The Component Structure of Memory during Development

Samantha S. Cohen,
Ingrid R. Olson,
Nora S. Newcombe
Department of Psychology and Neuroscience, Temple University, Philadelphia, PA, 19122

Abstract

Episodic memory, or memory for specific past experiences, can be measured in ways ranging from fully controlled laboratory paradigms to real world settings. However, it is unknown whether different measures of episodic memory are capturing the same construct. In other words, how do the different tasks that help to define episodic memory relate to one another, or what is the component structure of episodic memory as defined by the tasks that measure it? We examined this question in children between four and seven years (N=76), using a battery of assessments ranging from measurements of memory for mini-events presented in the lab, to measures of memory for a cartoon, and most naturalistically, recall of real-world events, i.e., autobiographical memory. Most measures of memory improved with age. After controlling for the effects of age and verbal IQ, a factor analysis revealed two distinct components of children's memory. The first component was comprised of our lab-based measures of memory and captured the majority of variance in the data. The second component included autobiographical memory and free recall of previously seen cartoons. This dichotomy may be driven by whether memory was assessed via recall or recognition, or by the presence or absence of narrative structure. Without changes to typically used experimental design, current lab-based tasks and more naturalistic assessments of memory may not be measuring the same underlying construct.

Keywords

Episodic memory; autobiographical memory; factor analysis; naturalistic; child development

Introduction

Memory is a multifaceted phenomenon that allows one to not only retell stories of past experiences, but also learn how to spell words, or to play an instrument (Kandel & Squire, 2001). Since memory is a broad construct, it has been divided into two main branches: one for memories that you are consciously aware of and one for memories that allow you to function in everyday life, but for which you cannot verbally demonstrate your knowledge, such as riding a bike (Squire, 1995). Among consciously reportable memories, there are

Address for correspondence: Samantha.cohen0008@temple.edu.

Conflict of Interest Statement: The authors declare no conflict of interests.

This study was not preregistered.

at least two forms: semantic and episodic memory (Tulving, 1972, 1983, 2002). Whereas semantic memory forms a general knowledge base, allowing one to remember facts and schemas, episodic memory allows one to recall specific experiences from the past (Tulving, 2002; Tulving & Lepage, 2000).

Here we investigated the structure of episodic memory during child development. Our use of the word "structure" refers to the various ways episodic memory is measured and how they relate to one another. Episodic memory has been captured by many tasks, in part because it is not a unitary construct (Nelson & Fivush, 2004). Although it constitutes memory for specific past events, this fundamental phenomenon of human consciousness can be measured in many different ways. For instance, there is a long tradition of examining memory for nonsense words or syllables, or lists of real words (Ebbinghaus, 1913; Kahana, 2012). Although these fundamental studies have helped us understand some aspects of episodic memory, some have argued that they do not capture the full range of rememberable events and experiences (Neisser, 1982; Virk et al., 2024). Some studies of episodic memory have thus utilized more naturalistic tasks that run the gamut from fully lab-based stimuli meant to simulate real world experiences to studies of naturally-occurring events (Virk et al., 2024).

There is an inherent tension, however, between precisely measured memories for items or events experienced in the laboratory context and memories for freely experienced events in the real world (Virk et al., 2024). In fact, there is a decades-long debate as to whether these two forms of episodic memory research are in fact measuring the same phenomena, and whether one form of research is more generalizable to the outside world than the other (Banaji & Crowder, 1989; Davies & Logie, 1993; Neisser, 1982). This tension is especially apparent in the study of the development of memory during childhood (Baker-Ward et al., 1993). Children may be less apt to learn arbitrary associations that are not made socially or personally relevant (Baker-Ward, Ornstein, and Holden, 1984). Thus, their overall memory capacity may be underestimated by purely lab-based measures of memory. Thus, having children learn a staged event or situation in the laboratory has advantages since the events learned at encoding are known, and can even be manipulated (Baker-Ward et al., 1993).

Some developmental memory researchers have posited that lab-based and real-world forms of memory exist on a continuum of experimental control (Baker-Ward et al., 1993; Virk et al., 2024). It is thus possible that different tests of the development of episodic memory all measure growth of the same construct and using lab-based tasks is therefore sufficient for measuring real world memory phenomena (Saragosa-Harris et al., 2021). In fact, several previous studies using multiple measures of episodic memory in children ranging in age from three to eight years have found that different tests of episodic memory are a part of the same latent construct (Canada et al., 2022; Cheke & Clayton, 2015). However, these previous studies cannot address whether lab-based and more naturalistic measures are measuring the same construct because they only employed batteries of fully lab-based tasks (Canada et al., 2022; Cheke & Clayton, 2015). It is therefore also possible that different assessments of episodic memory are assessing separable facets of a larger system. This may be especially true when directly comparing memories for items learned in the laboratory context and memories for experiences in the real world (Canada et al., 2022).

In this study, with children ranging from four to seven years of age, we utilized a battery of episodic memory tasks and examined the correlational and factor structure of these tasks. Although the particular tasks that we could have used is vast, we selected tasks that would be engaging to children while also spanning the gamut from more traditional laboratory-based tasks where the memoranda are still pictures and their associations, to real-life autobiographical memories. On one end of the laboratory-based to fully naturalistic spectrum, we used a mnemonic discrimination task designed to assess the process of pattern separation known to be essential for distinguishing similar episodic memories from one another (O'Reilly & Rudy, 2001; Yassa & Stark, 2011). In this task we assessed memories for events constructed from three simultaneously presented pictures (Horner & Burgess, 2013; Ngo et al., 2019, 2021). Each event consisted of an animal, an object, and a location that were accompanied by an auditory narrative. Children were subsequently tested on their ability to remember which particular items they had seen in comparison to similar lures and distracting foils. Through this task children were also assessed on relational binding, i.e., their ability to remember the associations between all pairs of items (Horner & Burgess, 2013; Ngo et al., 2019). Finally, this task also allowed us to assess the phenomenon of holistic recollection which measures whether individuals tend to remember or forget events as integrated units (Horner & Burgess, 2013; Ngo et al., 2019).

On the other end of the spectrum is memory for real-world events. To approximate this in the laboratory context we assessed children's memories for several novel cartoons (Furman et al., 2007; Yang et al., 2022). Memory was assessed one week after children watched the cartoons by asking children to freely recall the cartoons' narratives and asking the children specific questions about the stories. Finally, in the most naturalistic task, we asked children to freely recall memories from their real life past outside of the laboratory, or autobiographical memory (Nelson & Fivush, 2004; Virk et al., 2024).

Our goal was to compare our fully lab based assessments of the specific factors thought to be foundational for episodic memory function: mnemonic discrimination, relational binding, and holistic recollection, to the increasingly more naturalistic measures of probed questions and freely recalled narratives, and the most ecologically valid measure of autobiographical memory. We thus sought to uncover the structure of episodic memory during a period in life in which episodic memory is dynamically emerging (Newcombe & Nguyen, 2024).

Methods

Participants

A total of 104 four- to- seven-year-old children participated in the current study. Of these children, one was excluded due to language proficiency, 12 did not return for the second session when several of the memory measures were assessed, six subjects had over a one month gap between sessions, rendering it difficult to test their memory for the first session, and nine subjects are missing because their data was improperly recorded due to experimenter error. Overall, 76 children (5.81 ± 1.19 years, range = 4.1 – 7.9 years, 51 females, 25 males; see Supplementary Figure 1 for age and sex histogram) provided useable behavioral data and were included in the analyses. It is difficult to say if this sample size is sufficient for the analyses we conducted given the fact that our analyses are novel.

