adult Reading Test (AMNART [Grober & Sliwinski, 1991] —an American version of the National Adult Reading Test [Nelson, 1982]). This test measures the ability to read aloud irregular words. Pronunciation errors were tallied and AMNART-estimated verbal IQ scores were calculated using Grober and Sliwinski's formula, which accounts for years of education.

Estimating retrieval dependency

The retrieval dependency between retrieval successes for different associations within the same event was computed using the same methods as in previous studies (Horner & Burgess, 2013, 2014; Horner et al., 2015; Bisby, Horner, Bush, & Burgess, 2018). Six 2 x 2 contingency tables for the data and the predicted independent model were computed for each participant based on their retrieval accuracy for each pairwise association in order to assess dependency between retrieving two elements when cued by the remaining common element within an event (A_BA_C; i.e., cue with A and retrieve B, and cue with A and retrieve C), and the dependency between retrieving a common item when cued by the other two elements within an event (B_AC_A; i.e., cue with B and retrieve

A, and cue with C and retrieve A). Each 2x2 contingency table for the data, for every participant, shows the proportion of events that fall within the four categories: both A_B and A_C are correct or incorrect, A_B correct and A_C incorrect and vice versa. To examine retrieval dependency, we computed the proportion of joint retrieval for the data—defined as the proportion of events in which both associations were either correctly or incorrectly retrieved (cells 1,1 and 2,2 of each contingency table; see Figure 2B). We then averaged this measure across 6 contingencies tables (three tables for the A_BA_C analysis, for each element-type, and three tables for the B_AC_A analysis, for each element-type) for each participant.

The independent model of retrieval estimates the degree of statistical dependency if retrieval success for specific cue-test pairs (cue: person, test: scene) is independent of retrieval success of other cue-test pairs (cue: person, test: object) in relation to the participants' overall accuracy. The independent model predicts the proportion of joint retrieval, given a participant's overall level of performance, if retrievals of event pairs are independent, such that the probability of the successful retrieval for both (e.g.) A_B and A_C is equal to P_{AB}*P_{AC}, where P_{AB} is the probability of retrieving B when cued by A across all events, and similarly for P_{AC} (see Table 1 for full details). The proportion of joint retrieval for the independent model (calculated in the same manner as described above) serves as a predicted baseline for which we compare the proportion of joint retrieval in the data. Given that the proportion of joint retrieval for the data scales with accuracy, the main index of retrieval dependency was the difference between the proportion of joint retrieval in the data and independent model for each participant—referred to as dependency. If this dependency measure (data – independent model) is significantly greater than zero, this provides evidence for significant retrieval dependency (for the same approach, see Horner & Burgess, 2013; 2014). In addition, we take the magnitude of dependency to signify the extent of holistic retrieval. To probe the development of holistic retrieval, we tested for age effects in dependency—comparing the size of dependency among 4-year-olds, 6-year-olds, and young adults.

Statistical Analyses

All planned statistical analyses were performed using SPSS. Key null findings were tested with Bayesian hypothesis testing using JASP.

Data availability

Second-level data are publicly available through the Open Science Framework at https://osf.io/2pnu6/.

Table 1. Contingency table for the predicted independent model for proportion of correct and incorrect cued recognition over the total number of events for elements B and C when cued by A (P_{AB} denotes the probability of retrieving B when cued by A). The proportion of joint retrieval for the independent model is calculated by summing the main diagonal cells and dividing by the sum across all four cells.

Cued by A		**Retrieving B**	
		Correct	Incorrect
Retrieving C	Correct	$\Sigma_{i=1}^{N} P_{AB} P_{AC}$	$\Sigma_{i=1}^{N} P_{AC}(1 - P_{AB})$
	Incorrect	$\Sigma_{i=1}^{N} P_{AB}(1 - P_{AC})$	$\Sigma_{i=1}^{N} (1 - P_{AB})(1 - P_{AC})$

Results

Overall Accuracy

Overall accuracy is defined as the proportion of target selection across 144 test trials. First, we found no sex differences in any of the three age groups, all p's> .53, all BF_{01}'s > 2.46. Given that the task procedure was different for children and young adults, we compared accuracy between 4- and 6-year-olds. An independent t-test showed that 6-year-olds ($M= 0.82$, $SE= 0.03$) outperformed 4-year-olds ($M= 0.68$, $SE= 0.04$), $t(60) = -2.73$, $p = .01$, 95% CI [-0.25, -0.04], $d= -0.70$, 95% CI [-1.21, -0.18], suggesting age-related improvement in individual pairwise associative memory between the ages of 4 and 6 (see Figure 3). Results on the effects of age and block on accuracy are reported in SI (see SI 1.1).

As for young adults, we found overall accuracy was at 0.72 ($SD= 0.19$). However, we did not perform any statistics comparing overall accuracy between children and adults due to differences in task procedure. Results on the effects of cue and test item types on accuracy for each age group are reported in SI (see SI 1.2).

