



This postprint was originally published by Elsevier as:

Beneat, S. L., Ngo, C. T., Olson, I. R., & Newcombe, N. S. (2021). Understanding relational binding in early childhood: Interacting effects of overlap and delay. *Journal of Experimental Child Psychology*, 208, Article 105152.

<https://doi.org/10.1016/j.jecp.2021.105152>.

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, nontransferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. By using this particular document, you accept the above-stated conditions of use.

Provided by:

Max Planck Institute for Human Development
Library and Research Information
library@mpib-berlin.mpg.de

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

Susan L. Benear¹, Chi T. Ngo², Ingrid R. Olson¹, Nora S. Newcombe¹

¹ Department of Psychology, Temple University, USA

² Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin,
Germany

Acknowledgements

We would like to 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 Institute of Health grants to I. Olson [R01 MH091113, R21 HD098509, and R56 MH091113], N. Newcombe and I. Olson [R01 HD099165], and to C.T. Ngo [F31 HD090872] as well as by funding from Temple University to Dr. Nora 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.

Data Sharing and Data Accessibility

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, etc.) have been made available on a permanent third-party archive and can be accessed at https://osf.io/ckteb/?view_only=21aa31a17c0341dcbaa0c722984fc574.

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 (2013) suggested that extensive overlap is more problematic for younger children, and Darby and Sloutsky (2015) reported that a 48-hour delay period actually improves children's memory for overlapping pairs of items. In this 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-hour delay. As expected, older children performed better than younger children and both age groups performed worse on overlapping pairs. Surprisingly, the 24-hour 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 the younger children.

Keywords: episodic memory, development, delayed memory, memory interference, relational binding

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

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 age 4, 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 between the ages of 4 and 6 years (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 between 4 and 6 years, 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 paper, 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 the 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 only useful if recalled beyond the brief window just after an event occurs.

Overlap

Younger children have difficulty 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, Newcombe, & Olson, 2018; Ngo, Lin et al., 2019), 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 the items from the first list are all reused but the pairings are shuffled (AB-ABr). For example, in the AB-CD condition, List 1 = bike-cup, couch-cat, List 2 = backpack-fork, umbrella-football; AB-AC condition, List 1 = bike-cup, couch-cat, List 2 = bike-fork, couch-football; AB-ABr condition, List 1 = bike-cup, couch-

cat, List 2 = bike-cat, 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.'s (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 this 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 hours 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 a day or two. In adults, a delay period filled with sleep as compared to 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; 2020), likely due to sleep-related consolidation. Designs with delays up to 24 hours demonstrate that increases in the length of the delay (e.g. 30 minutes versus 12 or 24 hours) 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 old can recall actions learned 24 hours prior (Herbert & Hayne, 2000), but age-related improvements in memory performance after a delay continue across early childhood (Loucks & Price, 2019; Morgan & Hayne, 2010). Since delays of 24 hours increase forgetting, but sleep can attenuate forgetting, how a delay of 24 hours 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 while the other group was tested after a 48-hour delay. The group who experienced the delay had superior memory (not just memory maintenance) compared to children who were tested immediately, for the overlapping pairs that are most susceptible to interference. Darby and Sloutsky (2015) suggest 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 since 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's (2015) study were exposed to the same associative pairings until a learning criterion was met.

Methods

In this study, we sought to examine whether a 24-hour 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-subjects design using a one-shot learning task with both overlapping and unique associative pairs (see Figure 1).

