



Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp



Understanding relational binding in early childhood: Interacting effects of overlap and delay



Susan L. Benear^{a,*}, Chi T. Ngo^b, Ingrid R. Olson^a, Nora S. Newcombe^a

^a Department of Psychology, Temple University, Philadelphia, PA 19122, USA

^b Center for Lifespan Psychology, Max Planck Institute for Human Development, 14195 Berlin, Germany

ARTICLE INFO

Article history:

Received 19 November 2020

Revised 5 February 2021

Available online 22 April 2021

Keywords:

Episodic memory
Child development
Delayed memory
Memory interference
Relational binding
Cognitive development

ABSTRACT

Episodic memories typically share overlapping elements in distinctive combinations, and to be valuable for future behavior they need to withstand delays. There is relatively little work on whether children have special difficulty with overlap or withstanding delay. However, Yim, Dennis, and Sloutsky (*Psychological Science*, 2013, Vol. 24, pp. 2163–2172) suggested that extensive overlap is more problematic for younger children, and Darby and Sloutsky (*Psychological Science*, 2015, Vol. 26, pp. 1937–1946) reported that a 48-h delay period actually improves children's memory for overlapping pairs of items. In the current study, we asked how children's episodic memory is affected by stimulus overlap, delay, and age using visual stimuli containing either overlapping or unique item pairs. Children aged 4 and 6 years were tested both immediately and after a 24-h delay. As expected, older children performed better than younger children, and both age groups performed worse on overlapping pairs. Surprisingly, the 24-h delay had only a marginal effect on overall accuracy. Although there were no interactions, when errors were examined, there was evidence that delay buffered memory for overlapping pairs against cross-contextual confusion for younger children.

© 2021 Elsevier Inc. All rights reserved.

* Corresponding author.

E-mail address: susan.benear@temple.edu (S.L. Benear).

Introduction

Episodic memories for events anchored in a specific spatiotemporal context are a central aspect of our sense of personal identity, are important in social interactions, and support decision making about the future (Prebble, Addis, & Tippett, 2013; Szpunar, Addis, McLelland, & Schacter, 2013; Tulving, 2002). By 4 years of age, children show good episodic memory in many paradigms (reviewed in Bauer, Larkina, & Deocampo, 2011). However, the ability to bind associations between two items or between an item and its context—i.e. *relational binding*, a hallmark of episodic memory—develops in several important ways from 4 to 6 years of age (reviewed by Newcombe, Benear, Ngo, & Olson, *in press*). In addition, temporal-spatial specificity (Newcombe, Balcomb, Ferrara, Hansen, & Koski, 2014), mnemonic discrimination (Ngo, Lin, Newcombe, & Olson, 2019), and holistic recollection (Ngo, Horner, Newcombe, & Olson, 2019) show strikingly similar patterns of age-related growth, with marked improvement from 4 to 6 years of age and continued change up to 8 years in some more complex dimensions such as the discrimination of similar contexts (Ngo, Lin, et al., 2019).

Despite this growing knowledge of episodic memory development, several aspects of memory development remain under-explored. In this study, we investigated two issues: whether there are differential effects with age of contextual overlap and whether younger children's memory is more affected by delay. In addition, we examined the interaction of age, overlap, and delay. Overlap of experiences across contexts is important to consider because episodic memories often contain overlapping relational information. For example, a child might remember that the last time she visited her grandparents' house she did a puzzle with her grandmother, but during the visit prior to that she helped her grandfather make cookies, so that her grandparents' house is linked both to the puzzle and to the cookies. Overlap can cause memory interference ("Did I make the cookies with Grandma or Grandpa?"). Investigating delay is important because episodic memories are useful only if recalled beyond the brief window just after an event occurs.

Overlap

Younger children have difficulty in recalling associated pairs of stimuli even when they contain no overlapping elements (e.g., an AB-CD object pair paradigm; Lloyd, Doydum, & Newcombe, 2009; Sluzenski, Newcombe, & Kovacs, 2006), but overlap may create an extra burden. Nonhuman animals struggle with memory deficits in the face of such interference (Jitsumori, Wright, & Cook, 1988; Kubo-Kawai & Kawai, 2007), and research has shown that performance on tasks involving cross-contextual overlap relies on hippocampal functioning (e.g., Eacott & Norman, 2004). Thus, based on the protracted course of hippocampal development across early childhood (Canada, Ngo, Newcombe, Geng, & Riggins, 2019; DeMaster, Pathman, Lee, & Ghetti, 2014; Gogtay et al., 2006; Krogsrud et al., 2014; reviewed in Canada, Botdorf, & Riggins, 2020), one might expect that younger children would particularly suffer from increased interference across contexts.

Children have been tested in relational binding studies with overlapping elements (i.e., an AB-AC object pair paradigm; e.g., Darby & Sloutsky, 2015; Ngo, Lin, et al., 2019; Ngo, Newcombe, & Olson, 2018), and older children outperform younger children in these paradigms. However, direct contrasts between unique and overlapping pairings are rare. Yim, Dennis, and Sloutsky (2013) tested 4- and 7-year-old children on a paradigm involving three types of associative pairs: unique pairs (AB-CD) that share no elements within pairs between the two lists of items, overlapping pairs (AB-AC) that share one element within pairs between the two lists of items, and an even more overlapping kind of pair in which all the items from the first list are reused but the pairings are shuffled (AB-ABr). For example, in the AB-CD condition, List 1 = bike-cup and couch-cat, and List 2 = backpack-fork and umbrella-football; in the AB-AC condition, List 1 = bike-cup and couch-cat, and List 2 = bike-fork and couch-football; and in the AB-ABr condition, List 1 = bike-cup and couch-cat, and List 2 = bike-cat and couch-cup. Both age groups showed memory interference when pairs were overlapping (AB-AC and AB-ABr conditions) but not when they were unique (AB-CD condition), with no interaction of pair type with age. However, a multinomial processing tree (MPT) model suggested an interaction whereby 7-year-olds outperform 4-year-olds specifically on the most complex latent process of

three-way binding, indicated by successful memory for the most challenging associative pairs (AB–ABr). MPT models attempt to infer the contributions of latent processes or structures to categorical data, with Yim et al. (2013) model assuming that the proportion of a specific response was determined by the availability of four types of latent structures: item recognition, item–item binding, item–context binding, and three-way item–item–context binding. Given data only from that study, whether younger children are differentially impaired when episodic memory tasks involve shared information across contexts remains an open question.