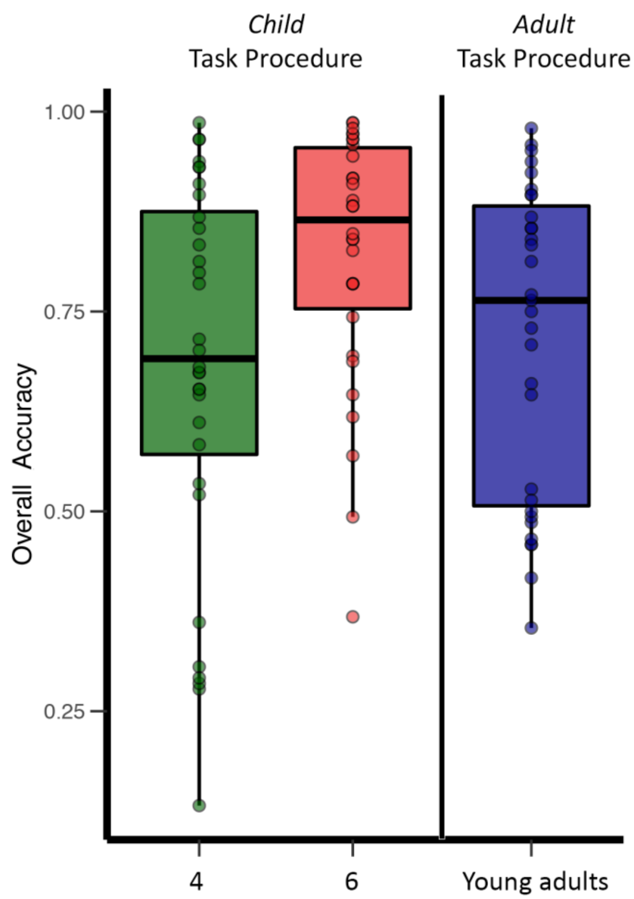

Figure 3. The distribution of the overall accuracy separated by age groups. Note that the overall task procedures differed between children and young adults, such that in children, 24 events were divided into 2 encoding-test sessions (12 events/each session), whereas young adults 24 events were administered in a single encoding-test session.

Retrieval Dependency

The primary questions of this research are whether holistic recollection is evident at the ages of 4 and 6, during a crucial developmental window of robust gains in episodic memory, and whether holistic recollection changes with age. To answer the first question, a one-sample t-test was conducted to test whether dependency (data—independent model) exceeded zero for each age group. As expected, we found that within-event retrieval accuracy was dependent in young adults, such that dependency (M=0.07; SE=0.01) was significantly greater than zero, t(30)= 7.37, p< .001, 95% CI [0.05, 0.08], d= 1.32, 95% CI [0.83, 1.80]. These results conceptually replicate previous studies showing that retrieval dependency is significant in young adults when using verbal stimuli (Horner & Burgess, 2013, 2014) (see Figure 4A).

Interestingly, 4-year-old children also showed significant retrieval dependency, such that dependency (M=0.03; SE=0.01) significantly exceeded zero, t(31)= 4.33, p< .001, 95% CI [0.02, 0.05], d= 0.77, 95% CI [0.37, 1.16]. Six-year-old children also showed significant dependency, in which dependency (M=0.05; SE=0.01) significantly exceeded zero, t(29)= 4.47, p< .001, 95% CI [0.03, 0.07], d= 0.82, 95% CI [0.40, 1.23] (see Figure 5A). Thus, evidence for holistic recollection seen in all three age groups demonstrates that memories for multi-element events may be represented as an integrated episodic unit even in early childhood.

To answer the second question, we tested for age effects on dependency to determine whether the magnitude of retrieval dependency differs across age groups. A one-way ANOVA showed a significant effect of age, F(2, 90)= 3.27, p= .04, *partial* η^2= 0.07. Tukey post-hoc tests showed that dependency was lower in 4-year-olds compared to young adults, t(61)= -2.54, 95% CI [-0.06, -0.002], p= .03, d= 0.69. The 6-year-olds were intermediate; they did not significantly differ from either the 4-year-olds, t(60)= -0.99, 95% CI [-0.04, 0.02], p= .59, d= 0.25, BF_{01}= 2.57, or young adults, t(59)= -1.52, p= 0.29, 95% CI [-0.05, 0.01], d= 0.37, BF_{01}= 1.63 (see Figure 5B). Although the difference in dependency between 6-year-olds and young adults did not reach significance, results from Bayesian statistics were equivocal, supporting neither the null nor the alternative hypotheses. These findings suggest that although dependency is present in all age groups, the degree to which retrieval success (or failure) of one association in a given event relates to other associations from the same event increases between age 4 and young adulthood.

Results on the effects of analysis (same cue vs. same test item) and age on dependency are reported in SI (see SI 1.3).