Participants

A total of 33 4-year-old children (19 females, 14 males; $M_{\text{month}} = 57.29 \pm 7.12$) and 32 6-year-old children (16 females, 16 males; $M_{\text{month}} = 74.36 \pm 8.78$) were recruited from Philadelphia and the surrounding suburbs. Children who participated in the study did not have any psychological, neurological, or developmental disorders, as reported by a parent. Informed consent was obtained from each child's parent or guardian. Ten additional children participated but were not included in the data analyses due to failure to complete at-home testing ($n=4$), participant non-compliance with at-home procedure instructions ($n=2$), child not meeting inclusion criteria (e.g. age 5, 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), five children (four 4-year-olds; two females, two males; $M_{\text{month}} = 54.54 \pm 4.72$) failed to perform at above-chance levels on the experimental procedure at immediate test—indicating they were guessing or responding randomly—and were removed from subsequent analyses (chance performance = proportion correct of .25 or less). Additionally, one child's standard score on the KBIT-2 (a standardized measure of verbal intelligence; see *Procedure*, section 2) was more than two standard deviations below the mean, and this child (6 years old; female) was also removed from subsequent analyses. Therefore, our final sample consisted of 59 children—29 4-year-olds (18 females, 11 males; $M_{\text{month}} = 57.53 \pm 7.44$) and 30 6-year-olds (14 females, 16 males; $M_{\text{month}} = 74.90 \pm 7.64$).

Materials

We developed a novel memory task based on previous studies (Ngo et al., 2018; Newcombe et al., 2014). 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 and a blue house), which 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 only appeared in 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 Figure 1a). Within each animation, unique and overlapping pairs appeared in an interleaved fashion.

Procedure

1. 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, pre-recorded audio informed participants they would visit two different locations and would have to remember the things they saw in each location. There were two locations per animation, which 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 5s with 12 transition frames (100ms/frame) before the next association appeared. The appearance of

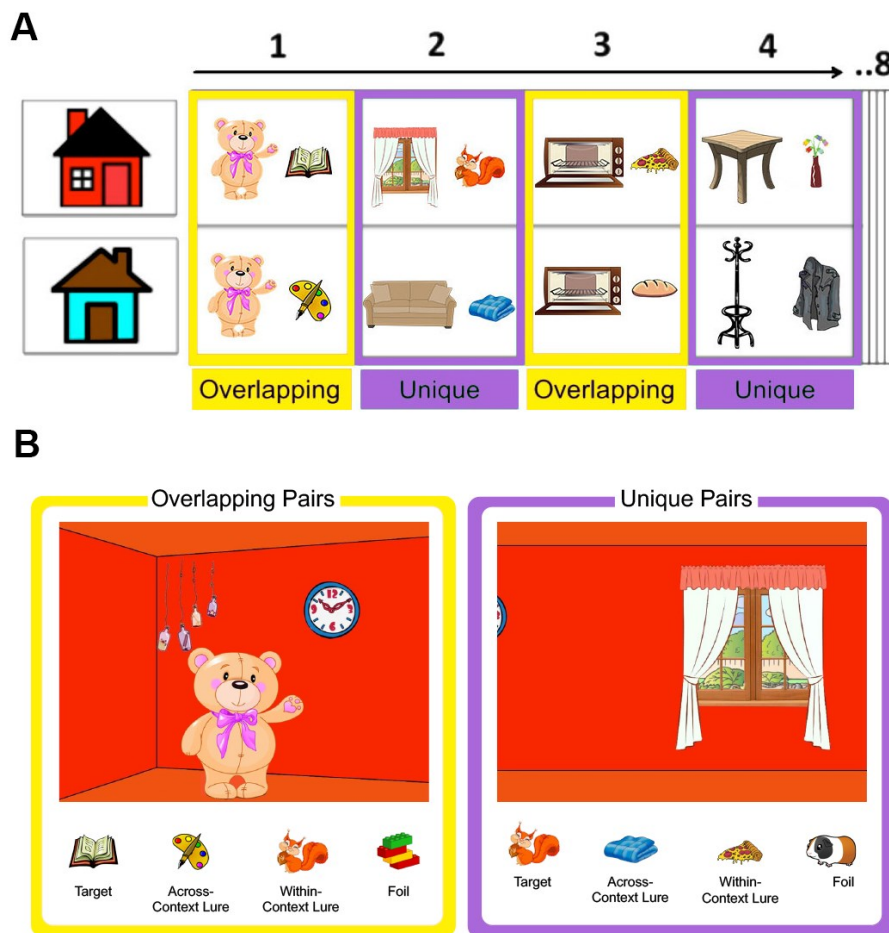


Figure 1. A schematic depiction of the encoding (A) and test phase (B). (A) Animations at encoding include two locations (e.g., a red and a blue house). Each version contains 4 overlapping (in yellow) and 4 unique (in purple) item pairs, for a total of 8 pairs per version and 16 pairs per animation. (B) At test, participants are 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 are a target, across-context lure, within-context lure, and foil.