Delay

Relatively little is known about how young children's episodic memory performance changes when tested immediately or after a delay of 24 h or more. Although there are many studies examining children's autobiographical memories for events after very long delays (e.g., Flin, Boon, Knox, & Bull, 1992; Poole & White, 1993), less is known about how children retain non-autobiographical episodic details over shorter delays of just 1 or 2 days. A recent study demonstrated that children ages 3–5 were able to retain memory for static object–scene associated pairs across a 24-h delay (Saragosa-Harris et al., 2021). In adults, a delay period filled with sleep as compared with an equivalent delay filled with wakefulness may protect against memory interference (Abel & Bäuml, 2014; Ellenbogen, Hulbert, Stickgold, Dinges, & Thompson-Schill, 2006; Sheth, Varghese, & Truong, 2012; Spencer, Sunm, & Ivry, 2006; but see also Bailes, Caldwell, Wamsley, & Tucker, 2020; Pöhlchen, Pawlizki, Gais, & Schönauer, 2021), likely due to sleep-related consolidation. Designs with delays up to 24 h demonstrate that increases in the length of the delay (e.g., 30 min vs. 12 or 24 h) between encoding and test lead to decrements in performance (Payne et al., 2012; Takashima et al., 2009), but this decline is attenuated by sleep during the delay, especially if it directly follows learning (Payne et al., 2012).

There is some evidence that children's memory also benefits from a sleep-filled delay period (Backhaus, Hoeckesfeld, Born, Hohagen, & Junghanns, 2008; Kurdziel, Duclos, & Spencer, 2013). We also know that children as young as 18 months can recall actions learned 24 h earlier (Herbert & Hayne, 2000), but age-related improvements in memory performance after a delay continue across early childhood (Loucks & Price, 2019; Morgan & Hayne, 2010; Saragosa-Harris et al., 2021). Because delays of 24 h increase forgetting, but sleep can attenuate forgetting, how a delay of 24 h that includes a night of sleep might affect children's performance is unclear.

Do overlap and delay interact?

How delays interact with differences in overlap across contexts is not well explored. In an important study, Darby and Sloutsky (2015) tested 4- and 5-year-old children on memory for object pairs they had previously learned to a criterion. One group of children was tested immediately after encoding, whereas the other group was tested after a 48-h delay. The group of children who experienced the delay had superior memory (not just memory maintenance) compared with children who were tested immediately for the overlapping pairs that are most susceptible to interference. Darby and Sloutsky suggested that, for young children who are highly susceptible to interference, an offline rest period allowed for consolidation that supported stable and precise configural memory traces. Whether these effects differ by age is not known because the authors did not compare across different age groups. In addition, whether these effects generalize to episodic forms of declarative memory is not known; the children in Darby and Sloutsky (2015) study were exposed to the same associative pairings until a learning criterion was met.

Method

In this study, we sought to examine whether a 24-h delay might stabilize memory for episodic material and whether such effects might differ by age and by the degree of overlap between associations. We studied younger and older children in a within-participants design using a one-shot learning task with both overlapping and unique associative pairs (see Fig. 1).

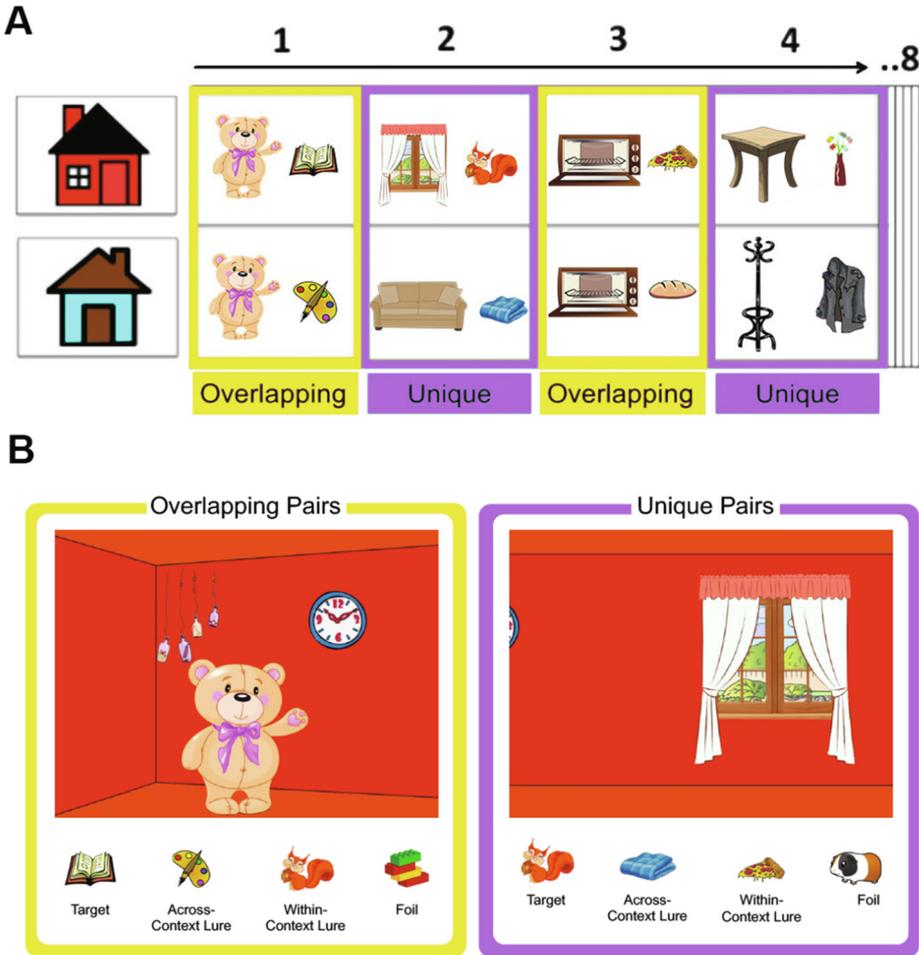


Fig. 1. Schematic depiction of the encoding phase (A) and test phase (B). (A) Animations at encoding included two locations (e.g., a red house and a blue house). Each version contained 4 overlapping item pairs (in yellow) and 4 unique item pairs (in purple) for a total of 8 pairs per version and 16 pairs per animation. (B) At test, participants were shown a still image of an item from one of the locations and four choices for the item with which it was paired; the options were a target, an across-context lure, a within-context lure, and a foil. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Participants

