Development of Holistic Episodic Recollection

Chi T. Ngo[1] iD, Aidan J. Horner[2,3] iD, Nora S. Newcombe[1], and Ingrid R. Olson[1]
[1]Department of Psychology, Temple University; [2]Department of Psychology, University of York; and [3]York Biomedical Research Institute, University of York

Abstract
Episodic memory binds 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. Previous work has shown that such holistic recollection is evident in young adults, as revealed by dependency in retrieval success for various associations from the same event. In addition, episodic memory shows clear quantitative increases during early childhood. However, the ontogeny of holistic recollection is uncharted. Using dependency analyses, we found here that 4-year-olds ($n = 32$), 6-year-olds ($n = 30$), and young adults ($n = 31$) all retrieved complex events in a holistic manner; specifically, 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 the age 4 to adulthood. Thus, extended refinement of multiway binding may be one aspect of episodic memory development.

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 by which 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, exposure to part of a past experience leads to recurrent connectivity in the 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, Stangl, Howard, & Wolbers, 2015; Yassa & Stark, 2011). One paradigm relies on the conceptualization that pattern completion allows for the recovery of the entire event on the basis of a partial cue so that all elements within an event can be retrieved (or not retrieved). Using multielement-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 consists 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

Corresponding Author:
Chi T. Ngo, Temple University, Department of Psychology, 1701 N. 13th St., Philadelphia, PA 19122
E-mail: chi.ngo@temple.edu

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, Bisby, Bush, Lin, & Burgess, 2015; for a review, see Horner & Doeller, 2017). Corroborating these findings, a recent study showed that retrieval improved long-term memory not only for the specific information tested but also for nontargeted 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 episodiclike memory capacities emerge early in development (for a review, see Bauer, Larkina, & Deocampo, 2010), several aspects of episodic memory are still far from mature in preschool-age children (for a review, see Olson & Newcombe, 2014). Evidence from research on episodic memory development employing laboratory-based paradigms suggests that children's relational memory improves robustly between the ages of 4 and 6 years, as shown in various tasks using interitem (e.g., *bear–book*), item–context (e.g., *bear–library*; Lloyd, Doydum, & Newcombe, 2009; Sluzenski, Newcombe, & Kovacs, 2006), or item–item–context (e.g., *bear–book–red house*; Ngo, Lin, Newcombe, & Olson, 2019; Ngo, Newcombe, & Olson, 2017; Yim, Dennis, & Sloutsky, 2013) associations. When asked to remember unique item–context pairs (e.g., bear–library), 6-year-olds showed superior relational memory compared with 4-year-olds (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, Ekstrom, & Ghetti, 2014), it is likely that pattern completion subserving holistic recollection also follows prolonged maturation between the early years of life and young adulthood.

The current research specifically focused 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 multielement-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 multielement events, each of which contained a scene, a person, and an object. In a cued recognition test, each element in turn served as the cue or as the retrieval item. Pattern completion was indexed by retrieval dependency—the degree to which the accuracy for within-event test trials was 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 relation between pairwise associative recognition and dependency by testing a separate group of 6-year-olds with a different task protocol that resulted in lower performance in pairwise associative-recognition memory. This experiment revealed that retrieval dependency among the 6-year-olds remained comparable with levels seen in Experiment 1, even when accuracy was dampened.

Experiment 1

Method

Participants. To ensure that we would have sufficient power (0.80) to detect an interaction between age and the proportion-of-joint-retrieval indices, we conducted an a priori power analysis of a repeated 3 (age) × 2 (proportion-of-joint-retrieval indices: data, independent model) mixed analysis of variance (ANOVA) using G*Power (Version 3.1; Faul, Erdfelder, Lang, & Buchner, 2007). The power analysis determined a total sample size of 81 (27 participants per age cohort) with sufficient power (.80) to detect a medium effect size ($f = .25$). The correlation coefficients between the repeated measures were set at 0, as this value was unknown. We slightly oversampled because we anticipated that some younger participants might fail to complete the task.