the paired item was accompanied by an audio clip of a chime to signal that an item was appearing on 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 emailed to parents/guardians. Parents/guardians were required to administer the test phase to their child on a desktop or laptop computer (they could not use a smartphone or tablet) to ensure the stimuli would be fully visible on screen and the resolution would be similar to when the task was shown in 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 could not bias their children's responses at home with knowledge of the correct answer.

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 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 beneath (see Figure 1b). The four options included 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 that were 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, as 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 minutes. The at-home (delayed) testing took approximately 10 minutes.

2. Test of verbal intelligence: Kaufman Brief Intelligence Test, Second Edition (KBIT-2)

We administered the Verbal Knowledge and Riddles subtests of the Kaufman Brief Intelligence Test, Second Edition (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’” – child points to 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?” — child responds, “star”, “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 minutes. 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.

3. Questionnaire data

Parents/guardians of children completed a demographics form asking for information such as parental education, and the child'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 child's sleep habits—such as average bedtime and wake time—as well as sleep difficulties—such as excessive daytime sleepiness and trouble 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, delayed) x 2 (pair type: unique, overlapping) x 2 (age: 4, 6) mixed design, with time point and pair type manipulated within subjects and age a between-subjects variable.

Analysis

Performance on our relational memory task was measured for each of the four tests separately, as proportion of target items selected out of 16 (for six 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 weren't 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, four children had only one of the

two tests administered at one of the two timepoints, i.e. they completed only 3 of the 4 total tests—we determined that nothing was systematically different about these children or their performance when compared to the rest of the group.).

We used JASP Version 0.13.1 for all of our analyses. First, we evaluated memory performance by conducting a repeated-measures ANOVA, with proportion of targets selected (i.e. accuracy) being our outcome variable. Within-subject factors were pair type and time point, and the between-subjects factor was age. We conducted three additional ANOVAs with the same within- and between-subject factors, but with proportion 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. In order to account for the effect of our covariate (KBIT-2) when plotting our data, we regressed KBIT-2 scores on all of 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 Figures 2 & 3).

Results

Preliminary Analyses

Four children were removed from our sample for performing below chance when tested immediately (described in “Methods”), but of the remaining children, average performance was well above the chance level of 25% (59% for 4-year-olds; 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 sex 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-hour 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 ($M = .216$, $SD = 0.479$) 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 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 $ps > .05$). Additionally, 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 were 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 of the following analyses.

Main Analyses

We conducted 2 (age) x 2 (pair type) x 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-hour 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 “Methods, Procedure”). We analyzed target selection as the most basic index of memory retention for our stimuli, but were particularly interested in across-context lure selection, as 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.