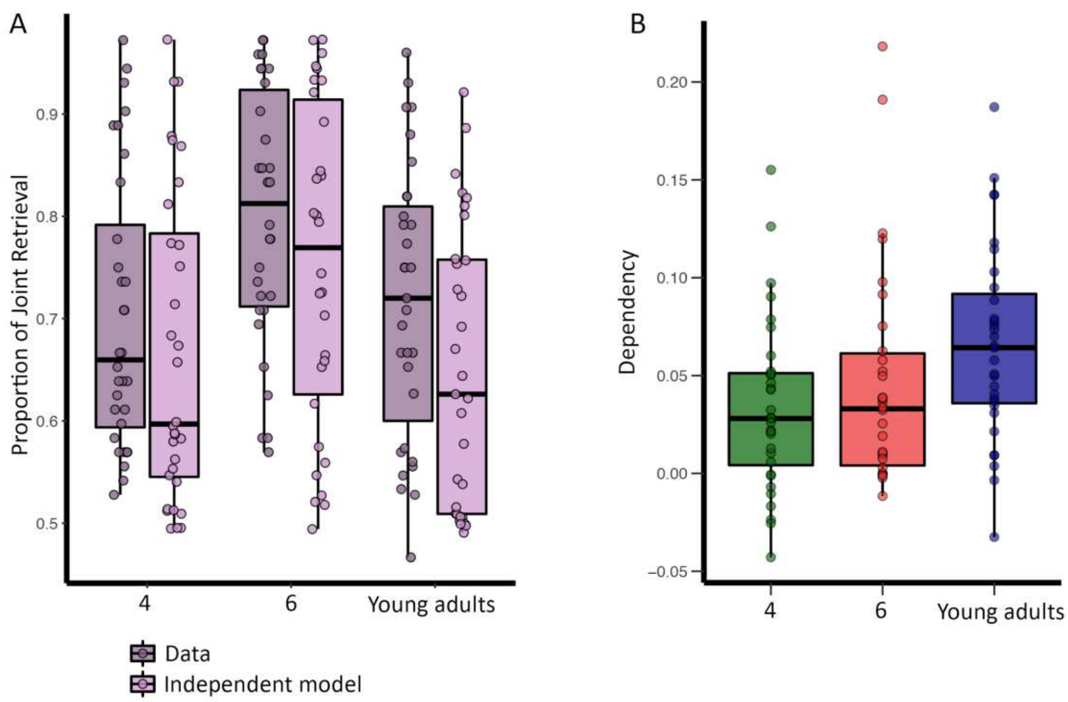

Figure 4. The distribution of the proportion of joint proportion of retrieval for the data and predicted independent model (A) and magnitude of retrieval dependency (B) separated by age groups.

Experiment 2

Overall accuracy was relatively high in the 6-year-old children, with a portion of the children hovering near ceiling-level performance. Thus, we examined whether dependency would be affected if overall accuracy was dampened in this age group. We tested an independent group of 30 6-year-olds (16F, M_{month}= 77.13, SD= 3.20) with the adult task version (same procedure administered in young adults in Experiment 1), in which children learned 24 events without the narratives, and were tested on all 144 test trials in a single encoding-test procedure.

Sex difference in accuracy did not reach significance, t(28)= 1.08, 95% CI [-0.06, 0.21], p= .29, d= .40, 95% CI [-0.33, 1.12], BF_{01}= 1.87. However, results from Bayesian statistics showed that we did not have evidence to support the null hypothesis. It is also worth noting that verbal intelligence did not differ between the two groups of 6-year-olds, t(57)= -0.11, p= .91, 95% CI [-7.96, 7.13], d= -0.03, 95% CI [-0.54, 0.48], BF_{01}= 3.77. As expected, 6-year-old children who performed the adult task procedure had lower overall accuracy compared to their same-aged peers who received the child task procedure, (M= 0.54, SE= 0.03 vs. M= 0.82, SE= 0.03, t(58)= 6.41, p< .001, 95% CI [0.19, 0.37], d= 1.65, 95% [1.06, 2.24] [1] (see Figure 6A). Results on accuracy dependent on cue and test item types for these children are reported in SI (see SI 2.1).

Again, our primary questions concerned dependency. This group of 6-year-olds also showed significant retrieval dependency, dependency (M=0.05; SE=0.01) significantly exceeded zero, t(29)= 4.17, p< .001, 95% CI [0.03, 0.08], d= 0.76, 95% CI [0.35, 1.16]. Critically, dependency did not differ between 6-year-old children in this

---

[1] The same analyses on accuracy (proportion correct) after an arcsine square root transformation yielded the same results and effect size.

experiment and those in Experiment 1, (M= 0.05, SE= 0.01 vs. M= 0.05, SE= 0.01), t(58)= -0.40, p= .69, 95% CI [-0.04, 0.03], d= -0.10, 95% CI [-0.61, 0.41], BF_{01}= 3.57 (see Figure 6B).

