

Children Develop Adult-Like Visual Sensitivity to Image Memorability by the Age of 4

Xiaohan (Hannah) Guo¹ and Wilma A. Bainbridge^{1, 2}

¹Department of Psychology, The University of Chicago

²Neuroscience Institute, The University of Chicago

Adults have been shown to consistently remember and forget certain images despite large individual differences, suggesting a population-wide sensitivity to an image's intrinsic memorability—a measure of how successfully an image is remembered. While a decade of research has focused on image memorability among adults, the developmental trajectory of these consistencies in memory is understudied. Here, we investigate by what age children gain adult-like sensitivity to the image memorability effect. We utilized data from Saragosa-Harris et al. (2021), where 137 children aged between 3 and 5 years old encoded animal-scene image pairs and then after a 5-min, 24-hr, or 1-week delay performed a cued recognition task for each scene target given its animal cue. We tested adults' memory of the same scene images using ResMem (Needell & Bainbridge, 2022), a pretrained deep neural network that predicts adult image memorability scores, and using an online behavioral continuous recognition task ($N = 116$). Results showed that ResMem predictions, as a proxy of adults' memory, predicted scene memory of children by the age of 4 and were the most predictive of children's memory across ages after a long, 1-week delay. Children at age 3 show nonadult-like consistent memory patterns, implying that the nonadult-like memory patterns were not due to poor memory performance. Instead, 3-year-olds may have consistently used certain visual memory strategies that become less optimal as they age. Our results suggest that adult-like sensitivity to image memorability emerges by the age of 4 through experience.

Public Significance Statement

This study strongly suggests that children older than 4 years old tend to remember and forget the same images as adults. We recommend teachers and caregivers utilize the ResMem deep neural network to select memorable images to be used in educational settings.

Keywords: memory development, memorability, scene image memory, deep neural network, associative memory task

Supplemental materials: <https://doi.org/10.1037/xge0001511.supp>

We experience a continuous, endless stream of visual sensory information daily, where some visual information is remembered, and other information is forgotten. Despite large individual differences in personal experience and memory performance, there is an overwhelming consistency in what visual information adults tend to remember and forget (Bainbridge et al., 2013; Isola et al.,

2011). Indeed, some scenes (Isola et al., 2011), faces (Bainbridge et al., 2013), words (Xie et al., 2020), videos (Cohendet et al., 2018), and even dance movements (Ongchoco et al., 2023) are intrinsically more memorable than others. This phenomenon whereby adults consistently remember and forget certain visual stimuli can be captured by memorability (Bainbridge et al., 2013; Isola

This article was published Online First December 7, 2023.

Xiaohan (Hannah) Guo  <https://orcid.org/0000-0002-8938-5458>

Wilma A. Bainbridge  <https://orcid.org/0000-0002-7554-0736>

This work was in part supported by the National Science Foundation (Grant 2329776). The authors thank the Hartley Lab at New York University for making the Saragosa-Harris et al. (2021) data set available and for providing useful feedback. The authors also thank the BLRB Group at the University of Chicago for providing insightful comments. The authors have no conflicts of interest to disclose. Data from this study can be accessed through Open Science Framework following this link: <https://osf.io/pz26c/>. Portions of these findings were presented as a poster at the 2022 Vision Sciences Society, St. Pete Beach, Florida, United States. These findings will be presented at the 2023 Object Perception, Visual Attention, and Visual Memory (OPAM), San Francisco, California,

United States. This work was not preregistered. A preprint was posted on *bioRxiv*: <https://doi.org/10.1101/2022.12.20.520853>.

Xiaohan (Hannah) Guo served as lead for data curation, formal analysis, investigation, project administration, validation, visualization, and writing—original draft and contributed equally to methodology. Wilma A. Bainbridge served as lead for supervision and served in a supporting role for data curation, formal analysis, investigation, methodology, resources, validation, visualization, and writing—original draft. Xiaohan (Hannah) Guo and Wilma A. Bainbridge contributed equally to conceptualization and writing—review and editing.