A total of 33 4-year-old children (19 girls and 14 boys; $M_{age} = 57.29 \pm 7.12$ months) and 32 6-year-old children (16 girls and 16 boys; $M_{age} = 74.36 \pm 8.78$ months) were recruited from Philadelphia, a large city in the eastern United States, and the surrounding suburbs. Children who participated in the study did not have any psychological, neurological, or developmental disorders, as reported by their parents. Informed consent was obtained from each child’s parent or guardian. An additional 10 children participated but were not included in the data analyses due to failure to complete at-home testing ($n = 4$), participant noncompliance with at-home procedure instructions ($n = 2$), not meeting inclusion criteria (e.g., 5 years of age, developmental disorder) ($n = 3$), and experimenter error ($n = 1$). Of the 65 children who met inclusion criteria and completed at least one of the two tests at each time point (immediate and delayed), 5 children (4 4-year-olds; 2 girls and 2 boys; $M_{age} = 54.5$

4 ± 4.72 months) failed to perform at above-chance levels on the experimental procedure at immediate test—indicating that they were guessing or responding randomly—and were removed from subsequent analyses (chance performance = proportion correct of .25 or less). In addition, 1 child's standard score on the Kaufman Brief Intelligence Test–Second Edition (KBIT-2, a standardized measure of verbal intelligence; see below) was more than 2 standard deviations below the mean, and this child (a 6-year-old girl) was also removed from subsequent analyses. Therefore, our final sample consisted of 59 children: 29 4-year-olds (18 girls and 11 boys; $M_{\text{age}} = 57.53 \pm 7.44$ months) and 30 6-year-olds (14 girls and 16 boys; $M_{\text{age}} = 74.90 \pm 7.64$ months).

Materials

We developed a novel memory task based on previous studies (Newcombe et al., 2014; Ngo et al., 2018). The stimuli consisted of four animated sequences in four different virtual environments—houses, parks, oceans, and fairs—that were created using Adobe Photoshop and Microsoft PowerPoint. Each animation consisted of a tour of two locations (e.g., a red house and a blue house) that had different salient background colors and ornamental details. Each location contained eight associated pairs (e.g., bear–book), with a total of 16 associations per animation. In each animation, half of the associations were assigned as *overlapping* (AB–AC), whereas the other half were assigned as *unique* (AB–CD). The overlapping pairs were made up of one common item (e.g., bear), an item that appeared in both locations, and one unique item (e.g., book, paint), an item that appeared in only one location. The unique pairs were made up of two items that were unique across locations; these pairs were seen in the same place within the two locations, but neither item overlapped with the corresponding pair in the other location (e.g., squirrel–window in red house living room, blanket–couch in blue house living room; see Fig. 1A). Within each animation, unique and overlapping pairs appeared in an interleaved fashion.

Procedure

Relational memory task

All participants were tested individually and randomly assigned to the different versions of the animations. Our goal was to design a task that would allow for two encoding phases to be tested at two different time points: immediately and after a delay. Therefore, we combined four animated videos that take children on “tours” of different places—house, park, fair, and sea—into two sets; one set of two videos watched sequentially and tested immediately, and the other set of two videos watched sequentially following the immediate test and later tested after a delay. At the beginning of each animation, prerecorded audio informed participants that they would visit two different locations and would need to remember the things they saw in each location. There were two locations per animation that were designed to be highly similar (e.g., red house and blue house, purple park and white park). Eight associated pairs were presented in each location, resulting in a total of 16 pairs per animation. Each association was presented statically for 5 s with 12 transition frames (100 ms/frame) before the next association appeared. The appearance of the paired item was accompanied by an audio clip of a chime to signal that an item was appearing on the screen. The order of the four animations and the two locations within each animation was counterbalanced across participants. Each encoding phase consisted of watching two animations sequentially. Each test phase followed each encoding phase.

All tests were administered via Qualtrics. There were two test phases. The first test phase focused on the first set of two animations and took place in the lab immediately following the first encoding phase. Then, the second encoding phase was administered. The second test phase focused on the second set of two animations and occurred the next day, administered in the home of the child by a parent or guardian. Links to the tests for the second set of animations were e-mailed to parents/guardians. Parents/guardians were required to administer the test phase to their children on a desktop or laptop computer (they could not use a smartphone or tablet) to ensure that the stimuli would be fully visible on the screen and the resolution would be similar to when the task was shown in the lab. We chose to have the delayed test phase take place at home rather than in our lab to reduce attrition; our attrition

rate was only 5.3%. Note that parents did not watch the animations with their children while in the lab, and therefore they could not bias their children's responses at home with knowledge of the correct answers.

Each test phase consisted of two sets of 16 four-alternative forced-choice trials, one for each of the two animations; in other words, each pair of items was tested only once, with half of the pairs being tested immediately after encoding (32 total pairs) and half being tested after a delay (32 total pairs). The test trials for each animation were presented in a pseudorandomized order; we created 8 versions of each animation test (8 red house–blue house animations, 8 white park–purple park animations, etc.) that contained different paired items in order to control for any effects on memory of the specific item. For each of these versions, the test question order was randomized in advance but then fixed. At test, participants were presented with a static screenshot of one item of each item pair in its location (e.g., bear in the red house) with four options shown below (see Fig. 1B). The four options were the target, an across-context lure, a within-context lure, and a foil. Targets (e.g., book) were the items that were indeed paired with the corresponding item shown in the static image (e.g., bear) in a specific location (e.g., red house). Across-context lures (e.g., paint palette) were the items paired with the corresponding common or unique item (e.g., bear) but seen in the other location (e.g., blue house). Within-context lures (e.g., squirrel) were unique items seen in the correct location (e.g., red house) but not paired with the common element (e.g., bear). Foils were novel items not seen at encoding. We were particularly interested in the across-context lures because we expected differences in performance depending on whether the tested item pair contained unique or overlapping elements because the latter type of pair included one item that was the same across contexts and would be more likely to result in memory interference.