A total of 32 four-year-old children (15 female; age: $M = 52.06$ months, $SD = 3.37$) and 30 six-year-old children (17 female: age: $M = 76.37$ months, $SD = 2.16$) from the Philadelphia area participated in the study at the Temple University Ambler Infant and Child Laboratory and at schools in 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 four-year-olds and 2 six-year-olds) were tested but were not included in data analyses because of experimenter error ($n = 2$) or because the procedure was not completed ($n = 4$). In addition to the final sample reported above, 2 six-year-old
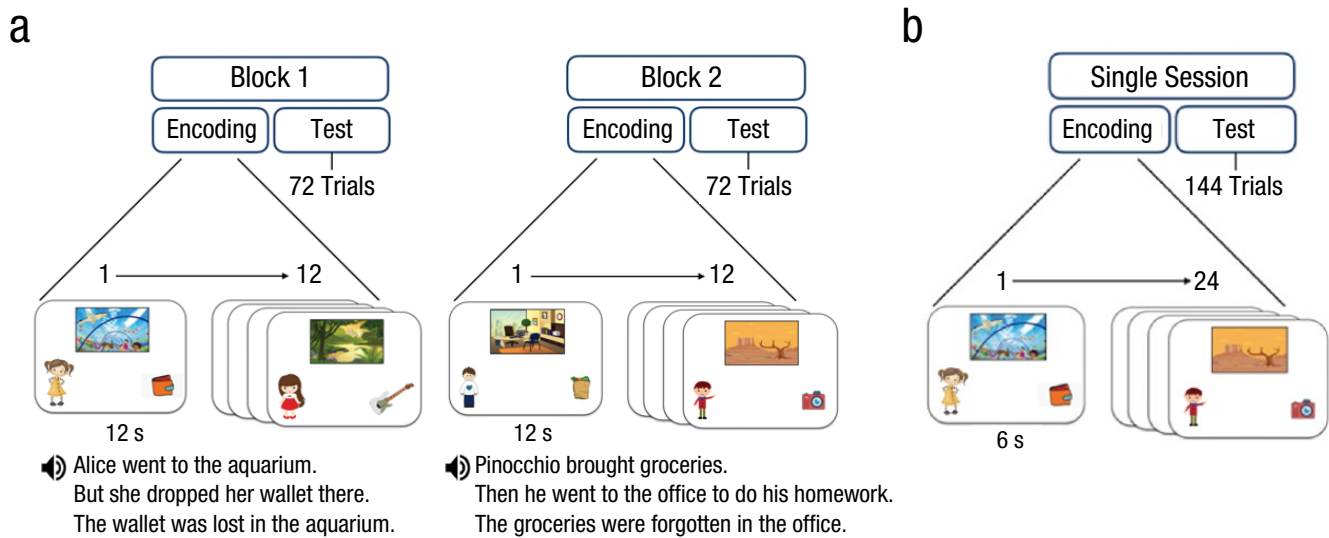
Fig. 1. Procedure of the child (a) and adult (b) multielement-event task. In the child task, participants viewed 24 events presented in two encoding sessions, each consisting of 12 events. Each event lasted 12 s and was accompanied by an audio-recorded narrative. The test phase of each block consisted of 72 test trials. In the adult task procedure, participants studied 24 events (6 s each) together and without the recorded narrative. The test phase consisted of 144 test trials. Note that the characters shown in each event were well-known cartoon characters (e.g., Alice, Pinocchio), which have been replaced in this illustration for copyright concerns.

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 (18 female; age: $M = 20.65$ years, $SD = 3.23$, range = 18–31) from Temple University who participated for partial course credit. All participants gave informed consent and reported having normal or corrected-to-normal vision. This experiment was completed in accordance with the guidelines of the Temple University Institutional Review Board, which also approved the experiment.

Memory task.

Materials. We sampled 24 cartoon images of distinct scenes (12 indoor scenes, e.g., an aquarium; 12 outdoor scenes, e.g., a playground), 24 cartoon images of common objects (e.g., a watch), and 24 images of cartoon characters from nonoverlapping movies or books (12 males, e.g., Pinocchio; 12 females, e.g., Alice) from the Google Images search engine. From this pool of selected images, we then constructed 24 "events," each consisting of a scene (e.g., an aquarium), a person (e.g., Alice), and an object (e.g., a wallet). The event assignment of the elements was randomized, with the exception that items with preexperimental 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 six test trials per event (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) and totaling 144 test trials.