However, our initially recruited sample is similar to a previous study that addressed a similar conceptual issue (Cheke & Clayton, 2015). Furthermore, we do not see any differences between the data that was included in our final sample and the data from children for whom data collection was incomplete (see Supplementary Figure 2). The interval between the two sessions was between one and 20 days (on average 6.59 +/− 3.00 days; see Supplementary Figure 3 for distribution of delays broken down by age). Children were recruited from a major metropolitan area in the northeastern United States through recruitment events and online advertisements. All participants were enrolled in a larger, ongoing, longitudinal study on memory and brain development in accordance with and approved by the Institutional Review Board at Temple University. Prior to enrollment, children were screened via parental report to ensure they were proficient in English, had not been diagnosed with a psychiatric, neurological, or developmental disorder (including a history of head trauma, seizures, autism, or ADHD), had no known metal in their body, and exhibited normal or corrected-to-normal sensory processing (e.g. vision and hearing). Informed consent was obtained from parents, and written assent was obtained for seven-year-old children. Most of the sample identified as white, from middle- to high-income households.

Materials

All visual stimuli were presented on a 14” MacBook pro screen.

Short cartoons: Six cartoons previously used to examine narrative proficiency in children in the same age range (Demir et al., 2014) were selected. The cartoons were short, silent stories about a small mouse and his friends (30–73 seconds; *Die Sendung mit der Maus*). These cartoons were selected partly since we expected our American subjects to be unfamiliar with them. Each cartoon had at least one goal, an initiating event (the problem), multiple episodes (attempts to achieve the goal), and an outcome or resolution (Demir et al., 2014).

Pattern separation (mnemonic discrimination) and pattern completion (relational binding and holistic recollection): The stimuli came from 216 child-friendly cartoon images used in a previous study (Ngo et al., 2021). Of these stimuli one-third were scenes (e.g., a lake), one-third were animals (e.g., a sheep), and one-third were objects (e.g., a mug). Each category of stimulus included 18 exemplar pairs (e.g., two similar images of lakes) to be used for the test of mnemonic discrimination. Children encoded 18 unique scene-animal-object triads (e.g., lake-sheep-mug). The assignment of an item to a triad was pre-determined based on the criterion that items with a potential pre-experimental semantic association (e.g., library and notebook) would not be assigned to the same triad. Following the methods of prior research in this age range (Ngo et al., 2019), a three-sentence narrative was pre-recorded for each triad, with each sentence highlighting one association within the triad. For example, the lake-sheep-mug triad was accompanied by the following narrative: “Sheep rested by the lake. He was thirsty so he got out his mug to drink from. The mug looked very nice in the reflection on the lake.” The order of the associations in each narrative (scene-animal, animal-object, object-scene) was consistent across all narratives. All narratives were audio-recorded with a child-friendly female voice, and with a fixed duration of 15 seconds. These narratives were used to maximize task engagement, and to increase the likelihood that attention was drawn to each item within a triad.

Procedure

Participants came to the lab for two sessions on different days. During the first visit demographic information including date of birth, sex, race, ethnicity, and socioeconomic information were reported by the child's parent using Redcap survey software (Harris et al., 2009, 2019).

Autobiographical memory: The procedure for the autobiographical interviews was modeled after those established in prior literature (Cleveland & Reese, 2005; Jack et al., 2009). Parents provided four positive or neutral non-routine events that they shared with their children between two days and six months ago (e.g., a trip to the zoo). Subsequently, two of the events were discussed between the parent and their child, and the other two were discussed between the child and an experimenter. The order of who the child spoke with first (parent or experimenter) was pseudo-randomized across participants.

For the two events recalled with the experimenter, experimenters began the conversation by saying: "Your (parent) told me that you (event) recently. Tell me about that." The experimenters were trained to make minimal prompts, limiting their responses to phrases such as: "mmhmm," "wow!", "tell me more", and repeating child's responses verbatim. For each event, the conversation continued until the child indicated that they were finished or responded "no" when asked if there was "anything else?". For the two events recalled with the parent, parents were instructed to discuss the events sequentially and as naturalistically as possible. The parent and child discussed the events without the experimenter present. All four event conversations were recorded via Audacity (Audacity, 2022).

Querying memory for short cartoons

Encoding. Encoding occurred during the child's first lab visit and was introduced as a "storytelling game." Before watching each cartoon, children saw a still frame from the cartoon and the researcher introduced the names of the characters and objects relevant to the story. The child watched each cartoon twice. During the second viewing the child was prompted to narrate the story in the cartoon. All participants saw the cartoons in the same order.

Recall. During the child's second visit they saw one frame from each cartoon and were prompted to tell the experimenter "as much as they can remember" about each story. While viewing the frame participants freely recalled the story to the best of their ability. Once the researcher asked, "anything else?" and the child responded "no", the participant said, "the end" or "I don't remember", researchers asked three questions about each story. Two of the questions prompted children to identify the specific problem and solution in the story (e.g. "Why doesn't the hose work?" or "What does Ellie do to fix the hose?"). The final question for all stories was "What is a good title for this story?". In total, across all stories, 17 questions were asked. All phases of the task were recorded via Audacity (Audacity, 2022).

Assessment of mnemonic discrimination, relational binding, and holistic recollection: All children were tested individually in a room with an experimenter present. The task was divided into two blocks (example block displayed in Figure 1). Each block consisted of an encoding phase, the pattern separation test (assessing mnemonic discrimination), and the pattern completion test (assessing relational binding and holistic recollection), administered in a fixed order. The two blocks were procedurally identical but with completely nonoverlapping stimuli (nine event triads/block). Children experienced one of three different encoding/retrieval conditions for this task. Seven children encoded and retrieved Block 1 during their first session, and Block 2 during their second session. Twenty-four children encoded and retrieved both blocks during their first lab visit. In other words, a total of 31 children encoded and retrieved the stimuli within the same session. The remaining 45 children encoded both blocks during their first lab visit and retrieved them during their second visit (see Supplementary Figure 4 for the range of possible delays for this task). Even after controlling for age, whether children did the mnemonic discrimination and relational binding task on the same day or after a delay had a significant effect on their memory accuracy (both p's < 1×10^−10; See Supplementary Table 1). To account for the effect of delayed recall, the delay between encoding in retrieval (in days) was regressed out of the mnemonic discrimination and relational binding data.

Encoding. Children were invited to play a "story game" about animals who went to places with objects. The encoding phase consisted of 13 events. The items and events in nine were subsequently tested on (Figure 1A). The remaining four stories served as examples and subsequent test practice material: one was shown at the beginning of the encoding phase to acquaint children with the procedure, and three were shown at the end to be used as practice material for the pattern separation test. In each event, a pre-designated scene, animal, and object appeared simultaneously on the top, bottom left, and bottom right of the screen, respectively. Each event was presented for 15 seconds with a 0.5-second inter-trial interval. A corresponding pre-recorded narrative accompanied each encoded event. The same block order and event order was administered to all children. The encoding phase for each block, including the example events, took approximately four minutes.

Pattern Separation Test. Children were invited to play the "which one" game wherein the experimenter wanted to know how well they had remembered the stories (Figure 1B). We tested children's memory discrimination for each item seen in each story with a self-paced four-alternative-forced choice (4AFC) task. In each block, this test phase was divided into three parts: scene, animal, and object (27 trials per block, nine items per item category). The order of the blocks and within-block trials were identical for all children.

In every test trial, children were shown four image options from a given item category (e.g., four scene images) and were asked to choose the same image that they had seen previously by pointing to it. The four options included a target: the same image that appeared at encoding (e.g., the same lake), a lure: a similar exemplar of the target (e.g., a different lake), and two foils: similar exemplars of an unstudied place (e.g., two lighthouses). The experimenter recorded the child's selection with a mouse click. The screen positions of the targets and lures were approximately equally distributed within each block and across the entire test phase. The pattern separation test phase for each block took on average 5.0 +/− 2.0 minutes.