Participants were asked to choose the item that they saw paired with the depicted item in a given scene by pointing to one of the four options presented on the screen. The experimenter or parent/guardian would select the corresponding button beneath that item and then move to the next question. Responses were automatically recorded by Qualtrics. All the tested items were counterbalanced such that they were assigned as each test item type an equal number of times across participants. The entire procedure, including encoding and the immediate test, took approximately 30 min. The at-home (delayed) testing took approximately 10 min.

Test of verbal intelligence: Kaufman Brief Intelligence Test–Second Edition

We administered the Verbal Knowledge and Riddles subtests of the KBIT-2 (Kaufman, 2004) to assess verbal intelligence. The KBIT-2 was administered in the lab before the relational memory task (detailed above). The KBIT-2 allowed us to control for any potential differences in memory performance that might be due to differences in verbal intelligence. For the Verbal Knowledge subtest, 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., “Point to ‘the one that goes with thunder’”—the child points to a picture of lightning). For the Riddles subtest, children responded verbally with a one-word answer to verbal riddles (e.g., “What is very far away, can only be seen at night, and twinkles in the sky?”—the child responds “star” or “planet”). The task was terminated when children incorrectly answered four consecutive questions. Standard scores were calculated based on age. The administration of the KBIT-2 took 10–20 min. In our initial examination of the data, we found that KBIT-2 scores were significantly associated with memory performance as well as with selection of all three error types (see Results), so it remained in our analyses as a covariate to account for variance in our outcome variables that was due to verbal intelligence.

Questionnaire data

Parents/guardians of children completed a demographics form asking for information such as parental education and children's gender, race, and ethnicity. Because our delay window included an overnight period, we also collected the Child Sleep Habits Questionnaire (CSHQ; Owens, Spirito, & McGuinn, 2000) completed by parents/guardians. This survey asked about their children's sleep habits, such as average bedtime and wake time, as well as sleep difficulties, such as excessive daytime sleepiness and difficulties in falling asleep. The demographic variables as well as sleep habits and

difficulties were not significantly associated with memory performance (see Results), so they were not included in our final analyses.

Design

Our design was a 2 (Time Point: immediate or delayed) \times 2 (Pair Type: unique or overlapping) \times 2 (Age: 4 or 6 years) mixed design, with time point and pair type manipulated within participants and age a between-participants variable.

Analysis

Performance on our relational memory task was measured for each of the four tests separately as the proportion of target items selected out of 16. (For 6 children in our sample, there was a technological error on one test that caused one question to yield a blank response; for these children, that test is calculated out of 15 items rather than 16.) We also calculated proportion selected for across-context lures, within-context lures, and foils to assess the types of errors children were making when they were not correctly selecting the target item. For every participant, the proportion of test item selection (target, across-context lure, within-context lure, and foil) was calculated for each pair type (32 unique vs. 32 overlapping) and for each testing session (32 immediate vs. 32 delay). (Due to experimenter error, 4 children had only one of the two tests administered at one of the two time points; i.e., they completed only three of the four total tests. We determined that nothing was systematically different about these children or their performance when compared with the rest of the group.)

We used JASP Version 0.13.1 for all our analyses. First, we evaluated memory performance by conducting a repeated-measures analysis of variance (ANOVA), with proportion of targets selected (i.e., accuracy) being our outcome variable. Within-participant factors were pair type and time point, and the between-participants factor was age. We conducted three additional ANOVAs with the same within-participant and between-participants factors but with proportions of across-context lures, within-context lures, and foils selected as the outcome variables.

For ANOVAs in which the results indicated significant main effects or interactions, we conducted post hoc tests with Holm-corrected p values to evaluate directionality of the pairwise effects and/or the nature of the interaction. Graphs of our results were created via RStudio for MacOS Version 1.1.463 using the *tidyr* and *ggplot2* packages. To account for the effect of our covariate (KBIT-2) when plotting our data, we regressed KBIT-2 scores on all our outcome variables and plotted the standardized residual values on the y axis; these values represent the proportion selected of each response type after removing variance accounted for by KBIT-2 scores (see Figs. 2 and 3 in Results).

Results

Preliminary analyses

Four children were removed from our sample for performing below chance when tested immediately (described in Method), but of the remaining children, average performance was well above the chance level of 25% (59% for 4-year-olds and 78% for 6-year-olds). According to independent-samples t tests, male and female participants did not differ on any of our outcomes variables of interest (selection of targets, across-context lures, within-context lures, and foils; all $ps > .05$), so effects of gender were not further considered.

"Sleep problems" (a summary score from the CSHQ of children's problematic sleep behaviors) and total number of hours slept in a 24-h window (nighttime hours plus nap hours if applicable; $M = 10.266$, $SD = 0.910$) were not significantly correlated with any of our outcome variables of interest (all $ps > .05$). The number of hours a child napped each day ($M = 0.216$, $SD = 0.479$) was not significantly correlated with across-context lure selection or foil selection (both $ps > .05$) but was significantly correlated with target selection, $r(57) = -.266$, $p = .042$, and within-context lure selection, $r(57) = .403$, $p = .002$. However, our sample included only 13 children who napped, all of whom were