We also compared this group of 6-year-olds to young adults, given that they received the same task procedure. Six-year-old had lower overall accuracy relative to young adults, t(59)= -3.66, p< .001, 95% CI [-0.27, -0.08] , d= 0.94 [2]. However, their dependency did not differ significantly from that of adults, t(59)= -0.85, p= .40, 95% CI [-0.04, 0.02] , d= -0.22, 95% CI [-0.72, 0.29], BF_{01}= 2.83.

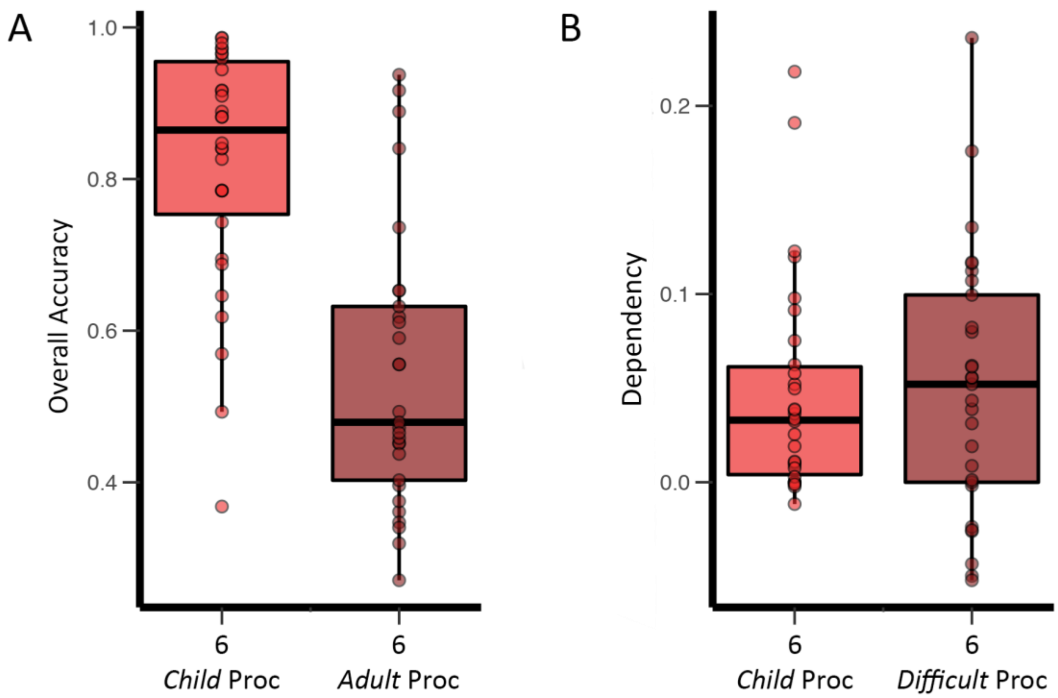

Figure 5. The distribution of the overall accuracy (A) and dependency (B) for the two groups of 6-year-old children in the child and adult task procedures.

Together, these results suggest that retrieval dependency does not simply reflect the overall retrieval accuracy of individual pairwise associative memories. Instead, it specifically assesses the nature of holistic recollection. Across the whole sample, overall accuracy did not scale with dependency, $r(123) = -0.13$, 95% CI [-0.30, 0.04], $p = 0.14$, $BF_{01} = 3.05$. Results on the correlation between dependency and verbal IQ are reported in SI (see SI 3.1).

General Discussion

A defining feature of episodic memory is that complex and multi-modal events are stored as coherent representations, so that episodic retrieval entails the holistic re-experience of all constituents of an event (Tulving, 2002). This work shows that as early as 4 years of age, children are capable of retrieving multi-element events as integrated units. Critically, however, there is a boost in the degree of event memory coherence from age 4 into young adulthood. That is, dependency is greater in adults than in 4-year-old children, whereas dependency in the 6-year-olds is intermediate—it does not differ significantly from either the younger or older participants. These results suggest that pattern completion is present in early childhood and undergoes critical refinements between early childhood and young adulthood.

The dissociation between overall accuracy and retrieval dependency is intriguing. Despite comparatively lower relational memory through preschool than later, young children's episodic memories still possess a significant degree of cohesion. Most studies of episodic memory development have used paired-associate paradigms to probe relational memory. However, memories of specific episodes are made up of an interlinked network of relational structures. By estimating the cohesiveness of event memory as an integrative unit, this work elucidates an important characteristic of how episodic memory refines from early life to adulthood.

In adult humans, one fMRI study used a variant of the multi-element event task in which participants learned all possible individual pairs in a three-element "event" in separate encoding trials (e.g., A—B, B—C, and A—C). Interestingly, hippocampal activity during encoding the final pair predicted memory performance on other pairs of the same event. Furthermore, during cued recognition of pairwise association (e.g., cue A; test B), neocortical activity corresponding to all event elements was reinstated, including the element that was incidental to a given trial (e.g., element C) (Horner et al., 2015). Importantly, the extent of neocortical reinstatement of non-target elements correlated with hippocampal activity at retrieval, consistent with the presence of pattern completion. In light of these results, the reported increased coherence of episodic retrieval from age 4 to adulthood in our work aligns with the findings that intra-hippocampal circuitry has a slow maturational rate (e.g., Lee et al., 2014; Keresztes et al., 2017).