Note that the test phase for each block was preceded by a brief practice phase that introduced all three kinds of pattern separation test trials: scene, animal, object. In this practice phase, children were tested on items from the example events that were presented during the end of the encoding phase.

Pattern Completion Test. The pattern completion task for each block immediately followed the pattern separation task and was introduced as the "Go-together" game (Figure 1C). In this game, children were instructed to choose the item that was previously presented in the same event as the cue. To estimate holistic recollection (defined in analysis subsection), we tested children on all possible cue-item pairs for each three-item event (e.g., retrieving the object that belonged to the same event as the animal cue). This resulted in six different retrieval types per event (see Figure 1C for illustration). The six different retrieval pairs were separated into blocks with nine trials per block for a total of 54 trials per block. To maintain attention, there was a self-paced break between each block. All subjects were tested in the same pseudo-randomized block order so that bidirectional test types did not occur in succession (e.g., A_B following B_A, where B and A come from the same event).

On each test trial, a cue and four options were presented simultaneously (see Figure 1C). The four options consisted of one target: the correct item associated with the same event as the cue, and three lures: same-category elements from different events in the same block. The position of the target was counterbalanced across the test phase. Each item served as a lure an equal number of times across all trials within each block and across the entire test phase. Given that holistic recollection is based on the accuracy of test trials that share the same cue (e.g., A_B and A_C) or the same to-be-retrieved item (e.g., B_A and C_A), we counterbalanced how often the lures on these trials overlapped in their membership. Across all nine events in each block, any given two test trials that had overlapping cue items (e.g., A_B and A_C, where the A's are the same scene), or tested items (e.g., B_A and C_A, again, where A is identical across pairs), or were bidirectional (e.g., A_B and B_A, where A and B are identical) only shared one lure with one another. For example, for the A_B test trial for event 1, the lures included the B elements from events 2, 3, and 4, whereas for the A_C trial for event 1, the lures included the elements from events 3, 5, and 7. Thus, only one B and one C lure both belonged to the same event (event 3). The pattern completion test phase for each block took on average 8.0 +/− 2.3 minutes.

As in the pattern separation test phase, a brief practice session preceded the pattern completion phase for each block. In the practice phase, children were shown an example event and immediately tested with three example trials where each item served as the cue or as a to-be-retrieved item once (e.g., cue: scene — retrieve: animal; cue: object — retrieve: scene; and cue: animal — retrieve: object).

Verbal intelligence: To assess age-adjusted verbal intelligence (IQ) participants took the verbal knowledge and riddles subtests from the Kaufman Brief Intelligence Test (second edition; KBIT-2; (Kaufman & Kaufman, 2004). The verbal knowledge subtest is comprised of 60 items. For each item, the child is presented with six images and is instructed to point to one image based on the experimenter's prompt (e.g. "Point to the one that goes with thunder" which is correctly answered by pointing to the image of lightning). The riddles subtest has 48 items. The first eight items require the participant to point to correctly identify an object, and the later items are typically questions (e.g. "What is very far away, can only be seen at night, and twinkles in the sky?" is correctly answered by either star or planet). For each subtest, the first item queried is determined by the participant's age. Participants are given a 1 for correct responses and a 0 for incorrect or omitted responses. For both subtests, if a participant cannot correctly answer four questions in a row, they are asked questions from a starting age one level below their actual age. Each subtest, which increases in difficulty in each section, was terminated after a child provided four consecutive incorrect responses.

Analysis

Autobiographical memory: Audio recordings of the conversations were transcribed using Otter.ai (https://otter.ai/) and checked manually for accuracy. The transcripts were then divided into the four discussed events. The dialogue within each event was then divided into utterances, defined as conversational turns containing a verb or implied verb, such that there was one verb per clause. The child's utterances were then coded for memory elaborations using an adapted version of a reliable coding scheme established in the literature (Reese et al., 1993). Utterances were coded exhaustively for structure in accordance with previously used coding schemes that measure memory elaborations, repetitions, confirmations, questions, meta-memory comments, and associated event talk (see (Haden, 1998; Reese et al., 1993; Reese & Brown, 2000), but ultimately only the child's memory elaborations, which provide new information relevant to the discussed event, were used in the present analyses (Cleveland & Reese, 2005). One coder coded all autobiographical memory transcripts. This coder was trained by becoming reliable with an unpublished sample of 34 autobiographical memory transcripts from four- to seven-year-olds, with a Cohen's Kappa of 0.836 (0.823 for the parent-child events, and 0.855 for the experimenter-child events), indicating substantial agreement between raters (J. Cohen, 1960; Landis & Koch, 1977). The number of memory elaborations elicited by the child were averaged across all four events they recalled to yield a measure of autobiographical memory.

Coding for memory of short cartoons: Audio recordings of the narratives were transcribed using Otter.ai (https://otter.ai/) and were checked manually for accuracy. Following previous research on the analysis of narrative ability (Demir et al., 2014), each participant's free recall was separated into clauses corresponding to a single situation or event. A clause is defined as a subject (noun phrase or its equivalent) and its predicate (verb phrase and other accompanying elements such as object or complement).

The answer key to determine each child's memory for the cartoons was derived from the union of all unique clauses written by ten adults who were otherwise unfamiliar with the task. The number of possible correct clauses in each cartoon ranged from 18–28 (23.5 +/– 3.8). Each child received one point for each of their clauses that was present in the answer key and received no points for repeated clauses or falsely remembered clauses. These points were summed within each story and then averaged across all six stories to yield a cartoon free recall score. The transcripts from 19 children were scored by two different research assistants yielding a between-rater ICC of 0.89.

All recall questions (excluding "What is a good title for this story?") were scored on a 0–2 scale. A 0 was given if the answer was false, irrelevant to the story, paraphrasing the question, or the child indicated that they forgot. A 1 was given if the child's answer was accurate to the story (provides details relating to the story) but incomplete (missing a detail to fully answer the question). A 2 was given if the child's answer was accurate and complete. The questions from 34 children (44.7% of the sample) were scored by two different research assistants to establish an inter-rater ICC of 0.90 (0.88–0.92). The scores of all 11 scored questions were averaged for each participant to obtain a measure of question-prompted memory for the cartoons.

Holistic recollection calculation: Following previous work (S. S. Cohen, Ngo, et al., 2025; Horner & Burgess, 2013; Ngo et al., 2019, 2021), we sought to determine the degree to which people remembered (or forgot) all of the pair-wise associations within the same event while controlling for their overall associative memory accuracy in the pattern completion task. We therefore calculated two values: (1) the proportion of joint retrieval the data (described below) and (2) the independent model (described below). The difference between the proportion of joint retrieval within the data and the independent model yields a measure of holistic recollection.

The proportion of joint retrieval. The proportion of joint retrieval in the data measures, within one participant's data, whether remembering (or forgetting) an item (e.g. a scene) when cued by another item from the same event (e.g. an object) was associated with remembering (or forgetting) the same item (the scene) when cued by the third item from the event (e.g. the animal). Thus, we calculated the frequency with which each participant correctly remembered or forgot each possible pair of cue-item pairs (either A_BA_C, a common cue, or B_AC_A, a common to-be-retrieved item) from the same event. For each event, this was done separately for all six possible pairs of cue-item pairs, and then these values were averaged across all events and all pairs.

The independent model. To calculate a measure of the overall likelihood of remembering (or forgetting) any given cue-item pair, regardless of whether that pair belongs to the same event, we calculated the independent model based on previous work (Horner & Burgess, 2013). For this calculation, for instance, the probability of remembering any scene-object pair is deemed independent of remembering any scene-animal pair. For each participant, we therefore found the average accuracy for one cue-item pair across all events and multiplied this value by the average accuracy for the other cue-item pair across all six possible situations with either a common cue or a common to-be-retrieved item, and then averaged these values across all pairs.