Procedure. All participants were tested individually. The task procedure administered to children consisted of 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-in. 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 in each story. Then, participants viewed a series of events (12 s each; 0.5 s intertrial interval). A short audio-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 Fig. 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 narrative was constructed this way to engage children in the task and to increase the likelihood that children would pay attention to all of the elements in an event. Prior to encoding, we provided one example (a playground, Elastigirl, a 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 of each studied event, resulting in 6 test trials per event, which totaled 72 test
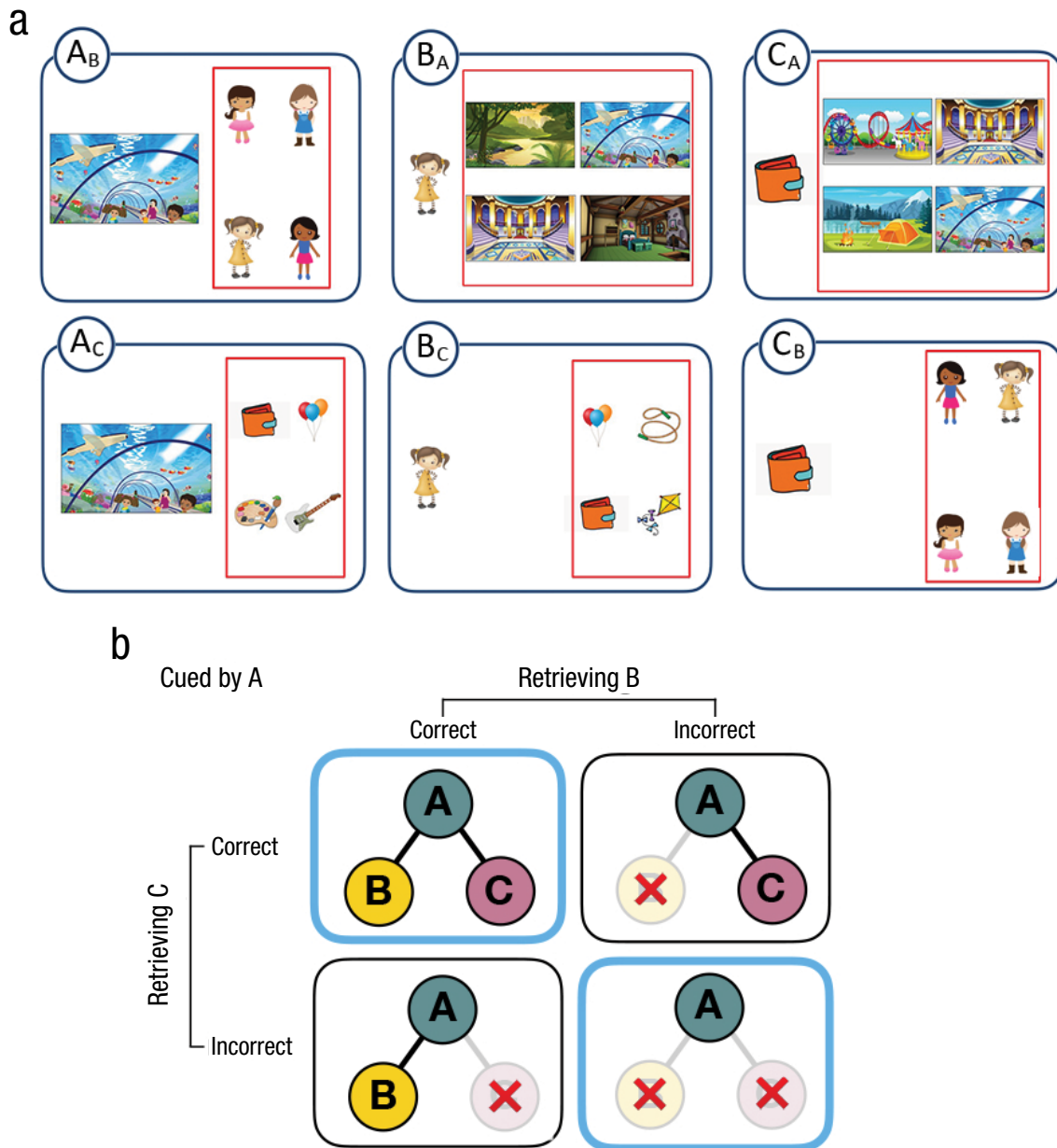
Fig. 2. A schematic depiction of the task design and the 2 × 2 contingency table used to estimate retrieval dependency. Examples of six retrieval types per event in the test phase are shown in (a). Each element of a studied event took a turn serving as the cue (item presented on the left side of the screen) and the tested element (one of the four options presented inside the red box). The schematic (b) shows how the proportion of joint retrieval for A_B and A_C pairs was computed for each participant. The contingency table shows the proportion of events that fell within each of the four categories: Both A_B and A_C pairs were retrieved correctly, both A_B and A_C pairs were retrieved incorrectly, A_B was retrieved correctly and A_C was retrieved incorrectly, and A_B was retrieved incorrectly and A_C was retrieved correctly. The proportion of events in the blue-outlined boxes (both pairs correct and both pairs incorrect) were added, and the sum was divided by the total number of events. Note that the characters shown in each event were well-known cartoon characters (e.g., Alice, Pinocchio), which have been replaced in this illustration for copyright concerns.