Correspondence concerning this article should be addressed to Xiaohan (Hannah) Guo, Department of Psychology, The University of Chicago, 301 Green Hall, 5848 South University Avenue, Chicago, IL 60637, United States. Email: hannahguo@uchicago.edu

et al., 2011), an objective stimulus measure indicating how well observers remember certain sensory information, across the adult population. How and when the susceptibility to memorability emerges across development, however, is understudied. It remains unclear whether children also consistently remember and forget certain visual stimuli similar to adults, and if so, whether memorability scores derived from adults can be used to predict children's memory patterns. In this study, we ask how and when in childhood humans develop adult-like consistent visual memory patterns.

One hypothesis is that humans develop susceptibilities to memorability in early childhood, driven by the development of specific brain structures or the accumulation of visual experience. Human and animal research together has highlighted a critical period in early childhood when the hippocampus undergoes a decline in neurogenesis and an increase in competence to store long-term memories (Akers et al., 2014; Alberini & Travaglia, 2017; Dennis et al., 2016; Josselyn & Frankland, 2012; Travaglia et al., 2016). Such neural development eliminates infantile amnesia, the phenomenon whereby episodic memories in early childhood quickly fade away and are rarely remembered in adulthood (Hayne, 2004). In humans, infantile amnesia fades away at 3–4 years of age (Bauer et al., 2011; Schneider & Pressley, 2013), implying a change in children's memory from nonadult-like to more adult-like in this period due to rapid hippocampal development. Human functional magnetic resonance imaging (fMRI) research has found that the medial temporal lobe including the hippocampus may be sensitive to the memorability of images (Bainbridge & Rissman, 2018; Bainbridge et al., 2017). Thus, it is possible that children develop a susceptibility to the memorability effect along with the resilience to infantile amnesia in the period of hippocampal development at ages 3 and 4. In addition, the accumulation of visual experience in the first few years of life could gradually train children's visual memory to become more adult-like. Memorability may reflect the statistical properties of stimuli in our world (Hovhannisyan et al., 2021), such as co-occurrence statistics (Xie et al., 2020) or across-item similarity (Lukavský & Děchtěrenko, 2017, although see Kramer et al., 2023). Thus, consistencies in memory could develop as children learn these statistical regularities of visual stimuli.

An alternative hypothesis is that children possess adult-like susceptibility to the memorability effect early on, and we would expect young children to have similar memory patterns to adults. Previous research has found that infants are born with capabilities to process faces (Farah et al., 2000; Tzourio-Mazoyer et al., 2002), scenes (Kamps et al., 2020), and edges (Kessen et al., 1972). These commonalities in visual processing between newborns and adults indicate early sensitivity to certain visual categories and properties, occurring in late visual areas in the brain (i.e., the inferotemporal cortex). Magnetoencephalography (MEG) studies have found that adults' sensitivity to memorability occurs rapidly at the time scale of late visual processing (170–250 ms; Khaligh-Razavi et al., 2016; Mohsenzadeh et al., 2019), and an fMRI study found adult sensitivity to memorability in inferotemporal cortex (Bainbridge et al., 2017). Thus, like their early sensitivity to visual categories, it is possible that infants could possess sensitivity to memorability in late visual areas from birth. Another piece of supportive evidence for early sensitivity to the visual memorability effect lies in primate research. Despite deprived visual experience compared to humans, laboratory-reared monkeys showed higher magnitudes of brain activity in late visual areas when viewing real-world images (e.g., fire hydrant and grocery store) that were more memorable to humans

(Jaegle et al., 2019). Such commonalities in visual processing across species suggest that limited visual experience is required to develop a susceptibility to the memorability effect, providing supportive evidence that infants could have similar visual sensitivity to adults and thus exhibit adult-like visual memory patterns very early on.

One power of memorability is that memorability scores are highly consistent across different tasks and observers (Bainbridge, 2020; Goetschalckx et al., 2019). The classic method of computing memorability scores is to use a continuous recognition task, where a timed sequence of stimuli is shown, and participants are asked to indicate stimulus repeats (Bainbridge et al., 2013; Isola et al., 2011). Another way of computing memorability scores is to use a paired association task, where a timed sequence of stimulus pairs is encoded, and after a delay, participants are asked to indicate the target stimulus based on its paired counterpart as a cue (Xie et al., 2020). Previous results suggest that the memorability of the target stimulus rather than the cue drives memory performance. Memorability can also be measured from one adult population to successfully predict the memory of another adult population (e.g., from online participants to patients with intracranial electrodes; Xie et al., 2020). Thus, memorability can be measured in one group or task and used to predict memory in another group or task, and we can apply this principle to predicting children's memory from adult behavior. It is also possible to predict memorability scores using a pretrained deep neural network (DNN) called ResMem (Needell & Bainbridge, 2022), which can significantly predict image memorability for adults (a correlation of .67 between predicted and true memorability scores). Knowing that memorability scores obtained across paradigms robustly reflect adults' memory, in the current study, we intend to test children's development of adult-like visual memory using itemized memorability obtained using neural network predictions and adult memory behavior.