It is worth noting that several paradigms have been employed to investigate the behavioral expression of pattern completion. In some cases, partial cues are defined as fragments of learned scenes (Vieweg, Stangl, Howard, & Wolbers, 2015).), whereas in the multi-element event task, partial cues are defined as elements within complex events. In other paradigms, pattern completion is inferred as the opposite expression of lure discrimination (pattern separation) —a computation that assigns distinct representations, even with a high degree of similarity in the service of reducing interference (Marr, 1971; McClelland et al., 1995). That is, a bias in pattern completion can result in overgeneralization to the detriment of fine-grained lure discrimination, causing interference among similar experiences (e.g., Mnemonic Similarity Task, reviewed in Yassa & Stark, 2011). However, there is evidence that casts doubt on the assumption that lure discrimination and false alarm rates on the MST actually index pattern separation and pattern completion, respectively (e.g., see Molitor, Ko, Hussey, & Ally, 2014). One study that used the MST operationalized pattern completion as the ability to identify lures with low levels of similarity to targets (i.e., degraded input) (Rollins & Cloude, 2018). In this study, younger children (ages 5-6, 8-9) were less able to identify lures that were dissimilar to targets compared to older children (ages 11-12) and young adults, suggesting a deficiency in pattern completion in early and middle childhood. However, the authors acknowledged that the MST is designed to test pattern separation, and thus interpretation regarding pattern completion should be cautious. There is a general consensus that the literature on pattern completion in humans necessitates tasks that are more process-pure and independent of pattern separation failure hallmarks (e.g., Rollins & Cloude, 2018; Vieweg et al., 2015).

Lastly, an intriguing question remains: when does holistic recollection emerge in development? Many studies on the development of autobiographical memories have examined the structural coherence of event recall to assess the quality of the memory trace (e.g., Bauer & Larkina, 2016; Peterson et al., 2014; Reese et al., 2011). Although the operational definition of coherence varies across studies, there is a general consensus that a coherent account includes components such as context (people, place, time), sequence (chronological ordering), and thematic coherence (understandable to naïve listeners). It has been suggested that infants may encode only bits and pieces of early life experiences, but not coherent representations of past experiences that they can later recall (Fivush, Gray, & Fromhoff, 1987; Umiker-Sebeok, 1979). Preschoolers show frail evidence of coherence (i.e., not at floor level) such that they are able to stay on topic. However, their memories of context and chronological ordering are still limited. Over the course of childhood and beyond, there is a clear developmental progression in all dimensions of coherence (Reese et al., 2011). Even more relevant to the idea of holistic recollection and pattern completion, measurement of narrative "completeness", quantified by tallying the number of different narrative categories (e.g., who, when, what), showed that 4-year-olds recalled a smaller proportion of the events compared to 6-, 8-year-olds and adults (Reese et al., 2011). These results converge with our findings on the different levels of holistic retrieval of multi-element events by 4-year-old children and young adults. Future research could investigate whether there is a link between the development of holistic recollection as estimated using within-event retrieval success contingency and improvements in real-life complex episodic memory narrative coherence.

In conclusion, holistic retrieval of a memory trace unites distinct aspects of the past event including where we were, who we met, and the objects we encountered altogether. The present work shows that by 4 years, memory for complex events is not stored as separate pairs of associations. However, the integration of the units continues to increase after 4 years.

Author Note

We would like to thank Ying Lin, Elizabeth Eberts, Lizi Zhong, Linda Hoffman, Jelani Mumford, and Rebecca Adler for their help with stimuli development and/or with data collection. We would also like to thank Elizabeth Eberts for the voice recording of the memory task. This work was supported in part by National Institutes of Health grants F31HD090872 to C.T.N and RO1 MH091113 to I.R.O. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Health. Aidan Horner is funded by the Wellcome Trust (204277/Z/16/Z) and Economic and Social Research Council (ES/R007454/1). The authors declare no competing or conflicting financial interests.

Author Contributions

C. T. Ngo and N. S. Newcombe developed the research questions and the design of experiments. A. J. Horner contributed to the study design and provided guidance for the data analytical plan. Data were collected by C. T. Ngo and others (see Author Note). Data analyses and results interpretation were led by C. T. Ngo under the supervision of N. S. Newcombe, A. J. Horner, and I. R. Olson. C. T. Ngo drafted the manuscript, and all co-authors provided critical revisions. All authors approved the final version of the manuscript for submission.

Open Practices Statement

Neither of the experiments in this article was formally preregistered. De-identified data for both experiments and experimental materials have been made publically available through the Open Science Framework at https://osf.io/ma5tv/.