The measure of holistic recollection for each subject is the proportion of joint retrieval in their data subtracted by the value calculated for their independent model, thus controlling for overall pair-wise retrieval accuracy in a world in which someone is no more likely to jointly retrieve pairs from the same event than can be explained by accuracy alone. For example, for elements A, B, and C, for an event with three elements, someone might be able to use A to remember B and C, or use C to remember A and B, but be unable to use B to cue either B or C. In this instance, where a cue allows someone to either completely recall the other items associated with it, or a cue does not allow them to remember any of the items associated with it, the person has a positive level of holistic recollection. In another situation, someone else might be able to recall C after being prompted by A, but not recall B, and be able to recall C after being prompted by B, but not recall A. They may also be able to use C to recall both A and B. Although this other person has the same overall accuracy level as the first individual, they have no holistic recollection because the items are sometimes able to cue one association but not the other.

Verbal intelligence: The standard score of verbal intelligence (IQ) is based on the summed raw score from each subtest adjusted by the child's age in months. The raw score for each subtest is derived by subtracting the total number of errors from the ceiling item. The assessments were scored and recorded by two separate research assistants and discrepancies were examined by a third research assistant.

Factor analysis: In sum, the variables subject to the initial investigation of the structure of our data were as follows: (1) three measures of mnemonic discrimination (the rate of target, lure, and foil selection in the pattern separation task), (2) relational binding (the overall accuracy on the pattern completion task), (3) holistic recollection (calculated from the pattern completion task), (4) free recall accuracy (the average number of correct clauses recalled across the six cartoons), (5) probed question accuracy (the average score across all memory questions about the cartoons), (6) the number of memory elaborations for the autobiographical memories averaged over all events discussed. See Supplementary Table 2 for a summary of all memory measures, Supplementary Table 3 for descriptive statistics for all memory measures, Supplementary Figure 5 for a depiction of the bivariate relationships between individual data points and histograms of all variables, and Supplementary Figure 6 for a depiction of the bivariate relationships between individual data points used for the factor analysis after regressing out age and IQ.

We note here that our results in terms of factor loading look very similar if we include a more exhaustive list of variables where we additionally include: (1) the rate of target, lure, and foil selection for the three item categories (object, animal, and scene), (2) the relational binding accuracy on the pattern completion task for the six possible pair types (see Figure 1C), (3) the number of memory elaborations for the autobiographical memories discussed with the parent and the experimenter separately. We first examined potential correlations between our memory measures and age and IQ, and subsequently controlled for these variables in all subsequent analyses.

The assumptions of exploratory factor analysis (EFA) include a linear relationship between variables as well as a lack of outliers in the data. We checked for linear relationships via visual inspection of the bivariate relationships between data points (see Supplementary Figure 6 for a depiction of these relationships). Another assumption is that the data are free of outliers. We therefore re-ran our EFA after detecting and removing outliers based on the interquartile range method of outlier detection. Our factor analysis results are quite similar with (see Table 1) or without (see Supplementary Table 4) these data points. We have decided to include them in the manuscript since the results are stable.

Next, all remaining variables were subject to an EFA. In preparation for the EFA all variables were normed by subtracting their mean and dividing by their standard deviation. Age and IQ were regressed out of all variables. Additionally, the variable delay period between encoding and retrieval was regressed out of the memory measures derived from the pattern separation and pattern completion task (for mnemonic discrimination and relational binding). We performed a Kaiser-Meyer-Olkin factor adequacy test and Bartlett test of homogeneity of variances prior to performing our EFA to ensure our data were appropriate for factor analysis.

To determine the number of factors to retain, since Kaiser criterion technique for determining the number of factors to be retained can be problematic (Zwick & Velicer, 1986), we performed a parallel analysis (Horn, 1965) in which we ran a simulation with 100 replications to determine what the eigenvalues would be if there were the same number of cases and variables, but the data were random. If the eigenvalues from our real data were lower than expected due to chance (i.e., those produced from the parallel analysis), then that factor would not be interpreted as capturing any latent traits present in the data. We also examined a plot of Velicer's Maximum Average Parcel to determine the relative amounts of systematic and unsystematic variance remaining in a correlation matrix after extractions of increasing numbers of components.

After obtaining the number of factors to retain, we performed principal components analysis (PCA) to determine how each variable loaded onto each factor. Our PCA was based on a custom R script using the *oblimin* rotation. Code was written in R and Python and is available on Github (https://github.com/samsydco/Factor_Analysis). The materials necessary to reproduce the findings reported in this manuscript are available at https://osf.io/s7ywr/ (S. S. Cohen, Olson, et al., 2025). The study's design and its analysis were not preregistered.

Results

We sought to determine the structure of episodic memory in children by analyzing how several measurable components of episodic memory correlate with one another and load onto latent factors. Our assessments of memory ranged from lab-based assessments of mnemonic discrimination (a measure of pattern separation), relational binding, and holistic recollection (a measure of pattern completion), to completely naturalistic assessments of memories experienced in the real world (autobiographical memory; See Supplementary Table 2 for a summary of all memory measures used).

Age-related memory improvements

As episodic memory ability is known to improve between 4–7 years of age (Newcombe et al., 2022; Newcombe & Nguyen, 2024), we first checked to see if this expected pattern was evident in our data. We indeed found age-related memory improvements in most of the major components of episodic memory that we examined (Figure 2 and Supplementary

Figure 5 for depiction of the relationship between individual data points). For our lab-based assessment of mnemonic discrimination, selecting the previously seen target item significantly increased with age ($r = 0.46$, $p = 3\times10^{-5}$) and likewise the rate at which children selected the similar lure and distracting foil items significantly decreased with age (lure: $r = -0.35$, $p = 0.002$, foil: $r = -0.42$, $p = 1\times10^{-4}$). Similarly, our lab-based assessment of relational binding, or remembering the associations between items seen together, significantly increased with age ($r = 0.45$, $p = 5\times10^{-5}$). However, in line with previous research (Ngo et al., 2019), we found that holistic recollection is an age invariant phenomenon within the age range of four to seven years ($p = 0.6$). Similarly, our more naturalistic memory measures also improved with age. This included freely recalled memory for short cartoons ("free recall accuracy", $r = 0.68$, $p = 2\times10^{-11}$), accuracy for short questions about those cartoons ("probed question accuracy", $r = 0.50$, $p = 3\times10^{-6}$), and the number of autobiographical memory elaborations made ($r = 0.31$, $p = 0.007$). The delay between encoding and retrieval had a significant effect on mnemonic discrimination and relational binding performance and was therefore regressed out of all analyses so as not to spuriously generate correlations that could influence the factor structure of the data. (see Supplementary Table 1 for a summary of these results). No measure of memory correlated with IQ except for probed question accuracy ($r = 0.26$, $p = 0.03$).

Cross-correlation between variables.

Next, we assessed the bivariate correlations between our measures of episodic memory after controlling for the potential effects of age and IQ. Since all measures of mnemonic discrimination (for target, lure, and foil) were highly correlated with one another even after controlling for age and IQ (correlation magnitudes ranging from 0.70 – 0.93) they are measuring the same underlying construct. We therefore only examine mnemonic discrimination for the target going forwards, as this measure positively correlates with the other memory measures.

As it was prior to controlling for age, relational binding between previously seen items remained highly correlated with mnemonic discrimination for the previously seen target (Figure 3). Probed question accuracy is now no longer significantly correlated with any other episodic memory measure. Interestingly, it does not even significantly correlate with free recall accuracy for the same cartoons. Finally, free recall accuracy and autobiographical memory are correlated, perhaps due to the similar nature of these tasks which involve describing experienced events or stimuli. As was seen prior to controlling for age and IQ, holistic recollection is not significantly correlated with any other variable (correlation range from 0.1 – –0.07, Figure 3).