The current study serves as one of the first studies to investigate the development of sensitivity to image memorability, investigating the two previously mentioned hypotheses. We ask by what age children have developed adult-like visual sensitivity to the image memorability effect. We ran a series of experiments comparing children's memory with adult memory measures for a set of scene images, and found that children develop adult-like sensitivity to image memorability by the age of 4. This memory pattern was best predicted by ResMem predictions of adult memorability, when the children were tested after a 1-week delay. Furthermore, we found evidence that children at the age of 3 consistently used certain encoding strategies that were not applied by the older age groups. Previous research has shown that young children have less accurate episodic memory (Zimmermann & Meier, 2006) and worse recognition memory performance (Lorsbach & Reimer, 2005) compared to adults. Despite children's worse memory, our results imply that at least one aspect of memory—the sensitivity to the memorability of images—becomes adult-like in early childhood while memory is still developing. Furthermore, we have identified a DNN able to predict children's memory, empowering educators and caregivers with a tool able to create more intuitive learning materials for young children.

Method

Child Memory Data Set

We utilized children's image memory data from Saragosa-Harris et al. (2021; Open Science Framework link: <https://osf.io/4m9kx/>),

where 3- to 5-year-old children encoded and retrieved animal-scene image pairs in an associative memory task. We measured children's memory accuracy (percent correct) for each scene and animal image using these memory test data.

In the original experiment (Saragosa-Harris et al., 2021), a sample of 137 children (N female = 72) aged 3–5 years old (forty-five 3-year-olds, forty-five 4-year-olds, and forty-seven 5-year-olds) first encoded eight animal-scene associative pairs and was then tested with a cued recognition task after a delay. Child participants in each age group were evenly assigned to three delay conditions, where memory was assessed after a 5-min ($N = 45$; fifteen 3-year-olds, fifteen 4-year-olds, and fifteen 5-year-olds), a 24-hr ($N = 45$; fifteen 3-year-olds, fifteen 4-year-olds, and fifteen 5-year-olds), or a 1-week ($N = 47$; fifteen 3-year-olds, fifteen 4-year-olds, and seventeen 5-year-olds) delay.

As shown in Figure 1, during the encoding phase, child participants were presented with eight unique trials of animal-scene associative pairs in a timed sequence. Within each association trial, an interactive storybook showed children an isolated animal image in full screen, then a scene image as the animal's "favorite place" on the next screen, followed by a third screen showing an image that placed the animal within the same scene image. Each screen was presented for 5 s with speech narration. The pool of images consisted of eight animals and 20 scenes. There were four storybook versions in total, and each contained the same eight animals paired with eight different scenes. Children were randomly assigned to one of the four storybook versions (Storybook 1: $N = 37$; Storybook 2: $N = 34$; Storybook 3: $N = 33$; Storybook 4: $N = 33$). During the memory test phase (after 5 min, 24 hr, or 1 week depending on condition), child participants completed eight 4AFC (alternative forced-choice) trials with speech narration. Within each test trial, child participants were shown one of the eight animal images (cue) on a screen for 5 s, followed by a screen for an unlimited time with the cue animal in the center and four scene options surrounding the animal. The four scene options consisted of the correct scene (target) associated with the animal during encoding, two lure scenes associated with

other animals during encoding, and a novel scene missing from the encoding phase. All scene choices came from the pool of 20 scene images. Child participants did not receive feedback on whether they chose the correct scene. Saragosa-Harris et al. (2021) observed that children's memory accuracy increased with age and decreased with longer delays.