References

1. Bauer, P. J., Larkina, M., & Deocampo, J. (2010). Early memory development. In U. Goswami (Eds), The Wiley-Blackwell Handbook of childhood cognitive development, Second edition, (pp. 153-179).
2. Bauer, P. J., & Larkina, M. (2016). Predicting remembering and forgetting of autobiographical memories in children and adults: A 4-year prospective study. Memory, 24(10), 1345-1368.
3. Bisby, J. A., Horner, A. J., Bush, D., & Burgess, N. (2018). Negative emotional content disrupts the coherence of episodic memories. Journal of Experimental Psychology: General, 147(2), 243-256.
4. Fivush, R., Gray, J. T., Fromhoff, F. A. (1987). Two year olds talk about the past. Cognitive Development, 2, 393-410.
5. Grober, E., & Sliwinski, M. (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. Journal of Clinical and Experimental Neuropsychology, 13(6),933-949.
6. Guzowski, J. F., Knierim, J. J., & Moser, E. I. (2004). Ensemble dynamics of hippocampal regions CA3 and CA3. Neuron, 44(4), 581-584.
7. Horner, A. J., & Burgess, N. (2013). The associative structure of memory for multi-element events. Journal of Experimental Psychology: General, 142(4), 1370-1383.
8. Horner, A. J., & Burgess, N. (2014). Pattern completion in multielement event engrams. Current Biology, 24, 988-992.
9. Horner, A. J., Bisby, J. A., Bush, D., Lin, W. J., & Burgess, N. (2015). Evidence

for holistic episodic recollection via hippocampal pattern completion. Nature Communications, 6:7462.

10. Horner, A. J., & Doeller, C. F. (2017). Plasticity of hippocampal memories in humans. Current Opinion in Neurobiology, 43, 102–109.
11. Jonker, T. R., Dimsdale-Zucker, H., Ritchey, M., Clarke, A., & Ranganath, C. (2018). Neural reactivation in parietal cortex enhances memory for episodically linked information. PNAS, 115, 43.
12. Kaufman, A. S., & Kaufman, N. L. (1990). Kaufman Brief Intelligence Test. Circle Pines, MN: American Guidance Service.
13. Keresztes, A., Bender, A. R., Bodammer, N. C., Lindenberger, U., Shing, Y. L., & Werkle-Bergner, M. (2017). Hippocampal maturity promotes memory distinctiveness in childhood and adolescence. PNAS, 114(34), 9212-9217.
14. Lee, J. K., Ekstrom, A. D., & Ghetti, S. (2014). Volume of hippocampal subfields and episodic memory in childhood and adolescence. NeuroImage, 1(94), 162-171.
15. Lloyd, M. E., Doydum, A. O., & Newcombe, N. S. (2009). Memory binding in early childhood: Evidence for a retrieval deficit. Child Development, 80(5), 1321–1328.
16. Marr, D. (1971). Simple memory: A theory for archicortex. Philosophical transactions of the Royal Society of London,, 262(841), 23-81.
17. McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why are there complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychological Review, 102(3), 419-457.

18. Molitor, R. J., Ko, P. C., Hussey, E. P., & Ally, B. A. (2014). Memory-related eye movements challenge behavioral measures of pattern completion and pattern separation. Hippocampus, 24(6), 666-672.
19. Morris, G., Baker-Ward, L., & Bauer, P. J. (2010). The survival of children's autobiographical memories across time. Applied Cognitive Psychology, 24, 527-544.
20. Nelson, H. E. (1982). National Adult Reading Test (NART): For the assessment of premorbid intelligence in patients with dementia: Test manual. Windsor: NFER-Nelson.
21. Ngo, C. T., Newcombe, N. S., & Olson, I. R. (2017). The ontogeny of relational memory and pattern separation. Developmental Science. doi.org/10.1111/desc.12556
22. Ngo, C. T.*, Lin, Y.*, Newcombe, N. S., & Olson, I. R. (2019). Building up and wearing down episodic memory: Mnemonic discrimination and relational binding. Journal of Experimental Psychology: General.
23. Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychological Review, 110(4), 611–646.
24. Olson, I. R., & Newcombe, N. S. (2014). Binding together the elements of episodes: Relational memory and the developmental trajectory of the hippocampus. *The Wiley Handbook and the Development of Children's Memory,* Volume I/II. Patricia J. Bauer and Robyn Fivush.
25. Peterson, C., Morris, G., Baker-Ward, L., & Flynn, S. (2014). Predicting which

childhood memories persist: Contributions of memory characteristics. Developmental Psychology, 50, 439-448.