Exploratory factor analysis

The variables included in the factor analysis were the measures of relational binding and mnemonic discrimination (as assessed by target selection rate), the measures of memory for short cartoons (free recall accuracy and probed question accuracy), and autobiographical memory. Holistic recollection was not included as it is not a measure of memory accuracy and is therefore unrelated to all other measures of episodic memory (Figure 2). Age and IQ were regressed out of all variables prior to performing the factor analysis. Since the effect of delay was only significant for relational binding and mnemonic discrimination, the effect of delay (in days) was only regressed out of these two variables.

An exploratory factor analysis (EFA) using the principal factor solution factoring method was employed to examine the factorial structure of the memory measures. The number of components to be extracted was determined by parallel analysis (Horn, 1965) as this method is thought to be a good detector of components across differing levels of component complexity (Zwick & Velicer, 1986). Parallel analysis compares the observed eigenvalues extracted from our data's correlation matrix to those obtained from uncorrelated normal variables. We adopted the rule that a factor is significant if its eigenvalue is larger than the mean of those obtained from the random uncorrelated data. This approach revealed a two-factor solution. explaining 62.9% of the total variance (See Supplementary Figure 7 for the parallel analysis scree plot).

The component loadings from a principal component analysis (PCA) with two components can be seen in Table 1 (see Supplementary Table 4 for a replication of these results after potential outliers were removed). The loadings reveal that the two factors roughly align with a division based on lab-based measures of memory and more naturalistic memory measures. However, probed question accuracy regarding previously seen cartoons, a seemingly more naturalistic measure of memory, loads strongly onto both the lab-based memory component (Factor 1) and the naturalistic memory component (Factor 2). Factor 1 explains 37.0% of the variance in the data and Factor 2 explains 25.9% of the variance in the data. Furthermore, the two factors are very weakly correlated with one another ($r = 0.07$). Thus, although lab-based memory measures explain more variance in the data, naturalistic measures of memory explain an unrelated part of episodic memory.

Discussion

This study examined whether different assessments of episodic memory are measuring the same underlying construct in children. Four to seven year old children's memory was assessed with a range of more traditional lab-based tasks assessing mnemonic discrimination, relational binding and holistic recollection (Horner & Burgess, 2013; Ngo et al., 2019, 2021), naturalistic tasks assessing memory for short cartoons (Furman et al., 2007; Yang et al., 2022), and an assessment of autobiographical memory (Nelson & Fivush, 2004; Virk et al., 2024). We found that all assessments of episodic memory improved with age within the range of four and seven years. Holistic recollection is a notable exception to this trend, but note that, unlike the other memory assessments, holistic recollection is not an index of memory accuracy but rather of memory cohesion (Horner & Burgess, 2013). After regressing out the effects of age and IQ, the memory measures clustered into two main components: one captured the lab-based measures of memory and the other represented autobiographical memory and the free recall of narratives.

The improvement in memory we saw between four and seven years aligns with previous literature. This is a dynamic period of development when children acquire the ability to encode episodic memories (Newcombe et al., 2022; Newcombe & Nguyen, 2024). After regressing out the effect of age, verbal IQ, and the variable delay between encoding and retrieval, some of the measures of memory were still highly correlated. Notably, the lab-based measures that derived from the same encoding experience of triads of elements – mnemonic discrimination and relational binding – correlated highly with each other. However, the questions asked about the cartoons were not significantly correlated with the ability to accurately freely recall the same cartoons. Memory for the same memoranda was measured in a different way in these two tasks. However, free recall accuracy correlated with our measure of autobiographical memory, our most naturalistic measure of episodic memory.

The lab-based measures load onto a single factor that explains most of the variance in the data. Autobiographical memory and free recall for cartoons load onto a second, unrelated factor. Thus, although the lab-based measures are not indicative of performance on our most naturalistic measure of episodic memory, namely the real-world assessment of autobiographical memory, this form of memory assessment is important as it captures most of the variance in memory behavior in our data. The free recall of cartoons loads onto both factors, but best on Factor 2, and thus may be an important bridge between purely lab-based and real-world assessments of memory.

One possible reason for the two-factor dichotomy is a difference in the way memory was assessed. In the first factor the tasks involved recognition memory; previously seen items were presented to the children. The tasks comprising the second factor required the children to freely recall past experiences, or to recall in response to questions. Recall and recognition may be distinct and separable memory processes that echo the distinction between recollection and familiarity (Yonelinas, 1994, 2002). Recall is similar to recollection in that it is an all-or-none retrieval process that either allows an individual to call a specific experience to mind or not. Although recognition tasks can rely on both recollection and familiarity, these two processes have been found to be dissociable in these tasks, and familiarity can be independently used to detect previously seen items in these paradigms (Yonelinas, 1994, 2002). There is evidence that recall and recognition tap neuroanatomically separable brain networks, with recall relying more heavily on the hippocampus, whereas recognition may rely on cortical regions in the surrounding temporal lobes (Yonelinas, 2002; Yonelinas et al., 2010). Supporting these theories, a study in four- to seven- year old children found that the free recall of cartoons had no relation with recognition of details from those cartoons after accounting for the effect of age (Benear et al., 2025).

An additional potential driver of the two-factor solution is the variation in encoding conditions across tasks. The memory measures that comprise the first factor are both derived from experiencing the same content. This factor may therefore be partly due to the fact that the encoding experience, and accompanying attention during that experience, was identical across measures. However, encoding of the cartoon stories and of the autobiographical events occurred at different times and places, and hence this factor does not depend on a common encoding experience.

Another difference between what the factors measure may involve personal relevance, which is very different for the triads of images presented in the lab and a real-world event attended with friends and family. Autobiographical and non-autobiographical episodic memories have been shown to depend on distinct, non-overlapping brain regions (Gilboa, 2004). Again, however, the cartoon stories are in between. They lack personal relevance but may be personally engaging. Furthermore, real-world events and watching short videos both inherently contain some form of narrative structure, whereas static images displayed on a screen lack this dynamism.

Future research should seek to distinguish whether our two-factor solution is due to use of recognition versus recall, a common encoding experience, or the presence of narrative structure. Memory for narratives could be assessed via recognition memory (Benear et al., 2023). Likewise, relational binding and mnemonic discrimination could be assessed via cued recall. By systematically varying the way stimuli are encoded and retrieved we will gain a better understanding of whether the mode of encoding, retrieval, or potentially both help to structure the components of episodic memory.

Autobiographical memory can be assessed in several different ways, and it is thus possible that a different measure of autobiographical memory would yield a different result. We measured autobiographical memory as the number of unique memory elaborations that children elicited across four different events. Although this measurement technique has been utilized in previous literature on children's autobiographical memory (Cleveland & Reese, 2005; Jack et al., 2009), this previous literature has also often accounted for how parents scaffold their children's recalls (Cleveland & Reese, 2005; Jack et al., 2009; Reese et al., 1993; Reese & Brown, 2000). Parental support during autobiographical recall has been shown to account for differences in how children recall their pasts (Jack et al., 2009). This difference may extend beyond autobiographical memories into other skills such as literacy (Reese, 1995). Furthermore, some measures of autobiographical memory in adults split autobiographical memory recalls into categories based on whether the details recalled constitute actual episodic memories or semantic details related to an individual's knowledge base more broadly (Levine et al., 2002). It would be interesting for future research to account for measures of parental support of recall as well as different ways to categorize memory elaborations. Additionally, our measure of autobiographical memory had no ground truth validation because we weren't present at the actual event. It is therefore difficult to know how our measure of autobiographical memory maps onto memory accuracy. Of course, our more naturalistic measure of freely recalling previously seen cartoons helps to account for this gap since in this case we can precisely know and control the experience that children were later recalling.