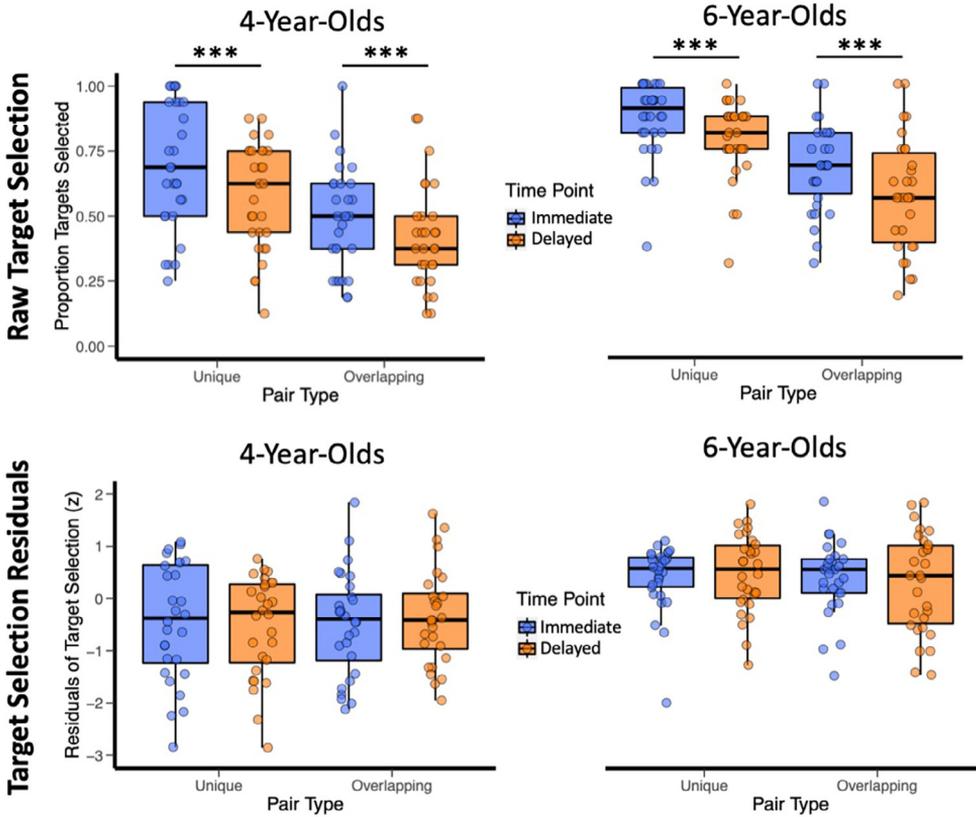


Fig. 2. Target selection (top panel) and residuals of target selection after removing variance accounted for by the Kaufman Brief Intelligence Test–Second Edition (KBIT-2) (bottom panel) for 4- and 6-year-olds plotted separately by time point and pair type. Before KBIT-2 was included in the analysis of variance, there was a significant effect of time point, but when KBIT-2 was included, the effect of time point was not significant. *** $p < .001$.

4 years old. We created a data subset of only 4-year-old participants, and there was no correlation between number of hours napped and any of our variables of interest for this subset of participants (all $ps > .05$). In addition, when number of hours napped was included in our ANOVAs of the full dataset as a covariate, it did not change the significance of any of the relationships between our other predictors and outcome variables of interest. Therefore, none of the sleep variables was included in our analyses.

Lastly, KBIT-2 score was significantly correlated with overall memory performance (i.e. target selection), $r(57) = .445, p < .001$, across-context lure selection, $r(57) = -.304, p = .020$, and within-context lure selection, $r(57) = -.452, p < .001$, and the correlation with foil selection was trending toward significance, $r(57) = -.255, p = .053$. Its inclusion in our ANOVAs also changed the relationship between some predictors and outcomes of interest. Therefore, KBIT-2 was included as a covariate in all the following analyses.

Main analyses

We conducted 2 (Age) \times 2 (Pair Type) \times 2 (Time Point) mixed ANOVAs for each dependent variable separately. Our goal was to evaluate the impact of age, the difficulty of each relational pair based on its overlap across contexts, and the time at which children were tested—immediately or after a 24-h

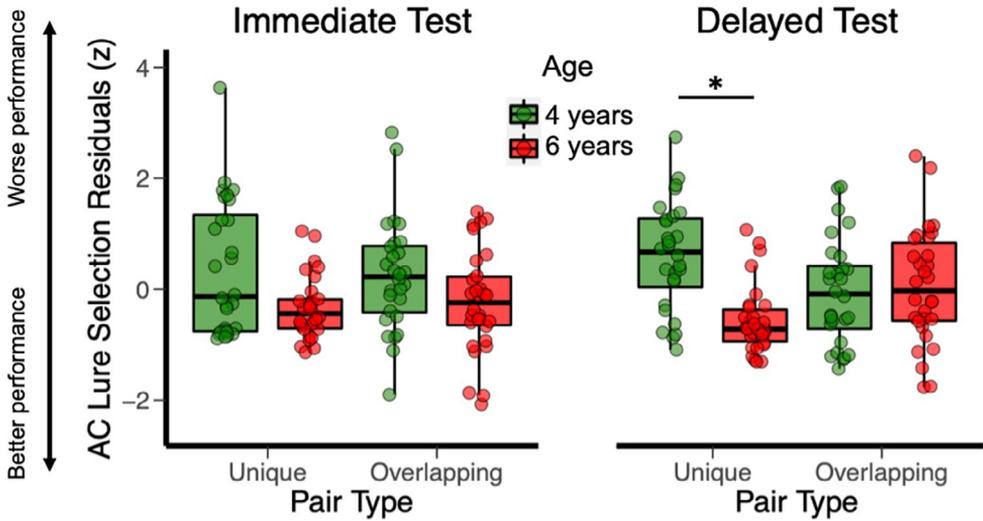


Fig. 3. Across-context (AC) lure errors plotted and grouped by age, time point, and pair type after removing variance accounted for by Kaufman Brief Intelligence Test–Second Edition (KBIT-2) scores. The significant age by time point by pair type three-way interaction is highlighted with a significance bar. Note that selection of a lure was an error, so higher values indicate worse performance. To obtain the residual values presented on the y axis, we regressed KBIT-2 scores on proportion selected at each time point for each pair type. * $p < .05$.

delay—on our dependent variables of interest. These dependent variables were target selection (i.e., overall memory accuracy), across-context lure selection, within-context lure selection, and foil selection (described in the “Procedure” section of Method). We analyzed target selection as the most basic index of memory retention for our stimuli, but we were particularly interested in across-context lure selection because across-context errors indicate intact item–item associative memory, with a specific failure of binding the association to its context, a crucial index of episodic memory. Higher values for target selection indicate better performance, whereas lower values for both types of lures and foils indicate better performance.

For target selection, we found a main effect of age, $F(1, 55) = 25.911, p < .001, \eta_p^2 = .320$, with a post hoc test (using the Holm correction to adjust p) showing that 6-year-olds outperformed 4-year-olds, $t = 5.090, p_{\text{Holm}} < .001$. There was also a main effect of pair type, $F(1, 55) = 4.337, p = .042, \eta_p^2 = .073$, with a post hoc test demonstrating that unique pairs were better remembered than overlapping pairs, $t = 12.593, p_{\text{Holm}} < .001$. The main effect of time point was trending toward significance, $F(1, 55) = 3.425, p = .070, \eta_p^2 = .059$. There were no significant interactions (all $ps > .14$). This suggests that older children performed better than younger children, and both groups of children better remembered the unique pairs than the overlapping pairs.