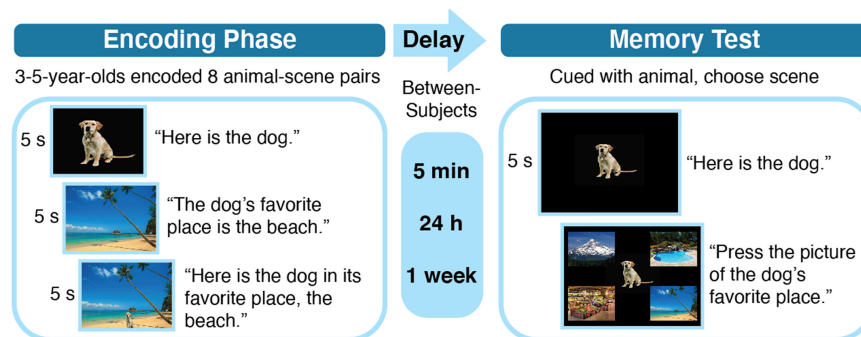
ResMem Memorability Score Predictions

We inputted the 59 target scenes and the 60 target animals from the adult experiment (see Adult Memory Experiment section), including images from the child data set, into ResMem and extracted a predicted memorability score (hit rate) for each image. ResMem is a pretrained residual DNN, which combines AlexNet feature layers that capture low-level visual information with ResNet feature layers that capture high-level semantic information of the inputted images (Needell & Bainbridge, 2022). ResMem was trained to predict image memorability using large-scale adult memory performance data obtained in similar continuous recognition tasks as the adult memory experiment in this study. The predicted memorability score ResMem assigns to each image, ranging from 0 to 1, indicates the probability that this image is successfully recognized among adults when repeated. Images with higher predicted memorability scores are better remembered by adults (Spearman correlation $\rho = .67$; Needell & Bainbridge, 2022). Because ResMem's correlation of .67 with adult memory performance is close to the noise ceiling of .68 estimated by correlating the memory of random split-halves of adults (Goetschalckx et al., 2019; Khosla et al., 2015), ResMem predictions can be treated as a proxy for adults' memory.

Adult Memory Experiment

To confirm that the memorability scores predicted by ResMem DNN are a reliable proxy for adults' memory performance, we ran two continuous recognition experiments among adults and calculated the memorability score (hit rates) for each of the 59 scene

Figure 1
Procedure to Obtain Children's Memory



Note. During the encoding phase (left panel), a speech narration described for child participants each animal, then its favorite place, followed by the animal in the place. Each screen was presented for 5 s. Each child participant encoded eight animal-scene pairs. After a 5-min, 24-hr, or 1-week delay (between subjects), child participants took a memory test (right panel) consisting of eight trials. Within each trial, participants first viewed an animal cue for 5 s with narration and were then asked to choose the correct associated scene (target) out of four scenes. See the online article for the color version of this figure.

images and 60 animal images. Note that while children's memory was measured with an associative memory task, here adult memory is measured with an item-based continuous recognition task. Prior work has shown that memorability measures are stable across different task parameters and image contexts, and can be measured using both item-based and associative tasks (Bainbridge, 2020; Bylinskii et al., 2015; Xie et al., 2020). However, we revisit the differences between these two tasks in the Discussion.

Participants

One hundred sixteen online participants on Amazon Mechanical Turk engaged in a scene memory experiment, where a timed sequence of scene images was shown, and 148 online participants engaged in an animal memory experiment, where a timed sequence of animal images was shown. Prior work showed that having 80 participants in a continuous recognition task was optimal for estimating memorability scores for each image (Isola et al., 2014). We recruited 160 online participants for each experiment to account for potential exclusions. After applying the exclusion criteria (see the end of the Procedure subsection), 116 participants from the scene experiment and 148 participants from the animal experiment were included for analysis. Participants consented to their participation following the guidelines of the Institutional Review Board (IRB) at the University of Chicago (IRB19-1395). Participants could participate in both experiments, and each participant was awarded a \$0.70 compensation upon completion. Each experiment took approximately 7 min, and participants were allowed to end the experiment at any time.