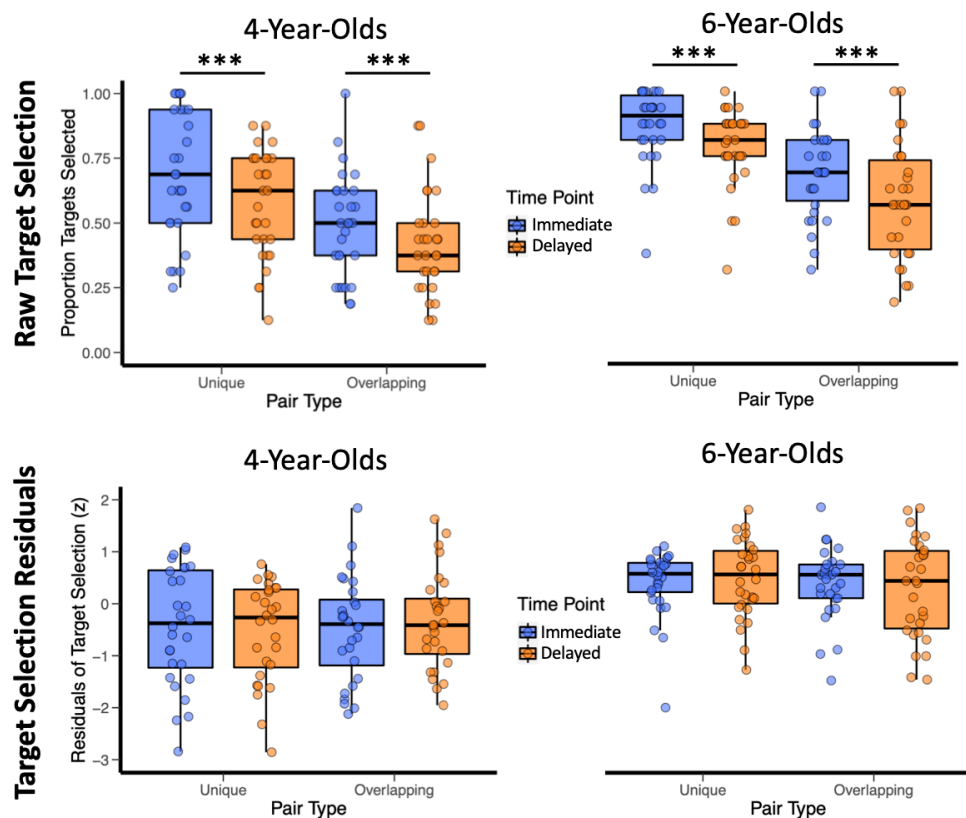


Figure 2. Target selection (top panel) and residuals of target selection after removing variance accounted for by KBIT-2 (bottom panel) for 4- and 6-year-olds, plotted separately by time point and pair type. Before KBIT-2 is included in the ANOVA, there is a significant effect of time point, but when KBIT-2 is included, the effect of time point is not significant. *** $p < .001$

.320)—a post-hoc test (using the Holm correction to adjust p) showed that 6-year-old children outperformed 4-year-old children ($t = 5.090$, $p_{\text{holm}} < .001$). 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 towards significance ($F_{1, 55} = 3.425$, $p = .070$, $\eta_p^2 = .059$). There were no

significant interactions (all p 's $> .14$). This suggests that older children performed better than younger children, and both groups of children better remembered the unique than the overlapping pairs.

It is worth noting that there was a significant effect of time point in an ANOVA in which 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 Figure 2). However, this effect did not hold with KBIT included as a covariate (see Figure 2), indicating that children's forgetting across a 24-hour 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 ($F_{1,55} = .378, p = .541, \eta_p^2 = .007$). Additionally, 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 overlapping pairs ($t = 1.13, p_{\text{holm}} = .261$).

Importantly, this pattern further interacted with time point, such that there was a significant three-way, age by pair type by time point interaction ($F_{1,55} = 5.064, p = .028, \eta_p^2 = .084$). Post-hoc tests showed 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 Figure 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 Figure 3). This result indicates that the delay period affected relational binding error rates differentially for younger vs.

older children, and that this effect depends on whether or not the relational pairs share an overlapping constituent across conditions.