26. Reese, E., Haden, C. A., Baker-Ward, L., Bauer, P., Fivush, R., & Ornstein, P. A. (2011). Coherence of personal narratives across the lifespan: A multidimensional model and coding method. Journal of Cognition and Development, 12(4), 424-462.
27. Rolls, E. T. (2016). Pattern separation, completion, and categorisation in the hippocampus and neocortex. Neurobiology of Learning and Memory, 129, 4-28.
28. Rollins, L., & Cloude, E. B. (2018). Development of mnemonic discrimination during childhood. Learning & Memory, 25, 294-297.
29. Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Baysian t-tests for accepting and rejecting the null hypothesis. Psychonomic Bulletin & Review, 16, 225-237.
30. Sluzenski, J., Newcombe, N. S., & Kovacs, S. L. (2006). Binding, relational memory, and recall of naturalistic events: a developmental perspective. Journal of Experimental Psychology: Learning, Memory, and Cognition, 32(1), 89–100.
31. Tulving, E. (2002). Episodic memory: From mind to brain. Annual Review of Psychology, 53, 1-25.
32. Umiker-Sebeok, D. J. (1979). Preschool children's intraconversation narratives. Journal of Child Language, 6, 91-109.
33. Vieweg, P., Stangl, M., Howard, L. R., & Wolbers, T. (2015). Changes in pattern completion — a key mechanism to explain age-related recognition memory deficits. Cortex, 64, 343-351.

34. Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. Trends in Neuroscience, 34(10), 515-525.

35. Yim, H., Dennis, S. J., & Sloutsky, V. M. (2013). The Development of Episodic Memory. Psychological Science, 24(11), 2163–2172.

Supporting Information for Ngo, Horner, Newcombe, & Olson

Experiment 1

Results

1.1 Accuracy across blocks

A 2 (age: 4, 6) x 2 (block: 1, 2) mixed ANOVA yielded a main effect of age, $F(1, 60) = 7.50$, $p = 0.01$, *partial* $\eta^2 = 0.11$, but there was neither a main effect of block, $F(1, 60) = 2.41$, $p = 0.13$, *partial* $\eta^2 = 0.04$, nor an age*block interaction, $F(1, 60) = 3.15$, $p = 0.08$, *partial* $\eta^2 = 0.05$.

1.2 Accuracy by cue type and test item type

In addition to examining overall accuracy, we also tested whether cue or test item types (scene, person, object) impacted accuracy for each age group separately. One-way ANOVAs revealed non-significant effects of cue type in all three age groups: 4-year-olds, $F(2, 62) = 1.88$, $p = .16$, *partial* $\eta^2 = 0.06$, 6-year-olds, $F(2, 58) = 2.42$, $p = .10$, partial $\eta^2 = 0.08$, and young adults, $F(2, 60) = 1.29$, $p = .28$, *partial* $\eta^2 = 0.04$ (see Figure 4A). For test item type, one-way ANOVAs revealed a significant effect of test item type for the two younger groups (4-year-old children, $F(2, 62) = 10.95$, $p < .001$, *partial* $\eta^2 = 0.26$, 6-year-old children, $F(2, 58) = 3.30$, $p = .04$, *partial* $\eta^2 = 0.10$) but not for the adults, $F(2, 60) = 1.50$, $p = .23$, *partial* $\eta^2 = 0.05$ (see Figure S1). Accuracy on person test trials was significantly lower than on the scene test trials for both child groups, with no differences between scene tests and object tests. Specifically, for 4-year-olds, person tests were lower than scene tests, $t(31) = 3.84$, $p = .001$, 95% CI [0.04, 0.14], $d = 0.68$, and object tests, $t(31) = -3.44$, $p = .002$, 95% CI [-0.10, -0.03], $d = 0.61$; scene and object test trials did not differ, $t(31) = 1.51$, $p = .14$, 95% CI [-0.01, 0.06], $d = 0.27$. For 6-year-olds, person test trials were lower than scene, t(29)= 2.26, p= .03, 95% CI [0.003, 0.07], d= 0.41, and object test trials, t(29)= -2.37, p= .02, 95% CI [-0.07, -0.01], d= 0.43, with no accuracy difference between the scene and object test trials, t(29)= 0.04, p= .97, 95% CI [-0.04, 0.04], d= 0.01.

1.3. Effects of Dependency Analysis and Age on Retrieval Dependency

Given that dependency was a composite index of pairs that either shared the same cue (A_BA_C) or the same test item from the same event (B_AC_A), it raised the question of whether dependency differed between associations that shared the same cue item and those that shared the same test item. Therefore, we tested whether dependency analyses (dependency for pairs with the common cue [A_BA_C] versus pairs with the common test item [B_AC_A]) and age impacted dependency. A 2 (analysis: same cue or same test item) x 3 (age) ANOVA on dependency revealed neither main effects of analysis type, F(1, 90)= 0.72, 95% CI [-0.01, 0.1], p= 0.40, *partial* η^2= 0.01, BF_{01}=5.28, nor an age*analysis interaction, F(2, 90)= 1.07, p= 0.35, *partial* η^2= 0.02, BF_{01}= 9.28. However, there was a main effect of age, F(2, 90)= 3.27, p= 0.04, *partial* η^2= 0.07. These results suggest that dependency did not differ between analyses of shared cue pairs and shared tested item pairs.