trials per block. On each trial, a cue and four options were presented simultaneously on the screen (see Fig. 2a). Among four options, one was a target—the correct item because it belonged to the same event as the cue. The three lures were same-category elements from different events. The lures always came from the events that contained same-sex characters, so that participants could not eliminate lures on the basis of general mnemonic heuristics (e.g., remembering that there was a female character who went to the aquarium). Across all 24 events, any two test trials that had overlapping cue items (e.g., A_B^1 and A_C^1) or in which tested items (e.g.,

$B_A{}^1$ and $C_A{}^1$) shared only one foil item (out of three) with respect 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 min.

The adult task procedure was similar to the child task procedure but with a few differences. First, the whole procedure was administered in a single session comprising 24 encoding events and 144 test trials. Second, no narratives were implemented at the encoding phase to avoid potential ceiling performance in young adults. Third, each encoding trial was presented for 6 s (see Fig. 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). All children took the Kaufman Brief Intelligence Test (second edition; 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?—a bunny"). The test, which increases in difficulty in each section, was terminated when a child provided four incorrect responses consecutively. A standard score was calculated for each child on the basis of his or her age.

Adults completed 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 irregular words aloud. 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 (Bisby, Horner, Bush, & Burgess, 2018; Horner et al., 2015; Horner & Burgess, 2013, 2014). Six 2 × 2 contingency tables for the data and the predicted independent model were computed for each participant on the basis of 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 ($A_B A_C$; 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 ($B_A C_A$; i.e., cue with B and retrieve A, and cue with C and retrieve A). Each 2 × 2 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 is correct and A_C is incorrect, and A_C is correct and A_B is incorrect. 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 Fig. 2b). We then averaged this measure across six contingencies tables (three tables for the $A_B A_C$ analysis for each element type and three tables for the $B_A C_A$ analysis for each element type) for each participant.

The independent model of retrieval estimated the degree of statistical dependency if retrieval success for specific cue–test pairs (cue: person, test: scene) was independent of retrieval success of other cue–test pairs (cue: person, test: object) in relation to participants' overall accuracy. The independent model predicted the proportion of joint retrieval given a participant's overall level of performance if retrievals of event pairs were independent such that the probability of the successful retrieval for both, for example, A_B and A_C was equal to $P_{AB} \times P_{AC}$, where P_{AB} was the probability of retrieving B when cued by A across all events, and similarly for P_{AC} (see Fig. 3 for full details). The proportion of joint retrieval for the independent model (calculated in the same manner as described above) served as a predicted baseline for which we compared the proportion of joint retrieval in the data. Given that the proportion of joint retrieval for the data scaled 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) was significantly greater than zero, this provided evidence for significant retrieval dependency (for the same approach, see Horner & Burgess, 2013, 2014). In addition, we took 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. All planned statistical analyses were performed using SPSS software. Key null findings were tested with Bayesian hypothesis testing using JASP software.

Probability of Retrieving B When Cued by A

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

Fig. 3. Contingency table for the predicted independent model for the 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 correct-correct and the incorrect-incorrect cells and dividing by the sum of all four cells.

Results

Overall accuracy. Overall accuracy was 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 ps > .53, all Bayes factors in favor of the null hypothesis (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-samples t test showed that 6-year-olds ($M = .82$, $SE = .03$) outperformed 4-year-olds ($M = .68$, $SE = .04$), $t(60) = -2.73$, $p = .01$, 95% confidence interval (CI) for the mean difference = [−.25, −.04], $d = -0.70$, 95% CI = [−1.21, −0.18], suggesting improvement in individual pairwise associative memory between the ages of 4 and 6 (see Fig. 4). Results for the effects of age and block on accuracy are reported in the Supplemental Material available online (see Section 1.1).

As for young adults, overall accuracy was at .72 ($SD = .19$). However, we did not compare overall accuracy between children and adults because of differences in task procedure. Results for the effects of cue and test-item types on accuracy for each age group are reported in the Supplemental Material (see Section 1.2).