It is worth noting that there was a significant effect of time point in an ANOVA where KBIT-2 was not included as a covariate, $F(1, 57) = 165.997, p < .001, \eta_p^2 = .744$, with both groups of children performing better immediately than after a delay (see Fig. 2). However, this effect did not hold with KBIT-2 included as a covariate (see Fig. 2), indicating that children’s forgetting across a 24-h window could be accounted for at least in part by verbal intelligence.

For across-context lure errors, we found a main effect of age, $F(1, 55) = 8.826, p = .004, \eta_p^2 = .138$, a trend toward a main effect of pair type, $F(1, 55) = 3.118, p = .083, \eta_p^2 = .054$, and no main effect of time point, $F(1, 55) = 0.378, p = .541, \eta_p^2 = .007$. In addition, there was a significant age by pair type interaction, $F(1, 55) = 4.14, p = .047, \eta_p^2 = .070$, with a Holm-corrected post hoc test revealing that the interaction was driven by 4-year-olds making more across-context errors than 6-year-olds on unique pairs, $t = 3.60, p_{\text{Holm}} < .001$, but not on overlapping pairs, $t = 1.13, p_{\text{Holm}} = .261$.

Importantly, this pattern further interacted with time point, such that there was a significant age by pair type by time point three-way interaction, $F(1, 55) = 5.064$, $p = .028$, $\eta_p^2 = .084$, with post hoc tests showing that this interaction was driven by 4-year-olds making more across-context errors than 6-year-olds on unique pairs after a delay, $t = 3.652$, $p_{\text{Holm}} = .004$, but not when tested immediately, $t = 2.055$, $p_{\text{Holm}} = .370$ (see Fig. 3). In contrast, the across-context errors on the overlapping pairs did not differ between the two age groups on either the immediate test, $t = 2.223$, $p_{\text{Holm}} = .273$, or the delayed test, $t = 0.544$, $p_{\text{Holm}} = 1.000$ (see Fig. 3). This result indicates that the delay period affected relational binding error rates differentially for younger versus older children and that this effect depends on whether or not the relational pairs share an overlapping constituent across conditions.

For within-context lures, there was only a main effect of age, $F(1, 55) = 19.178$, $p < .001$, $\eta_p^2 = .259$, with post hoc tests revealing that 4-year-olds made more within-context lure errors than 6-year-olds, $t = 4.379$, $p_{\text{Holm}} < .001$, whereas pair type, $F(1, 55) = 0.680$, $p = .413$, $\eta_p^2 = .012$, and time point, $F(1, 55) = 0.071$, $p = .790$, $\eta_p^2 = .001$, did not have an impact on the frequency of within-context lure selection. There were no significant interactions (all $ps > .26$).

Finally, we conducted a repeated-measures ANOVA for foils. There were main effects of age, $F(1, 55) = 8.041$, $p = .006$, $\eta_p^2 = .128$, and time point, $F(1, 55) = 6.013$, $p = .017$, $\eta_p^2 = .099$, and there was a trend toward significance for the main effect of pair type, $F(1, 55) = 3.426$, $p = .070$, $\eta_p^2 = .059$. Post hoc tests of the main effects demonstrated that 4-year-olds made more foil errors than 6-year-olds, $t = 2.836$, $p_{\text{Holm}} = .006$, and children made more foil errors when tested after a delay than when tested immediately, $t = 3.500$, $p_{\text{Holm}} < .001$. There was also a pair type by age interaction, $F(1, 55) = 4.211$, $p = .045$, $\eta_p^2 = .071$, with a post hoc test showing that 6-year-olds' foil selection errors were greater for unique pairs than for overlapping pairs, $t = 3.446$, $p_{\text{Holm}} = .005$, whereas 4-year-olds' foil errors did not differ by pair type, $t = 0.401$, $p_{\text{Holm}} = .690$.

Discussion

This study investigated several questions about age-related differences in binding capacities. First, we asked whether associations with overlapping constituents are more challenging than associations of unique pairs and whether the difference is especially marked for younger children. We found that accuracy is indeed lower with overlapping pairs but that the effect does not differ for 4- and 6-year-olds. Second, we asked whether delay reduced accuracy and whether any reduction differed by age. We found only a marginal effect of delay and no interaction with age. Third, we asked whether across-context lure errors varied by overlap, age, delay, and their interactions. We found a triple interaction. The 4-year-olds did surprisingly well on avoiding such lures for overlapping pairs after a delay, although on unique pairs 4-year-olds made more errors than 6-year-olds. These findings are reminiscent of data reported by Darby and Sloutsky (2015), who found that 4- and 5-year-old children tested after a 48-h delay retained overlapping associated pairs to a greater extent than children tested immediately. We did not find an actual boost in relational memory for overlapping pairs, but we did see a pattern of memory *maintenance* for these complex relational structures after a delay in younger children. Older children conversely showed maintenance for both unique and overlapping pairs, with no statistically significant increases in errors across a delay period for either pair type.

Work in adults suggests that memory replay—a process that occurs during rest after learning—tends to replay the “weakest” memories or those most vulnerable to forgetting (Schapiro, McDevitt, Rogers, Mednick, & Norman, 2018). In our study, the most vulnerable memories would be those for the overlapping pairs, which are subject to the highest degree of interference across learning episodes. We do not have a clear picture of how memory replay functions in young children, but we speculate that this process does occur in these children in some capacity. The still-developing systems of the youngest children cannot yet retain all learned associations, so the weakest memories that are often the most replayed in adults (Schapiro et al., 2018) might have been the *bulk* of the memories that were replayed in the 4-year-olds' brains during our delay period and then subsequently recalled. The 6-year-olds, on the other hand, have more mature memory systems and thus are able to successfully maintain memory for both types of pairs even after a delay. Numerically, 6-year-olds' across-context error rates on both types of pairs remained relatively stable from the immediate test to the

delayed test, whereas the error rate was maintained only on the overlapping pairs for 4-year-olds. Due to the episodic nature of our task and the relational binding of item to context that is required to perform the task successfully, it is likely that relatively late-maturing hippocampal subregions are related to this age difference in mnemonic performance. This is supported by prior work showing associations between development of these hippocampal subregions and episodic memory performance in early childhood (Canada et al., 2019; Riggins et al., 2018). Although 6-year-olds outperformed 4-year-olds, memory performance was still above chance after a 24-h delay for the younger children, demonstrating that they can retain relational memories over time. Recent work has shown that 4- and 5-year-olds can retain associated pairs of items beyond 24-h up to one-week delays, whereas 3-year-olds' performance is below chance after a week (Saragosa-Harris et al., 2021). These findings provide further evidence that the mnemonic system shows critical developmental across early childhood, with episodic memory capacity being fragile before age 4 and more robust by age 6.