Stimuli

A pool of 136 scene images was shown in the scene experiment, including the 20 scene images from the child data set and 116 from the Scene Understanding (SUN) database (Xiao et al., 2010). From the pool of 136 scene images, 76 were used as filler images, and 60 were used as target images, including the 20 from the child data set. We included more images in the adult continuous recognition tasks beyond the 20 images from the child data set in order to increase task difficulty. The higher number of images allows for a larger time interval between target encoding and recognition, and also ensures enough statistical power during image-based analyses. We balanced the proportions of indoor, natural outdoor, and manmade outdoor images between the target and the filler sets. To avoid images coming from the child data set versus from the SUN database being visually different and thus leading to systematic variations in visual memory, we prioritized selecting scenes from the SUN database which contained themes (e.g., forest, sea, and house) that were similar to the 20 scenes from the child data set. For both experiments, we ensured that the image set from the child data set versus the one we adapted online did not differ in low-level visual features using the Natural Image Statistical Toolbox (Bainbridge & Oliva, 2015; all $ps > .05$). All images were resized to 256×256 pixels. One target scene image from the SUN database was misaligned in dimensions with the rest of the scene images and was excluded from analysis in the Results section, resulting in 59 scene images that we used for analyses.

A pool of 136 animal images was shown in the animal experiment, including eight animal images from the child data set and 128 from Google Images (with Creative Commons licenses). To ensure that the two sets of animals were visually similar, we selected

Google images that contained isolated animals with neutral poses and expressions. We photoshopped the Google images following the format of the eight animal images from the child data set by placing each isolated animal in the middle of a black background. The 136 animals came from different species that are common in American zoos and aquariums. Similar to the scene experiment, 76 animal images were used as fillers, and 60 were used as targets, including the eight from the child data set. We balanced the proportions of mammals, fish, reptiles, insects, and birds between the target and the filler sets. All images were resized to 256×256 pixels.

Procedure

After answering three common-sense questions at the beginning of each experiment (e.g., "What is the name of the last month of the year?" and "What is the capital of the United States?"), participants were directed to the main task. The main task adopted a continuous recognition paradigm frequently used to test image memorability (Bainbridge et al., 2013), where participants were shown a timed sequence of images and were instructed to press the *r* key when they recognized an image from earlier in the sequence (Figure 2). Each image was presented for 750 ms with an 800-ms interstimulus interval where participants looked at a fixation cross. Each of the 136 images within each experiment was repeated at most once. All target images, but not all filler images, were repeated. Target repetitions were spaced at least 30 s apart. Filler repetitions were spaced 0–4 images apart and were included to maintain participant attention.

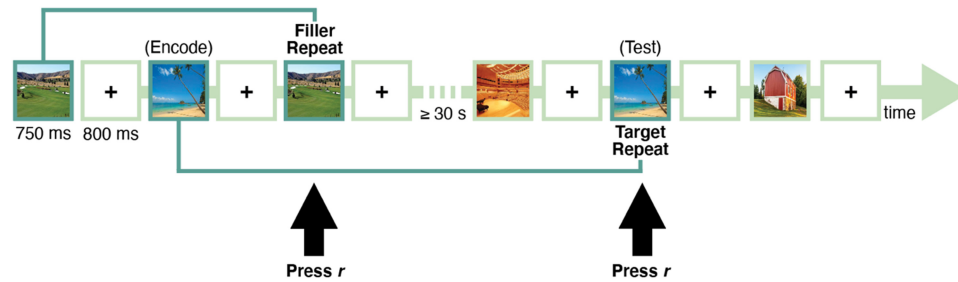
Adult participants did not receive feedback during the continuous recognition experiments for two reasons. While some studies (Almog et al., 2021; Bylinskii et al., 2015; Isola et al., 2011) provide feedback when testing image memorability, other work has not provided feedback (Bainbridge et al., 2013; Ongchoco et al., 2023), and even identified memorability effects during implicit memory, when participants were not intentionally memorizing the images (Bainbridge, 2020; Goetschalckx et al., 2019). These findings confirm a robust memorability effect regardless of whether feedback is provided. We opted to not provide feedback in order to align the design with the child experiment, where feedback was also not provided.

To ensure the quality of participants, only workers with an Amazon Mechanical Turk approval rate of at least 95% and the number of approved tasks of at least 50 were allowed to participate. Only workers with an IP address within the United States were allowed to participate. Participants who answered any of the three common-sense questions incorrectly were excluded. Participants who paid no attention and indicated repetitions for over 95% of the images in the timed sequence (i.e., kept pressing the *r* key) were excluded. We also used filler repetitions as a vigilance test to exclude participants who paid little attention during the experiments. Participants with over 70% false alarms on the first presentation of filler images and over 70% misses on vigilance repeats were excluded.