Retrieval dependency. The primary questions of this research were whether holistic recollection is evident at the ages of 4 and 6 years, during a crucial developmental window of robust gains in episodic memory, and whether holistic recollection changes with age. To answer the first question, we conducted a one-sample t test to determine 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: Dependency ($M = .07$, $SE = .01$, 95% CI = [.05, .08]) was significantly greater than zero, $t(30) = 7.37$, $p < .001$, $d = 1.32$, 95% CI = [0.83, 1.80]. These results conceptually replicate those of previous studies showing that retrieval
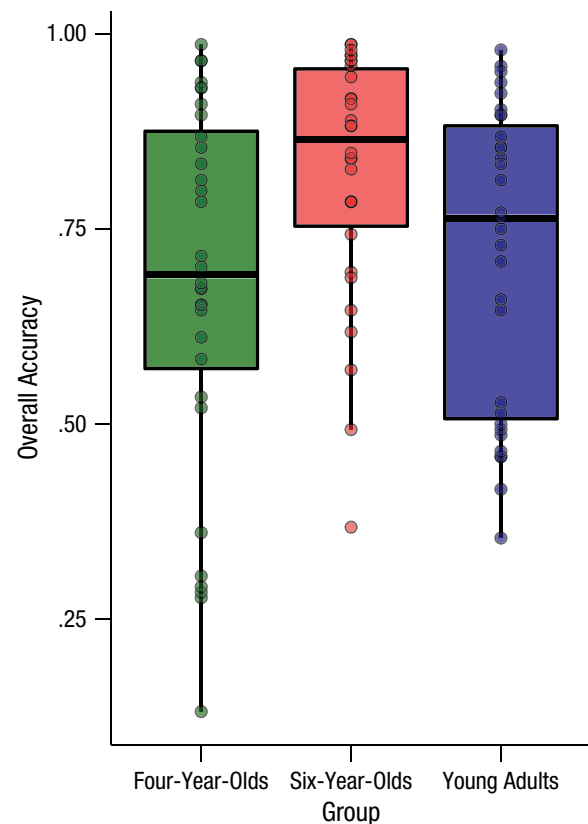
Fig. 4. Box-and-whisker plots showing the distribution of overall accuracy, separately for each age group (Experiment 1). The bottom and top edges of each box indicate the interquartile range of the data, the horizontal line indicates the median, the whiskers extend to 1.5 times the interquartile range, and the dots represent individual data points. The overall task procedures differed between children and young adults: For children, 24 events were divided into two encoding-test sessions (12 events in each session), whereas for young adults, 24 events were administered in a single encoding-test session.
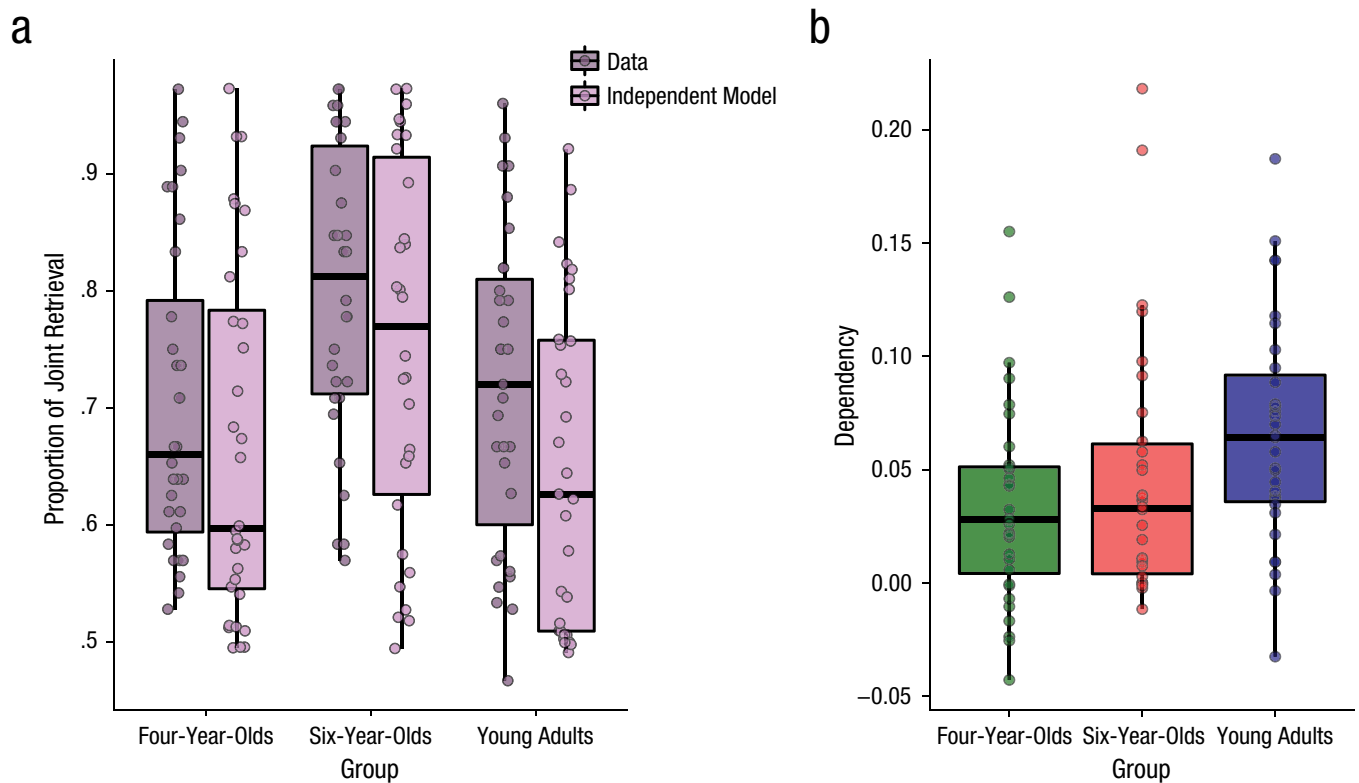
a