There are some important differences in the design of our study compared with prior studies. First, most studies of children's memory consolidation use a task design in which children are repeatedly exposed to static pairs of items until they reach a prespecified criterion (e.g., Backhaus et al., 2008; Darby & Sloutsky, 2015; Kurdziel et al., 2013). This kind of procedure is more akin to a semantic learning paradigm than an episodic one. We instead implemented a single-acquisition learning paradigm because in real life episodic memory typically involves experiencing an event only once. The average accuracy for children on the immediate test for our task was well above chance level, so although children were not trained to a specific performance criterion, they were still able to learn the associations in our task.

The amount of initial learning bears on an important broader question: What kinds of memory representations benefit from stabilization over a delay period? Although we often associate the ability to distinguish between similar contexts with episodic memory, overlap is also common in semantic memory; for example, a cow and a horse both live on a farm. Semantic memory has an earlier developmental trajectory than episodic memory (Drumme & Newcombe, 2002) and may be acquired through the process of generalization, relying on the cortex and parts of the hippocampus that are early developing (Keresztes, Ngo, Lindenberger, Werkle-Bergner, & Newcombe, 2018). Generalizing across multiple experiences and incorporating new learning into existing semantic stores is often hypothesized to be a pivotal role of consolidation during delay periods and is relevant for processes such as learning new concepts in school (e.g., Vlach & Sandhofer, 2012) and language generalization for word learning in younger children (e.g., Werchan & Gómez, 2014). Perhaps the reported benefits of delay periods on memory in children are more marked for tasks that tap semantic memory systems. Some tasks may superficially appear to be episodic because they involve paired associates but are actually semantic because the tasks use conceptually rich verbalizable stimuli that are repeated many times (e.g., Backhaus et al., 2008; Darby & Sloutsky, 2015; Kurdziel et al., 2013).

In conclusion, this study suggests two important similarities in episodic memory across 4 to 6 years of age. Although 6-year-olds perform better than 4-year-olds overall, overlapping pairs are similarly difficult at the two ages and age effects do not differ across a 24-h delay for memory accuracy. In addition, the data show that a delay window allowing for a period of consolidation does not always provide a protective benefit for relational memories; accuracy for unique and overlapping pairs declined equivalently across a delay. However, errors for across-context lures were greater in younger children than in older children only after the delay and only for non-overlapping pairs, suggesting that younger children's less-developed memory systems might best consolidate the stimuli most sensitive to interference. These data should affect our understanding of the source of age-related change in episodic memory and how varying delays affect memory in children.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Jelani Medford, Rebecca Adler, Linda Hoffman, Elizabeth Eberts, Ying Lin, and Alexandra Cavallo for their assistance with stimuli development and data collection. This work was supported by National Institutes of Health grants to I. R. Olson (R01 MH091113, R21 HD098509, and R56 MH091113), N. S. Newcombe and I. R. Olson (R01 HD099165), and C. T. Ngo (F31 HD090872) as well as by funding from Temple University to N. S. Newcombe (OVPR 161706-24607-02). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Institutes of Health.

1. Data Sharing and Data Accessibility1.1.

The study reported in this article was not formally preregistered. De-identified data with explanations of variables, files used for task procedures, and documents detailing experimental design choices (e.g., counterbalancing stimuli) have been made available on a permanent third-party archive and can be accessed at the Open Science Framework ([https://osf.io/ckteb/?view_only = 21aa31a17c0341dcbaa0c722984fc574](https://osf.io/ckteb/?view_only=21aa31a17c0341dcbaa0c722984fc574)).