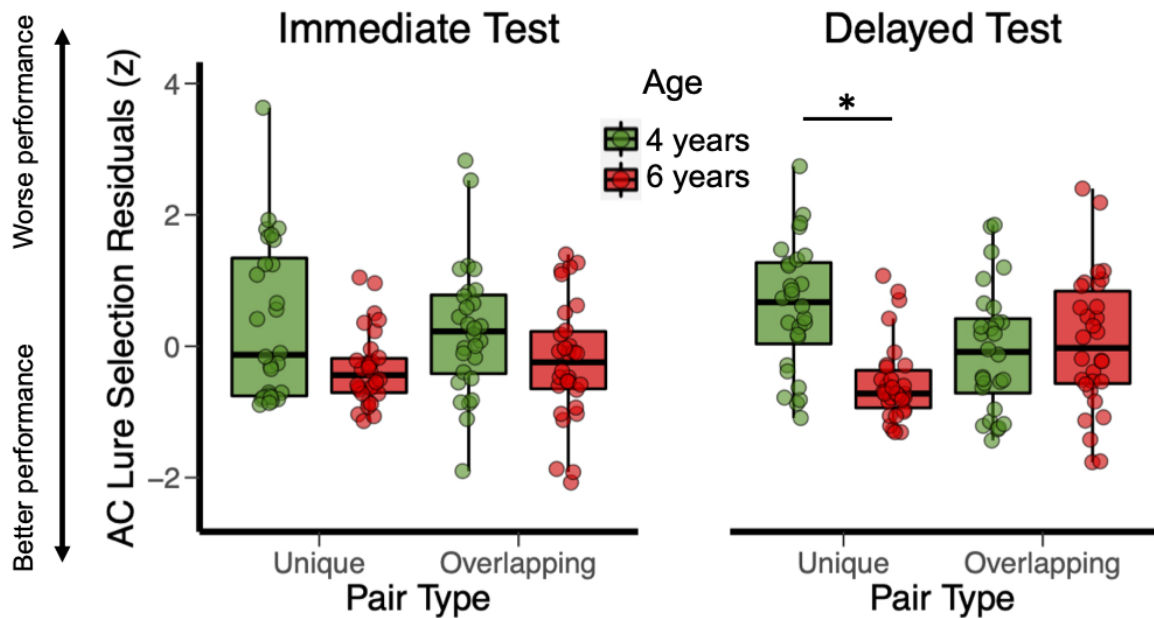


Figure 3. Across-context lure errors plotted and grouped by age, time point, and pair type after removing variance accounted for by KBIT-2 scores. The significant three-way age by time point by pair type interaction is highlighted with a significance bar. Note that selection of a lure is an error, so 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. $*p < .05$

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 ($F_{1,55} = .680, p = .413, \eta_p^2 = .012$) and time point ($F_{1,55} = .071, p = .790, \eta_p^2 = .001$) 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 two-way pair type by age interaction ($F_{1,55} = 4.211, p = .045, \eta_p^2 = .071$)—a post-hoc test showed that 6-year-olds' foil selection errors were greater for unique than 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 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. Four-year-old children 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-hr delay retained overlapping associated pairs to a greater extent than the 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-old children's brains during our delay period and then subsequently recalled. Six-year-olds, on the other hand, have more mature memory systems and are thus 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 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).

There are some important differences in the design of our study compared to 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 pre-specified 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? While 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 (Drummey & 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-hour 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 than 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.

References

- Abel, M., & Bäuml, K.-H. T. (2014). Sleep can reduce proactive interference. *Memory*, 22(4), 332–339. <https://doi.org/10.1080/09658211.2013.785570>
- 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(1), 76–80. <https://doi.org/10.1016/j.nlm.2007.08.010>
- Bailes, C., Caldwell, M., Wamsley, E., & Tucker, M. (2020). Does sleep protect memories against interference? A failure to replicate. *Sleep Medicine*, 64, S393–S393. <https://doi.org/10.1016/j.sleep.2019.11.1094>
- Bauer, P. J., & Larkina, M. (2019). Predictors of age-related and individual variability in autobiographical memory in childhood. *Memory*, 27(1), 63–78. <https://doi.org/10.1080/09658211.2017.1381267>
- Bauer, P. J., Larkina, M., & Deocampo, J. (2011). Early memory development. *The Wiley-Blackwell Handbook of Childhood Cognitive Development*, 2, 153–179.
- Canada, K. L., Botdorf, M., & Riggins, T. (2020). Longitudinal development of hippocampal subregions from early- to mid-childhood. *Hippocampus*, 1–14. <https://doi.org/10.1002/hipo.23218>
- 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(8), 3427–3433. <https://doi.org/10.1093/cercor/bhy211>
- Chanales, A. J. H., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among Spatial Memories Triggers Repulsion of Hippocampal Representations. *Current Biology*, 27(15), 2307-2317.e5. <https://doi.org/10.1016/j.cub.2017.06.057>