b

Fig. 5. Box-and-whisker plots showing the distribution of (a) the proportion of joint retrieval in the data and predicted by the independent model and (b) the magnitude of retrieval dependency, separately for each age group (Experiment 1). The bottom and top edges of each box indicate the interquartile range of the data, the horizontal line indicates the median, the whiskers extend to 1.5 times the interquartile range, and the dots represent individual data points.

dependency is significant in young adults when verbal stimuli are used (Horner & Burgess, 2013, 2014).

Interestingly, 4-year-old children also showed significant retrieval dependency: Dependency ($M = .03$, $SE = .01$, 95% CI = [.02, .05]) significantly exceeded zero, $t(31) = 4.33$, $p < .001$, $d = 0.77$, 95% CI = [0.37, 1.16]. Dependency ($M = .05$, $SE = .01$, 95% CI = [.03, .07]) also significantly exceeded zero in 6-year-old children, $t(29) = 4.47$, $p < .001$, $d = 0.82$, 95% CI = [0.40, 1.23] (see Fig. 5a). Thus, evidence for holistic recollection seen in all three age groups demonstrates that memories for multielement 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 differed across age groups. A one-way ANOVA showed a significant effect of age, $F(2, 90) = 3.27$, $p = .04$, $\eta_p^2 = 0.07$. Tukey post hoc tests showed that dependency was lower in 4-year-olds compared with young adults, $t(61) = -2.54$, 95% CI for the mean difference = [−.06, −.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 for the mean difference = [−.04, .02], $p =$

.59, $d = 0.25$, $BF_{01} = 2.57$, or young adults, $t(59) = -1.52$, $p = 0.29$, 95% CI for the mean difference = [−.05, .01], $d = 0.37$, $BF_{01} = 1.63$ (see Fig. 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 was 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 for the effects of analysis (same cue vs. same test item) and age on dependency are reported in the Supplemental Material (see Section 1.3).

Experiment 2

Overall accuracy in Experiment 1 was relatively high in the 6-year-old children, with a portion of the children hovering near ceiling-level performance. Thus, we next examined whether dependency would be affected if overall accuracy were dampened in this age group. We tested an independent group of 30 six-year-olds
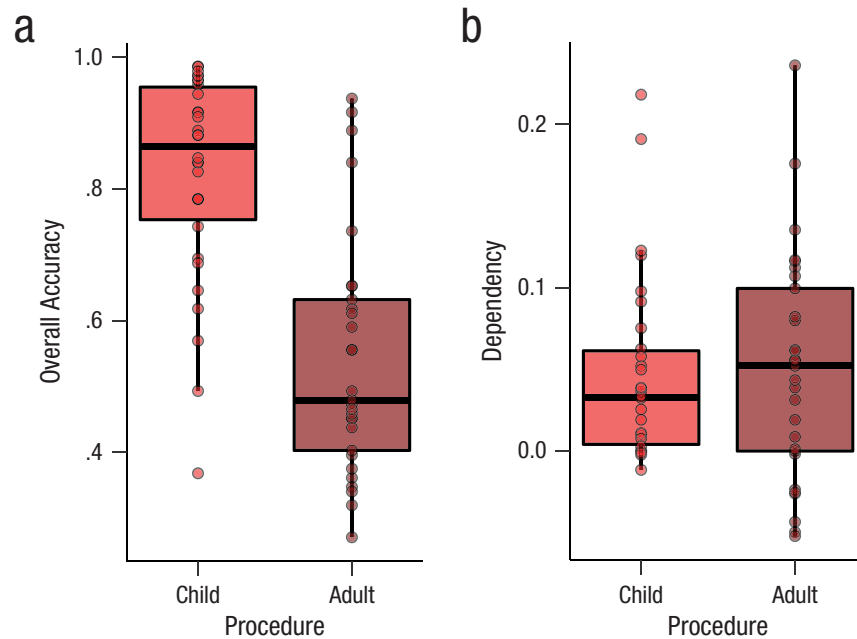
Fig. 6. Box-and-whisker plots showing the distribution of (a) overall accuracy (a) and (b) retrieval dependency, separately for the two groups of 6-year-old children who were given the child and adult task procedures (Experiment 2). The bottom and top edges of each box indicate the interquartile range of the data, the horizontal line indicates the median, the whiskers extend to 1.5 times the interquartile range, and the dots represent individual data points.