Analyses

Calculating Memory Performance

To compare children's and adults' memory, we generated item-wise memory measures between 0 and 1 for target images from the child data set and the adult experiment. We obtained children's memory accuracy for each of the 20 scene images by calculating the

Figure 2*A Flowchart of the Adult Visual Memory Experiment for Scene Images*

Note. Each image was presented for 750 ms, followed by an 800-ms fixation cross. Participants pressed *r* when they saw an image repeat. Memorability scores were derived for target images, which were repeated at least 30 s apart. Filler repetitions were spaced 0–4 images (<6 s) apart. The adult experiment for animal images followed the same paradigm. See the online article for the color version of this figure.

percentage of trials a scene was chosen when it was the target scene (percent correct). We obtained children's memory accuracy for each of the eight animal images by calculating the percentage of trials the target scene was chosen when cued with an animal. To obtain adults' memory accuracy, we measured the adult experiment hit rates (adult HR) on image repetitions of the 59 target scene images and the 60 target animal images which included images from the child data set. Finally, we also obtained predictions of memorability scores (hit rates) for each of the images through the ResMem DNN (ResMem HR).

We measured adult experiment false alarm rates (FAR) on the first presentation of the 59 scene images and the 60 animal images (see [Results SR1 and Figure 1 in the online supplemental materials](#)). However, we decided to compare children's and adults' memories using accuracy and HR rather than false-alarm-corrected hit rates (FA-corrected HR) because we could not compute valid FAR from ResMem or the child data set. Specifically, ResMem predicts image hit rates that are not corrected for false alarms. Furthermore, in the child experiment, foil images were not independent of the other images due to nature of the 4AFC task, preventing valid FAR from being measured with the child data. That being said, if we calculated proportion of times an image was chosen when it was a novel foil image, we also did not find consistency in behavior within any age or delay condition (see [Results SR2 in the online supplemental materials](#)), suggesting the use of children FAR may not be viable here.

Figure 3 shows examples of scene and animal images that are memorable and forgettable according to memory scores extracted from ResMem, the child data set, and the adults' experiments. We set an False Discovery Rate (FDR) threshold ($q < 0.05$) a priori for multiple comparisons correction when testing children's memory across different ages and delays.

Logistic Regression Model

We generated a logistic regression model to predict children's memory accuracy using trial-wise test data from the child data set, where each child answered eight test trials. Logistic regression was chosen because trial outcomes were binary (i.e., correct vs. incorrect answers). Age group (3-year-olds, 4-year-olds, and 5-year-olds) and the three delay conditions (5-min, 24-hr, and 1-week) were entered as

categorical variables into the regression model. We z-transformed continuous data before fitting the model. The model included ResMem-predicted memorability score (ResMem HR) of the target scene in each test trial, age, and delay as predictors:

$$\begin{aligned} \text{Children's memory accuracy} \sim & \beta_0 + \beta_1(\text{ResMem HR}_{\text{scenes}}) \\ & + \beta_2(\text{Age } 4) + \beta_3(\text{Age } 5) \\ & + \beta_4(\text{Delay } 2) + \beta_5(\text{Delay } 3) \\ & + (1 | \text{subject ID}) \\ & + (1 | \text{associated animal ID}) \end{aligned}$$

Note that we included two random intercepts. The random intercept of child participant ID accounted for individual differences across children. The random intercept of the associated animal image ID (i.e., 1–8) in each test trial accounted for the fact that child participants encoded scene images in association with animal images. The rationale was that even if different children encoded an animal associated with different scenes, the same animal might affect children's memories of different scenes in a similar way. We call this model the base model (see [Results SR3 and SR4 and Table 1.1 in the online supplemental materials](#) for other models we tested; this base model had the best fit metrics, adjusted R^2 and AIC, the Akaike Information Criterion, of all

Figure 3*Example Images From the Child Data Set and Memory Measures*

Note. Left: Examples of memorable and forgettable scene images, marked with single-score measures of ResMem-predicted memorability (ResMem HR), children's memory accuracy, and adults' memory (adult HR) for each image. Right: More examples of scene and animal images from the child memory data set. ResMem HR = ResMem hit rates; adult HR = adult hit rates. See the online article for the color version of this figure.

models). We also tested the same model with continuous rather than categorical children's age as a predictor (see Table 1.2 in the online supplemental materials for statistics).