Darby, K. P., & Sloutsky, V. M. (2015). When delays improve memory: Stabilizing memory in children may require time. *Psychological Science, 26*(12), 1937–1946.

<https://doi.org/10.1177/0956797615607350>

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*(11), 3036–3045.

<https://doi.org/10.1093/cercor/bht160>

Drummey, A. B., & Newcombe, N. S. (2002). Developmental changes in source memory. *Developmental Science, 5*(4), 502–513. <https://doi.org/10.1111/1467-7687.00243>

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*(8), 1948–1953.

<https://doi.org/10.1523/JNEUROSCI.2975-03.2004>

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*(13), 1290–1294.

<https://doi.org/10.1016/j.cub.2006.05.024>

Favila, S. E., Chanales, A. J. H., & Kuhl, B. A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nature Communications, 7*. <https://doi.org/10.1038/ncomms11066>

Flin, R., Boon, J., Knox, A. and Bull, R. (1992). The effect of a five-month delay on children's and adults' eyewitness memory. *British Journal of Psychology, 83*, 323-336.

<https://doi.org/10.1111/j.2044-8295.1992.tb02444.x>

Gogtay, N., Hugent, T. F., Herman, D. H., Ordonez, A., Greenstein, D., Hayashi, K. M., ... Thompson, P. M. (2006). Dynamic Mapping of Normal Human Hippocampal Development. *Hippocampus, 16*, 664–672. <https://doi.org/10.1002/hipo>

- Herbert, J., & Hayne, H. (2000). Memory retrieval by 18-30-month-olds: age-related changes in representational flexibility. *Developmental Psychology*, 36(4), 473–484.
<https://doi.org/10.1037/0012-1649.36.4.473>
- Herrmann, D., & Guadagno, M. A. (1997). Memory performance and socio-economic status. *Applied Cognitive Psychology*, 11(2), 113–120. [https://doi.org/10.1002/\(SICI\)1099-0720\(199704\)11:2<113::AID-ACP424>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1099-0720(199704)11:2<113::AID-ACP424>3.0.CO;2-F)
- 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(2), 146–154. <https://doi.org/10.1037/0097-7403.14.2.146>
- 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(8), 676–686. <https://doi.org/10.1016/j.tics.2018.05.004>
- Krogsrud, S. K., Tamnes, C. K., Fjell, A. M., Amlien, I., Grydeland, H., Sultvedt, U., ... Walhovd, K. B. (2014). Development of hippocampal subfield volumes from 4 to 22 years. *Human Brain Mapping*, 35(11), 5646–5657. <https://doi.org/10.1002/hbm.22576>
- 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(2), 189–197. <https://doi.org/10.1037/0735-7036.121.2.189>
- Kurdziel, 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*, 110(43), 17267–17272. <https://doi.org/10.1073/pnas.1306418110>

Lloyd, M. E., Doydum, A. O., & Newcombe, N. S. (2009). Memory binding in early childhood: Evidence for a retrieval deficit. *Child Development, 80*(5), 1321–1328.

<https://doi.org/10.1111/j.1467-8624.2009.01353.x>

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(2), 263–273. <https://doi.org/10.1037/dev0000637>

Morgan, K., & Hayne, H. (2010). Age-related changes in visual recognition memory during infancy and early childhood. *Developmental Psychobiology, 53*(2), 157–165.