One potential limitation of our work is the relatively small number of tasks that we used given the numerous ways in which one could measure episodic memory. Episodic memory research ranges from using highly abstracted stimuli that bear little resemblance to what is experienced in real life (Ebbinghaus, 1913), to the exploration of autobiographical memory discussed above. Given that we were assessing memory from children with limited attention spans with which to encode memoranda, we were unable to give them as wide a variety of memory tasks as would be ideal to fully flesh out the component structure of episodic memory. Additionally, each task had a variable number of trials, videos, or events upon which memory performance was based. This variation between the demands of each task may have added additional noise to our measures. A final limitation is our relatively small sample size that may yield spurious results driven by random noise. One encouraging finding is that our factor results do not change when we eliminate potential outliers in our data (see Supplementary Table 4). Future research should explore these potential complexities and limitations by administering additional tasks to a larger sample of both children as well as other ages. In addition, it is possible that the component structure of memory changes with age.

In sum, we found that episodic memory can be divided into a two-factor structure, supporting the notion that lab-based and real-world measures of episodic memory may not be measuring the same underlying construct (Baker-Ward et al., 1993; Neisser, 1982; Virk et al., 2024). This conclusion is in contrast to previous analyses of the structure of episodic memory (Canada et al., 2022; Cheke & Clayton, 2015; Saragosa-Harris et al., 2021). Most of the variance in measurable episodic memory comes from lab-based measures of memory, indicating that these kinds of tasks constitute an important facet of memory behavior that is not captured in more naturalistic paradigms (Banaji & Crowder, 1989). However, naturalistic tasks measure a distinct, but important, part of episodic memory. In fact, these tasks may be most indicative of how we recall our lived experiences (Baker-Ward et al., 1993; Neisser, 1982; Virk et al., 2024).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to acknowledge Tom Olino for his consultation regarding data analysis and Josh Litwin and Kate Hill for their help in collecting the data. This work was supported by a National Institute of Health grant to N.S. Newcombe and I.R. Olson [R01HD099165]. 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 Availability Statement:

The data that support the findings of this study are openly available on github at https://github.com/samsydco/Factor_Analysis.

References

Baker-Ward L, Ornstein PA, & Gordon BN (1993). Chapter 1 A Tale of Two Settings: Young Children'S Memory Performance in The Laboratory and The Field. In Advances in Psychology (Vol. 100, pp. 13–41). Elsevier. 10.1016/S0166-4115(08)61089-4

Banaji MR, & Crowder RG (1989). The Bankruptcy of Everyday Memory. American Psychologist.

Benear SL, Onwukanjo OJ, Olson IR, & Newcombe NS (2025). Children's Memory for Events: The Challenge of Free Recall. Journal of Cognitive Neuroscience, 1–12.

Benear SL, Popal HS, Zheng Y, Tanrıverdi B, Murty VP, Perlman SB, Olson IR, & Newcombe NS (2023). Setting boundaries: Development of neural and behavioral event cognition in early childhood. Developmental Science, 26(6). 10.1111/desc.13409

Canada KL, Hancock GR, & Riggins T (2022). Developmental changes in episodic memory across early-to mid-childhood: Insights from a latent longitudinal approach. Memory, 30(3), 248–261. [PubMed: 34825853]

Cheke LG, & Clayton NS (2015). The six blind men and the elephant: Are episodic memory tasks tests of different things or different tests of the same thing? Journal of Experimental Child Psychology, 137, 164–171. 10.1016/j.jecp.2015.03.006 [PubMed: 25931424]

Cleveland ES, & Reese E (2005). Maternal Structure and Autonomy Support in Conversations About the Past: Contributions to Children's Autobiographical Memory. Developmental Psychology, 41(2), 376–388. 10.1037/0012-1649.41.2.376 [PubMed: 15769193]

Cohen J (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20, 37–46.

Cohen SS, Ngo CT, Olson IR, & Newcombe NS (2025). Pattern separation and pattern completion in early childhood. Proceedings of the National Academy of Sciences, 122(11). 10.1073/pnas.2416985122

Cohen SS, Olson IR, & Newcombe NS (2025). The Component Structure of Memory during Development. [Dataset]. osf.io/ywr

Davies G, & Logie R (1993). Introduction. In Advances in Psychology (Vol. 100, pp. 1–9). Elsevier. 10.1016/S0166-4115(08)61088-2

Demir ÖE, Fisher JA, Goldin-Meadow S, & Levine SC (2014). Narrative processing in typically developing children and children with early unilateral brain injury: Seeing gesture matters. Developmental Psychology, 50(3), 815–828. 10.1037/a0034322 [PubMed: 24127729]

Ebbinghaus HE (1913). Memory: A contribution to experimental psychology. Teachers college, Columbia University.

Furman O, Dorfman N, Hasson U, Davachi L, & Dudai Y (2007). They saw a movie: Long-term memory for an extended audiovisual narrative. Learning & Memory, 14(6), 457–467. 10.1101/lm.550407 [PubMed: 17562897]

Gilboa A (2004). Autobiographical and episodic memory—One and the same? Neuropsychologia, 42(10), 1336–1349. 10.1016/j.neuropsychologia.2004.02.014 [PubMed: 15193941]

Haden CA (1998). Reminiscing With Different Children: Relating Maternal Stylistic Consistency and Sibling Similarity in Talk About the Past. Developmental Psychology, 34(1), 99. [PubMed: 9471008]

Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, & REDCap Consortium. (2019). The REDCap consortium: Building an international community of software platform partners. Journal of Biomedical Informatics, 95, 103208. [PubMed: 31078660]

Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, & Conde JG (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics, 42(2), 377–381. [PubMed: 18929686]

Horn JL (1965). A rationale and test for the number of factors in factor analysis. Psychometrika, 30, 179–185. [PubMed: 14306381]

Horner AJ, & Burgess N (2013). The associative structure of memory for multi-element events. Journal of Experimental Psychology: General, 142(4), 1370–1383. 10.1037/a0033626 [PubMed: 23915127]

Jack F, MacDonald S, Reese E, & Hayne H (2009). Maternal Reminiscing Style during Early Childhood Predicts the Age of Adolescents' Earliest Memories. Child Development, 80(2), 496–505. [PubMed: 19467006]

Kahana MJ (2012). Introduction. In Foundations of human memory (pp. 3–33). OUP.

Kandel ER, & Squire LR (2001). Neuroscience: Breaking down scientific barriers to the study of brain and mind. In Annals of the New York Academy of Science: Vol. 935. Unity of knowledge: The convergence of natural and human sciences (Vol. 935, pp. 118–135). New York Academy of Science.

Kaufman AS, & Kaufman NL (2004). Kaufman Brief Intelligence Test, Second Edition. Pearson, Inc.

Landis JR, & Koch GG (1977). The Measurement of Observer Agreement for Categorical Data. Biometrics, 33(1), 159. 10.2307/2529310 [PubMed: 843571]

Levine B, Svoboda E, Hay JF, Winocur G, & Moscovitch M (2002). Aging and autobiographical memory: Dissociating episodic from semantic retrieval. Psychology and Aging, 17(4), 677–689. 10.1037/0882-7974.17.4.677 [PubMed: 12507363]

Neisser U (1982). Memory: What are the important questions? In Memory observed: Remebering in natural contexts (pp. 3–19). Academic Press.

Nelson K, & Fivush R (2004). The Emergence of Autobiographical Memory: A Social Cultural Developmental Theory. Psychological Review, 111(2), 486–511. 10.1037/0033-295X.111.2.486 [PubMed: 15065919]

Newcombe NS, Benear S, Ngo CT, & Olson IR (2022). Memory in Infancy and Childhood. In Oxford Handbook on Human Memory.