(16 female; age: $M = 77.13$ months, $SD = 3.20$) with the adult task version (the same procedure administered to 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 for the mean difference = [−.06, .21], $p = .29$, $d = 0.40$, 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 for the mean difference = [−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 ($M = .54$, $SE = .03$) compared with their same-age peers who received the child task procedure ($M = .82$, $SE = .03$), $t(58) = 6.41$, $p < .001$, 95% CI for the mean difference = [.19, .37], $d = 1.65$, 95% CI = [1.06, 2.24][1] (see Fig. 6a). Results for accuracy dependent on cue and test item types for these children are reported in the Supplemental Material (see Section 2.1).

Again, our primary questions concerned dependency. Dependency ($M = .05$, $SE = .01$, 95% CI = [.03, .08]) significantly exceeded zero in this group of 6-year-olds, $t(29) = 4.17$, $p < .001$, $d = 0.76$, 95% CI = [0.35, 1.16]. Critically, dependency did not differ between

6-year-old children in this experiment and those in Experiment 1 ($M = .05$, $SE = .01$), $t(58) = −0.40$, $p = .69$, 95% CI for the mean difference = [−.04, .03], $d = −0.10$, 95% CI = [−0.61, 0.41], $BF_{01} = 3.57$ (see Fig. 6b).

We also compared this group of 6-year-olds with young adults, given that they received the same task procedure. Six-year-olds had lower overall accuracy relative to young adults, $t(59) = −3.66$, $p < .001$, 95% CI for the mean difference = [−.27, −.08], $d = 0.94$ (Note 1). However, their dependency did not differ significantly from that of adults, $t(59) = −0.85$, $p = .40$, 95% CI for the mean difference = [−.04, .02], $d = −0.22$, 95% CI = [−0.72, 0.29], $BF_{01} = 2.83$.

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, $r(123) = −.13$, 95% CI = [−.30, .04], $p = 0.14$, $BF_{01} = 3.05$. Results for the correlation between dependency and verbal IQ are reported in the Supplemental Material (see Section 3.1).

General Discussion

A defining feature of episodic memory is that complex and multimodal events are stored as coherent representations, so episodic retrieval entails the holistic reexperience

of all constituents of an event (Tulving, 2002). The present work shows that as early as 4 years of age, children are capable of retrieving multielement 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 6-year-olds is intermediate—it does not differ significantly from either the younger or older individuals. 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 relatively lower relational memory through preschool years than later, young children's episodic memories still possess a significant degree of cohesion. Most studies of episodic-memory development have used paired-associates 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, we elucidated an important characteristic of how episodic memory is refined from early life to adulthood.

In adult humans, one functional MRI study used a variant of the multielement-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 of 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 nontarget 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 intrahippocampal circuitry has a slow maturational rate (e.g., Keresztes et al., 2017; Lee et al., 2014).

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 et al., 2015), whereas in the multielement-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, or MST; for a review, see 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). In one study that used the MST, pattern completion was operationalized 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 with 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 done with caution. 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, Morris, Baker-Ward, & Flynn, 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 naive 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 a slight degree of evidence of coherence (i.e., not at floor level) in 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

with 6-year-olds, 8-year-olds, and adults (Reese et al., 2011). These results converge with our findings on the different levels of holistic retrieval of multielement 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. The present work shows that by the age of 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.

Action Editor

Caren Rotello served as action editor for this article.

Author Contributions

C. T. Ngo and N. S. Newcombe developed the research questions and designed the experiments. A. J. Horner contributed to the experimental design and provided guidance for the data-analysis plan. Data were collected by C. T. Ngo. Data were analyzed and interpreted 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 coauthors provided critical revisions. All authors approved the final version of the manuscript for submission.