<https://doi.org/10.1002/dev.20503>

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*(5), 743–756. <https://doi.org/10.1111/desc.12162>

<https://doi.org/10.1111/desc.12162>

Newcombe, N.S., Benear, S.L., Ngo, C.T., & Olson, I.R. (in press). Memory in Infancy and Childhood. *Handbook on Human Memory*. Oxford University Press.

Ngo, C. T., Horner, A. J., Newcombe, N. S., & Olson, I. R. (2019). Development of holistic episodic recollection. *Psychological Science, 30*(12), 1696–1706.

<https://doi.org/10.1177/0956797619879441>

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*.

<https://doi.org/10.1037/xge0000583>

Ngo, C. T., Newcombe, N. S., & Olson, I. R. (2018). The ontogeny of relational memory and pattern separation. *Developmental Science, 21*(2), 1–11.

<https://doi.org/10.1111/desc.12556>

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). John Wiley & Sons Ltd. <https://doi.org/10.1002/9781118597705.ch13>
- 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(8), 1–9. <https://doi.org/10.1093/sleep/23.8.1d>
- 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), 1–7. <https://doi.org/10.1371/journal.pone.0033079>
- Piccolo, L. da R., Arteché, A. X., Fonseca, R. P., Grassi-Oliveira, R., & Salles, J. F. (2016). Influence of family socioeconomic status on IQ, language, memory and executive functions of Brazilian children. *Psicologia: Reflexão e Crítica*, 29.
- Pöhlchen, D., Pawlizki, A., Gais, S., & Schönauer, M. (2020). Evidence against a large effect of sleep in protecting verbal memories from interference. *Journal of Sleep Research*. <https://doi.org/10.1111/jsr.13042>
- 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(5), 844-853.
- Prebble, S. C., Addis, D. R., & Tippett, L. J. (2013). Autobiographical memory and sense of self. *Psychological Bulletin*, 139(4), 815–840. <https://doi.org/10.1037/a0030146>
- Richardson, J. T. E. (1998). Socio-economic status, social class and memory performance: A critical response to Herrmann and Guadagno (1997). *Applied Cognitive Psychology*, 12(6), 593–609. [https://doi.org/10.1002/\(SICI\)1099-0720\(199812\)12:6<593::AID-ACP579>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1099-0720(199812)12:6<593::AID-ACP579>3.0.CO;2-P)

- Riggins, T. (2014). Longitudinal investigation of source memory reveals different developmental trajectories for item memory and binding. *Developmental Psychology, 50*(2), 449–459. <https://doi.org/10.1037/a0033622>.Longitudinal
- 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. <https://doi.org/10.1016/j.neuroimage.2018.03.009>
- 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*(1), 1–20. <https://doi.org/10.1038/s41467-018-06213-1>
- Sheth, B. R., Varghese, R., & Truong, T. (2012). Sleep shelters verbal memory from different kinds of interference. *Sleep*. <https://doi.org/10.5665/sleep.1966>
- Sluzenski, J., Newcombe, N. S., & Kovacs, S. L. (2006). Binding, relational memory, and recall of naturalistic events: A developmental perspective. *Journal of Experimental Psychology: Learning Memory and Cognition, 32*(1), 89–100. <https://doi.org/10.1037/0278-7393.32.1.89>
- Spencer, R. M. C., Sunm, M., & Ivry, R. B. (2006). Sleep-dependent consolidation of contextual learning. *Current Biology, 16*(10), 1001–1005. <https://doi.org/10.1016/j.cub.2006.03.094>
- 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(32), 10087–10093.

<https://doi.org/10.1523/JNEUROSCI.0799-09.2009>

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

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. <https://doi.org/10.1111/cdev.12149>

Yim, H., Dennis, S. J., & Sloutsky, V. M. (2013). The development of episodic memory: Items, contexts, and relations. *Psychological Science*, 24(11), 2163–2172.
<https://doi.org/10.1177/0956797613487385>