Newcombe NS, & Nguyen KV (2024). Developmental Sequences Constrain Models of the Mind. PsyArXiv. 10.31234/osf.io/vhuwc

Ngo CT, Horner AJ, Newcombe NS, & Olson IR (2019). Development of Holistic Episodic Recollection. Psychological Science, 30(12), 1696–1706. [PubMed: 31672085]

Ngo CT, Michelmann S, Olson IR, & Newcombe NS (2021). Pattern separation and pattern completion: Behaviorally separable processes? Memory & Cognition, 49(1), 193–205. 10.3758/s13421-020-01072-y [PubMed: 32728851]

O'Reilly RC, & Rudy JW (2001). Conjunctive Representations in Learning and Memory: Principles of Cortical and Hippocampal Function. Psychological Review, 108(2), 311. [PubMed: 11381832]

Reese E (1995). Predicting children's literacy from mother-child conversations. Cognitive Development, 10(3), 381–405. 10.1016/0885-2014(95)90003-9

Reese E, & Brown N (2000). Reminiscing and recounting in the preschool years. Applied Cognitive Psychology, 14(1), 1–17. 10.1002/(SICI)1099-0720(200001)14:1<1::AID-ACP625>3.0.CO;2-G

Reese E, Haden CA, & Fivush R (1993). Mother-child conversations about the past: Relationships of style and memory over time. Cognitive Development, 8(4), 403–430. 10.1016/S0885-2014(05)80002-4

Saragosa-Harris NM, Cohen AO, Shen X, Sardar H, Alberini CM, & Hartley CA (2021). Associative memory persistence in 3- to 5-year-olds. Developmental Science, 24(5), e13105. 10.1111/desc.13105 [PubMed: 33626196]

Squire LR (1995). Biological foundations of accuracy and inaccuracy in memory. In Memory distortions: How minds, brains, and societies reconstruct the past (pp. 197–225). Harvard University Press.

Tulving E (1972). Episodic and semantic memory. In Organization of memory (pp. 382–403). Academic Press.

Tulving E (1983). Elements of Episodic Memory. Oxford University Press.

Tulving E (2002). Episodic Memory: From Mind to Brain. Annual Review of Psychology, 53(1), 1–25. 10.1146/annurev.psych.53.100901.135114

Tulving E, & Lepage M (2000). Where in the brain is the awareness of one's past? In Memory, brain, and belief (pp. 208–228). Harvard University Press.

Virk T, Letendre T, & Pathman T (2024). The convergence of naturalistic paradigms and cognitive neuroscience methods to investigate memory and its development. Neuropsychologia, 196, 108779. 10.1016/j.neuropsychologia.2023.108779 [PubMed: 38154592]

Yang BW, Deffler SA, & Marsh EJ (2022). A comparison of memories of fiction and autobiographical memories. Journal of Experimental Psychology: General, 151(5), 1089–1106. 10.1037/xge0001125 [PubMed: 34735187]

Yassa MA, & Stark CEL (2011). Pattern separation in the hippocampus. Trends in Neurosciences, 34(10), 515–525. 10.1016/j.tins.2011.06.006 [PubMed: 21788086]

Yonelinas AP (1994). Receiver-Operating Characteristics in Recognition Memory: Evidence for a Dual-Process Model. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(6), 1341. [PubMed: 7983467]

Yonelinas AP (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. Journal of Memory and Language, 46(3), 441–517. 10.1006/jmla.2002.2864

Author Manuscript
Author Manuscript
Author Manuscript

Yonelinas AP, Aly M, Wang W, & Koen JD (2010). Recollection and familiarity: Examining controversial assumptions and new directions. Hippocampus, 20(11), 1178–1194. 10.1002/hipo.20864 [PubMed: 20848606]

Zwick WR, & Velicer WF (1986). Comparison of five rules for determining the number of components to retain. Psychological Bulletin, 99(3), 342.

Public Significance Statement

The development of episodic memory is typically studied with highly controlled lab-based tasks. Here we find that these tasks have little relationship to real-world autobiographical memory and memory for short videos. This fact suggests that lab-based tasks may need significant modification to be informative about real-world episodic memory for events.

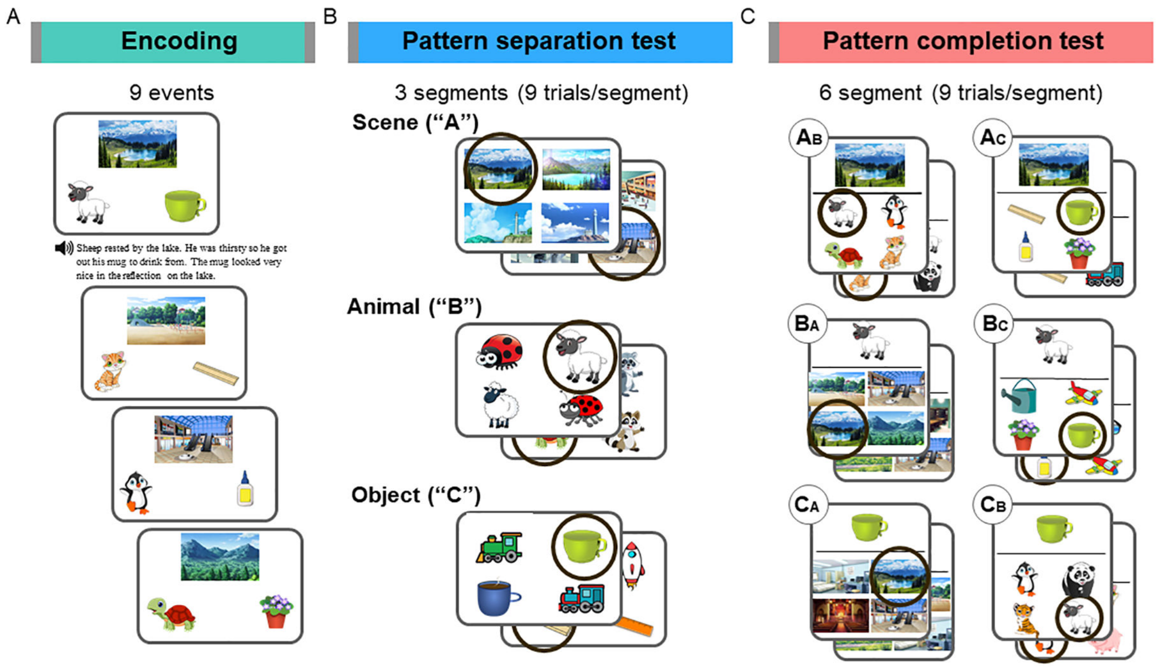

Figure 1.

A schematic depiction of the pattern separation and pattern completion task procedure for an example block. (A) Each block consisted of encoding nine event triads which were comprised of a scene, an animal, and an object. The triads and the items in them did not overlap between the sessions. Each event was presented for 15s while a pre-recorded three-sentence audio narrative was played. The event triads in each session were flanked by four example triads at the beginning and end of the encoding phase (represented as gray bars). These examples were then used to instruct the pattern separation and pattern completion tests (instruction indicated by grey bands in B and C). (B) The pattern separation test consisted of three segments, one for scenes, animals, and objects, each of which contained nine trials (to test for the items in each event). On each four-alternative forced choice (4AFC) trial, the four options included the correct target, a similar lure, and two exemplar foils. (C) The pattern completion test consisted of six segments that each tested the pairwise memory for the three items in each event. Thus, each segment contained nine trials (one per event). On each 4AFC trial, the cue was presented on top and the four options for the to-be-retrieved element were presented on the bottom of the screen. Scenes, animals, and objects are denoted as elements A, B, and C, respectively. Green circles indicate the correct option and are shown here for visual illustration (Figure adapted from S. S. Cohen, Ngo, et al., 2025).