Experiment 2

Results

2.1. Accuracy by cue type and test item type

We again tested whether cue and test item type had a main effect on accuracy. For cue type, we found an effect, F(2, 58)= 5.15, p= .01, *partial* η^2= 0.15, with higher accuracy for person cue trials (M= .58, SE= .03) than for scene (M= .53, SE= .04), t(29)=

-2.61, p= .01, 95% CI [-0.09, -0.01], d= 0.48, or object cue trials (M=.52, SE=.03), t(29)= 2.88, p= .01, 95% CI [0.02, 0.11], d= 0.53. Scene and object cue trials did not differ from one another, t(29)= 0.63, p= .53, 95% CI [-0.03, 0.06], d= 0.12.

Test item type also affected accuracy, F(2, 58)= 5.68, p= .01, *partial* η^2= 0.16, with higher accuracy for scene test trials (M= .58, SE= .03) compared to person (M= .53, SE= .04), t(29)= 3.02, p= .01, 95% CI [-0.02, 0.08], d= 0.55, or object test trials (M= .52, SE= .04), t(29)= 3.03, p= .01, 95% CI [-0.02, 0.10], d= 0.55. There was no difference in accuracy between person and object test trials, t(29)= 0.58, p= .57, 95% CI [-0.03, 0.05], d= 0.11 (see Figure S1).

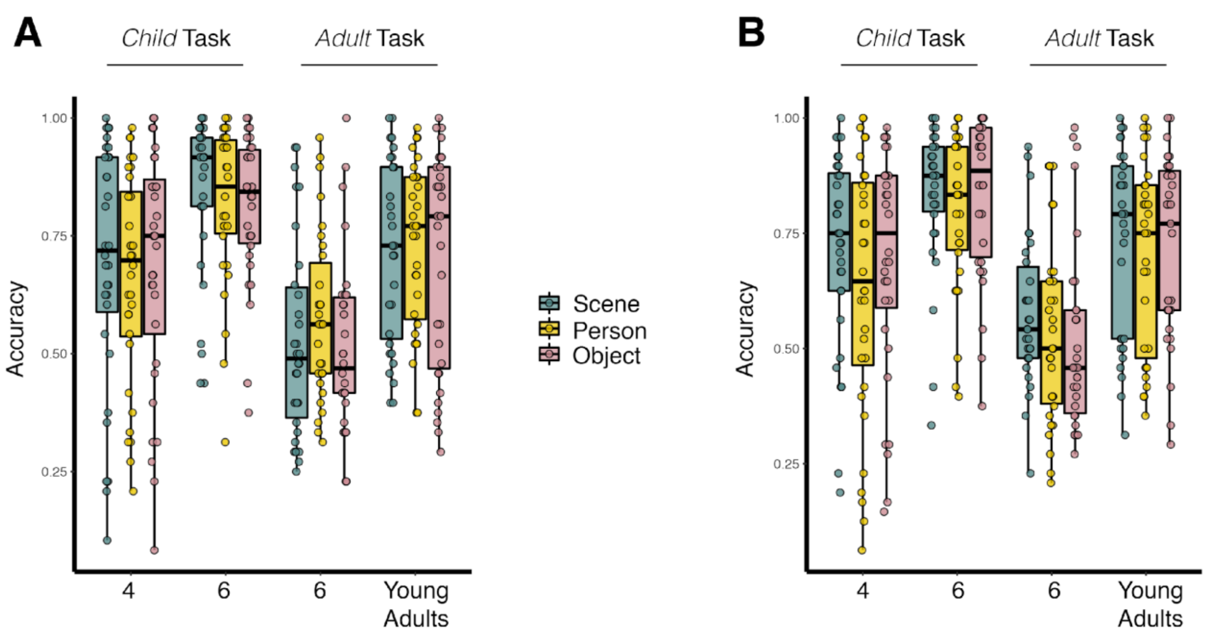

Figure S1. The distribution of accuracy across 3 cue types (A) and 3 test item types (B) separated by age groups.

Experiments 1 & 2

3.1. Dependency and Verbal IQ

Further, we tested whether event retrieval coherence simply tracks individual differences in verbal skills in each age group using Pearson correlations. We did not detect a significant correlation between dependency and verbal intelligence (measured by KBIT and AMNART in children and adults, respectively) in the 4-year-olds, $r(30)= -0.24$, 95% CI [-0.56, 0.13], $p= .19$, $BF_{01}= 1.96$, the 6-year-olds, $r(59)= -0.22$, 95% CI [-0.45, 0.04], $p= .09$, $BF_{01}= 1.56$, or young adults, $r(31)= 0.21$, 95% CI [-0.16, 0.52], $p= .26$, $BF_{01}= 2.45$. It is important to note that post-hoc power analyses estimated that we only obtained a modest power of 0.39, 0.65, and 0.38, for 4-, 6-year-olds, and young adults, respectively, to detect a significant correlation. Therefore, the relation between verbal intelligence and retrieval dependency should be addressed in future investigations with adequate statistical power.