References

- Abel, M., & Bäuml, K.-H. T. (2014). Sleep can reduce proactive interference. *Memory*, *22*, 332–339.
- Backhaus, J., Hoeckesfeld, R., Born, J., Hohagen, F., & Junghanns, K. (2008). Immediate as well as delayed post learning sleep but not wakefulness enhances declarative memory consolidation in children. *Neurobiology of Learning and Memory*, *89*, 76–80.
- Bailes, C., Caldwell, M., Wamsley, E., & Tucker, M. (2020). Does sleep protect memories against interference? A failure to replicate. *PLoS One*, *15*(2), e220419.
- Bauer, P. J., Larkina, M., & Deocampo, J. (2011). Early memory development. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (pp. 153–179). Hoboken, NJ: Wiley-Blackwell.
- Canada, K. L., Botdorf, M., & Riggins, T. (2020). Longitudinal development of hippocampal subregions from early- to mid-childhood. *Hippocampus*, *30*, 1098–1111.
- Canada, K. L., Ngo, C. T., Newcombe, N. S., Geng, F., & Riggins, T. (2019). It's all in the details: Relations between young children's developing pattern separation abilities and hippocampal subfield volumes. *Cerebral Cortex*, *29*, 3427–3433.
- Darby, K. P., & Sloutsky, V. M. (2015). When delays improve memory: Stabilizing memory in children may require time. *Psychological Science*, *26*, 1937–1946.
- DeMaster, D., Pathman, T., Lee, J. K., & Ghetti, S. (2014). Structural development of the hippocampus and episodic memory: Developmental differences along the anterior/posterior axis. *Cerebral Cortex*, *24*, 3036–3045.
- Drumme, A. B., & Newcombe, N. S. (2002). Developmental changes in source memory. *Developmental Science*, *5*, 502–513.
- Eacott, M. J., & Norman, G. (2004). Integrated memory for object, place, and context in rats: A possible model of episodic-like memory? *Journal of Neuroscience*, *24*, 1948–1953.
- Ellenbogen, J. M., Hulbert, J. C., Stickgold, R., Dinges, D. F., & Thompson-Schill, S. L. (2006). Interfering with theories of sleep and memory: Sleep, declarative memory, and associative interference. *Current Biology*, *16*, 1290–1294.
- Flin, R., Boon, J., Knox, A., & Bull, R. (1992). The effect of a five-month delay on children's and adults' eyewitness memory. *British Journal of Psychology*, *83*, 323–336.
- Gogtay, N., Hugent, T. F., Herman, D. H., Ordóñez, A., Greenstein, D., Hayashi, K. M., ... Thompson, P. M. (2006). Dynamic mapping of normal human hippocampal development. *Hippocampus*, *16*, 664–672.
- Herbert, J., & Hayne, H. (2000). Memory retrieval by 18–30-month-olds: Age-related changes in representational flexibility. *Developmental Psychology*, *36*, 473–484.
- Jitsumori, M., Wright, A. A., & Cook, R. G. (1988). Long-term proactive interference and novelty enhancement effects in monkey list memory. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*, 146–154.
- Kaufman, A. S. (2004). *Kaufman Brief Intelligence Test-Second Edition (KBIT-2)*. Circle Pines, MN: American Guidance Service.
- Keresztes, A., Ngo, C. T., Lindenberger, U., Werkle-Bergner, M., & Newcombe, N. S. (2018). Hippocampal maturation drives memory from generalization to specificity. *Trends in Cognitive Sciences*, *22*, 676–686.
- Krogsrud, S. K., Tamnes, C. K., Fjell, A. M., Amlien, I., Grydeland, H., Sulutvedt, U., ... Walhovd, K. B. (2014). Development of hippocampal subfield volumes from 4 to 22 years. *Human Brain Mapping*, *35*, 5646–5657.
- Kubo-Kawai, N., & Kawai, N. (2007). Interference effects by spatial proximity and age-related declines in spatial memory by Japanese monkeys (*Macaca fuscata*): Deficits in the combined use of multiple spatial cues. *Journal of Comparative Psychology*, *121*, 189–197.
- Kurdzziel, L. B. F., Duclos, K., & Spencer, R. M. C. (2013). Sleep spindles in midday naps enhance learning in preschool children. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 17267–17272.
- 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.
- Loucks, J., & Price, H. L. (2019). Memory for temporal order in action is slow developing, sensitive to deviant input, and supported by foundational cognitive processes. *Developmental Psychology*, *55*, 263–273.
- Morgan, K., & Hayne, H. (2010). Age-related changes in visual recognition memory during infancy and early childhood. *Developmental Psychobiology*, *53*, 157–165.
- Newcombe, N. S., Balcomb, F., Ferrara, K., Hansen, M., & Koski, J. (2014). Two rooms, two representations? Episodic-like memory in toddlers and preschoolers. *Developmental Science*, *17*, 743–756.

- Newcombe, N. S., Benear, S. L., Ngo, C. T., & Olson, I. R. (in press). Memory in infancy and childhood. In M. Kahana & A. Wagner (Eds.), *Handbook on human memory*. New York: Oxford University Press.
- Ngo, C. T., Horner, A. J., Newcombe, N. S., & Olson, I. R. (2019). Development of holistic episodic recollection. *Psychological Science*, *30*, 1696–1706.
- 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. (2018). The ontogeny of relational memory and pattern separation. In *Developmental Science*. Advance online publication. <https://doi.org/10.1111/desc.12556>.
- Owens, J. A., Spirito, A., & McGuinn, M. (2000). The Children's Sleep Habits Questionnaire (CSHQ): Psychometric properties of a survey instrument for school-aged children. *Sleep*, *23*, 1–9.
- Payne, J. D., Tucker, M. A., Ellenbogen, J. M., Wamsley, E. J., Walker, M. P., Schacter, D. L., & Stickgold, R. (2012). Memory for semantically related and unrelated declarative information: The benefit of sleep, the cost of wake. *PLoS ONE*, *7*(3) e33079.
- Pöhlchen, D., Pawlitzki, A., Gais, S., & Schönauer, M. (2021). Evidence against a large effect of sleep in protecting verbal memories from interference. *Journal of Sleep Research*, *30* e13042.
- Poole, D. A., & White, L. T. (1993). Two years later: Effect of question repetition and retention interval on the eyewitness testimony of children and adults. *Developmental Psychology*, *29*, 844–853.
- Prebble, S. C., Addis, D. R., & Tippett, L. J. (2013). Autobiographical memory and sense of self. *Psychological Bulletin*, *139*, 815–840.
- Riggins, T., Geng, F., Botdorf, M., Canada, K., Cox, L., & Hancock, G. R. (2018). Protracted hippocampal development is associated with age-related improvements in memory during early childhood. *NeuroImage*, *174*, 127–137.
- Saragosa-Harris, N. M., Cohen, A. O., Shen, X., Sardar, H., Alberini, C. M., & Hartley, C. A. (2021). Associative memory persistence in 3- to 5-year-olds. *Developmental Science*, e13105. <https://doi.org/10.1111/desc.13105>.
- Schapiro, A. C., McDevitt, E. A., Rogers, T. T., Mednick, S. C., & Norman, K. A. (2018). Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance. *Nature Communications*, *9*, 3920.
- Sheth, B. R., Varghese, R., & Truong, T. (2012). Sleep shelters verbal memory from different kinds of interference. *Sleep*, *35*, 985–996.
- 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.
- Spencer, R. M. C., Sunm, M., & Ivry, R. B. (2006). Sleep-dependent consolidation of contextual learning. *Current Biology*, *16*, 1001–1005.
- Szpunar, K. K., Addis, D. R., McLelland, V. C., & Schacter, D. L. (2013). Memories of the future: New insights into the adaptive value of episodic memory. *Frontiers in Behavioral Neuroscience*, *7*. <https://doi.org/10.3389/fnbeh.2013.00047>.
- Takashima, A., Nieuwenhuis, I. L. C., Jensen, O., Talamini, L. M., Rijpkema, M., & Fernández, G. (2009). Shift from hippocampal to neocortical centered retrieval network with consolidation. *Journal of Neuroscience*, *29*, 10087–10093.
- Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Reviews in Psychology*, *53*, 1–25.
- Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development*, *83*, 1137–1144.
- Werchan, D. M., & Gómez, R. L. (2014). Wakefulness (not sleep) promotes generalization of word learning in 2.5-year-old children. *Child Development*, *85*, 429–436.
- Yim, H., Dennis, S. J., & Sloutsky, V. M. (2013). The development of episodic memory: Items, contexts, and relations. *Psychological Science*, *24*, 2163–2172.