Author Manuscript
Author Manuscript
Author Manuscript

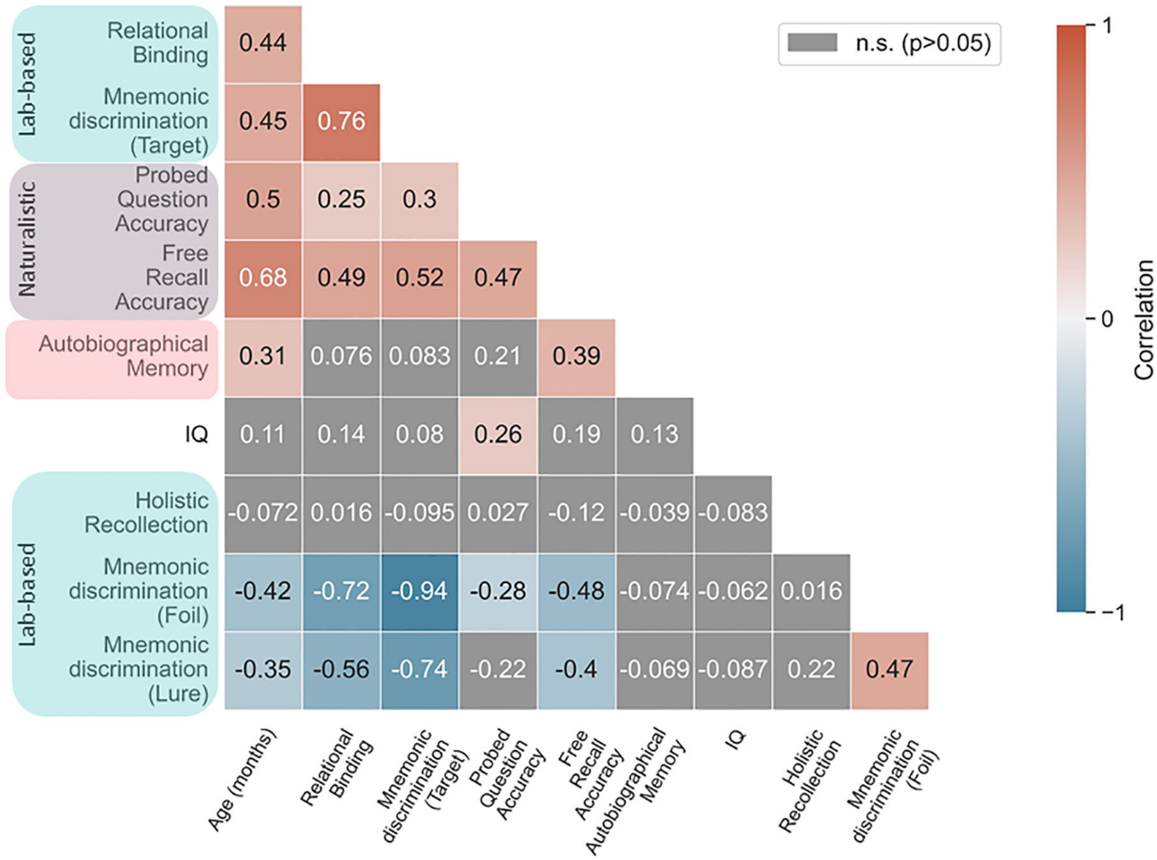

Figure 2.

Bivariate correlations between lab-based, naturalistic, and autobiographical memory measures, age, and IQ. The lab-based measures are relational binding, mnemonic discrimination (target), holistic recollection, mnemonic discrimination (foil), and mnemonic discrimination (lure) (highlighted in blue, “Lab-based”). The naturalistic measures are measures of memory for short cartoons assessed with probed question accuracy and free recall accuracy (highlighted in purple, “Naturalistic”). Our most naturalistic measure of memory is autobiographical memory, or memory for real world events that occurred outside of the lab (highlighted in pink). Significant correlations (p<0.05) are colored according to correlation value and non-significant correlations (n.s., p>0.05) are colored in grey.

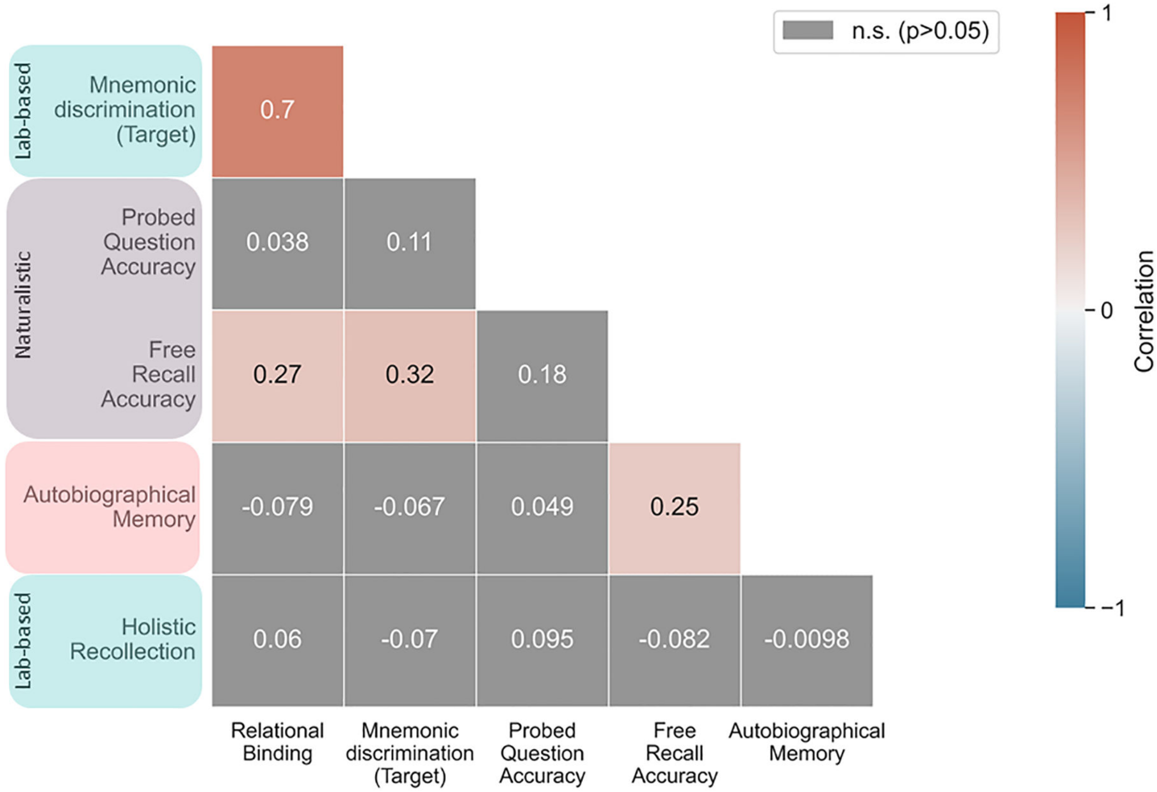

Figure 3.

Cross correlations between measured variables of episodic memory after controlling for participant age and IQ ("Lab-based" tasks highlighted in blue, "Naturalistic" tasks highlighted in purple, and Autobiographical Memory highlighted in pink). Significant correlations ($p<0.05$) are colored according to correlation value and non-significant correlations (n.s., $p>0.05$) are colored in grey.

Table 1.

Exploratory factor analysis factor loadings. Factor 1 is mostly comprised of lab-based measures of memory (relational binding and mnemonic discrimination). Factor 2 is mostly comprised of naturalistic and real-world measures of memory such as free recall accuracy and autobiographical memory.

	Memory measure	Factor 1	Factor 2
	Relational Binding	**0.90**	– 0.05
Lab-based Measures	Mnemonic Discrimination	**0.90**	0.02
Memory for naturalistic events	Probed Question Accuracy	0.10	**0.46**
	Free Recall Accuracy	0.39	**0.67**
Real-world memory	Autobiographical Memory	– 0.24	**0.79**