ORCID iDs

Chi T. Ngo (iD) https://orcid.org/0000-0001-6962-7168
Aidan J. Horner (iD) https://orcid.org/0000-0003-0882-9756

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797619879441

Open Practices

Deidentified data as well as materials for both experiments have been made publically available through the Open Science Framework at https://osf.io/2pnu6/ and https://osf.io/arphg/, respectively. The design and analysis plans for the experiments were not preregistered. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797619879441. This article has received the badges for Open Data and Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

Note

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

References

Bauer, P. J., & Larkina, M. (2016). Predicting remembering and forgetting of autobiographical memories in children and adults: A 4-year prospective study. *Memory*, *24*, 1345–1368.

Bauer, P. J., Larkina, M., & Deocampo, J. (2010). Early memory development. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 153–179). Hoboken, NJ: Blackwell.

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*, 243–256.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Fivush, R., Gray, J. T., & Fromhoff, F. A. (1987). Two-year-old talk about the past. *Cognitive Development*, *2*, 393–409.

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*, 933–949.

Guzowski, J. F., Knierim, J. J., & Moser, E. I. (2004). Ensemble dynamics of hippocampal regions CA3 and CA3. *Neuron*, *44*, 581–584.

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*, Article 7462. doi:10.1038/ncomms8462

Horner, A. J., & Burgess, N. (2013). The associative structure of memory for multi-element events. *Journal of Experimental Psychology: General*, *142*, 1370–1383.

Horner, A. J., & Burgess, N. (2014). Pattern completion in multielement event engrams. *Current Biology, 24*, 988–992.

Horner, A. J., & Doeller, C. F. (2017). Plasticity of hippocampal memories in humans. *Current Opinion in Neurobiology, 43*, 102–109.

Jonker, T. R., Dimsdale-Zucker, H., Ritchey, M., Clarke, A., & Ranganath, C. (2018). Neural reactivation in parietal cortex enhances memory for episodically linked information. *Proceedings of the National Academy of Sciences, USA, 115*, 11084–11089.

Kaufman, A. S., & Kaufman, N. L. (1990). *Kaufman Brief Intelligence Test*. Circle Pines, MN: American Guidance Service.

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. *Proceedings of the National Academy of Sciences, USA, 114*, 9212–9217.

Lee, J. K., Ekstrom, A. D., & Ghetti, S. (2014). Volume of hippocampal subfields and episodic memory in childhood and adolescence. *NeuroImage, 1*, 162–171.

Lloyd, M. E., Doydum, A. O., & Newcombe, N. S. (2009). Memory binding in early childhood: Evidence for a retrieval deficit. *Child Development, 80*, 1321–1328.

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society B: Biological Sciences, 262*, 23–81.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419–457.

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*, 666–672.

Nelson, H. E. (1982). *National Adult Reading Test (NART): For the assessment of premorbid intelligence in patients with dementia: Test manual*. Windsor, England: NFER-Nelson.

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, 148*, 1463–1479.

Ngo, C. T., Newcombe, N. S., & Olson, I. R. (2017). The ontogeny of relational memory and pattern separation.

Developmental Science, 21, Article e12556. doi:10.1111/desc.12556

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review, 110*, 611–646.

Olson, I. R., & Newcombe, N. S. (2014). Binding together the elements of episodes: Relational memory and the developmental trajectory of the hippocampus. In P. J. Bauer & R. Fivush (Eds.), *The Wiley handbook on the development of children's memory* (pp. 285–308). Hoboken, NJ: Wiley.

Peterson, C., Morris, G., Baker-Ward, L., & Flynn, S. (2014). Predicting which childhood memories persist: Contributions of memory characteristics. *Developmental Psychology, 50*, 439–448.

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*, 424–462.

Rollins, L., & Cloude, E. B. (2018). Development of mnemonic discrimination during childhood. *Learning & Memory, 25*, 294–297.

Rolls, E. T. (2016). Pattern separation, completion, and categorisation in the hippocampus and neocortex. *Neurobiology of Learning and Memory, 129*, 4–28.

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*, 89–100.

Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology, 53*, 1–25.

Umiker-Sebeok, D. J. (1979). Preschool children's intraconversation narratives. *Journal of Child Language, 6*, 91–109.

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.

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences, 34*, 515–525.

Yim, H., Dennis, S. J., & Sloutsky, V. M. (2013). The development of episodic memory: Items, contexts, and relations. *Psychological Science, 24*, 2163–2172.