Constraints on Generality

We used an online sample representative of the U.S. population to collect adults' behavioral data. ResMem was also trained using adults' behavioral data from an online sample representative of the U.S. population. However, these online participants were adults in the United States with internet access, which may limit the generalization of results in this study to all human adults, even though memorability has been shown to be robust across populations. In addition, the original child data set came from a New York area sample, which could limit the generalizability of the results.

Transparency and Openness

We report how we determined our sample size, all data exclusion criteria, all manipulations, and all measures in the study. All data are available at <https://osf.io/pz26c/>. Data were analyzed using MATLAB, Version 9.11.0.1769968 (R2021b). This study's design and its analysis were not preregistered.

Results

We have generated single-score measures of children's memory (percent correct), ResMem-predicted memorability (ResMem HR), and adults' memory (adult HR) for all target images. To find out how early children have developed adult-like visual memory patterns, we compared children's memory accuracy with image memorability extracted from ResMem and the visual memory behavioral experiment (adult experiment). Because images with higher hit rates are more likely to be remembered consistently across adults (Bainbridge et al., 2013; Isola et al., 2011), we could use whether images with higher adult HR were also better remembered by children as an indicator of whether children have developed adult-like visual memory by a certain age. As an overview, we first compared ResMem HR with adult HR on the same set of images to ensure that ResMem predictions successfully captured the image memorability effects on adults. We then compared children's memory accuracy to ResMem HR and adult HR, respectively, on the same set of images to see whether ResMem predictions and adults' memory could predict children's image memory.

Does ResMem Predict Adults' Memory?

As a sanity check for the adults' memory data, we tested the reliability of adult HR for the 59 scene targets ($M = 0.58$, $SD = 0.08$, $\min = 0.43$, $\max = 0.74$) by splitting adults into random halves and running a Spearman rank correlation between the 59 memorability scores from each half, as shown in Figure 4A. A higher similarity between the two score ranks would indicate higher consistency in adults' memory. Across 1,000 of these random split-half trials, adults were significantly consistent in the scenes they remembered, with an averaged Spearman–Brown corrected ρ of .73 ($p < .001$, calculated using a permutation test with 1,000 shuffled rankings). They were also significantly consistent in their memory performance for the 60 animal targets (Figure 4B; $M = 0.59$, $SD = 0.08$, $\min = 0.43$, $\max = 0.76$, averaged Spearman–Brown corrected, $\rho = .79$,

$p < .001$). Adults were also significantly consistent in their memory for just the 20 scene images from the child data set ($M = 0.65$, $SD = 0.07$; averaged Spearman–Brown corrected $\rho = .69$, $p = .018$) and the 28 animal and scene images from that data set ($M = 0.67$, $SD = 0.07$; averaged Spearman–Brown corrected, $\rho = .69$, $p = .003$; see Figure 2 in the online supplemental materials). These results showed that adults were indeed highly consistent in what scene and animal images they remembered and forgot.

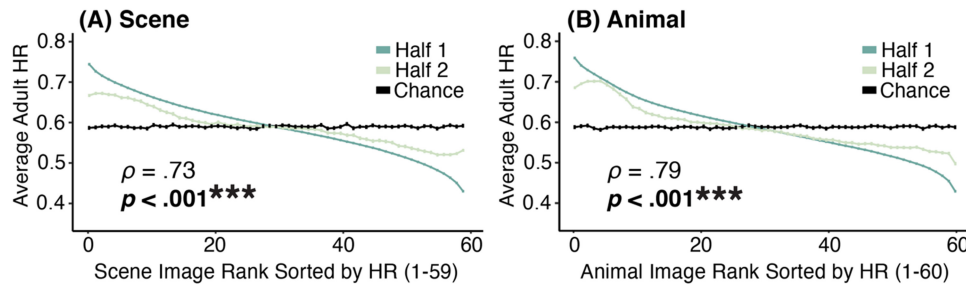
We then tested whether ResMem predicted higher memorability scores for images that were better remembered by adults, given that ResMem was trained to predict image memorability using large-scale adult's memory data on photographs. Such a sanity check provides a benchmark of whether ResMem predictions can be treated as a proxy of adults' memory. To see whether ResMem ($M = 0.67$, $SD = 0.11$, $\min = 0.41$, $\max = 0.92$) successfully assigned higher memorability scores to more memorable scene images, we ran a Spearman correlation between the 59 memorability scores extracted from ResMem and those calculated from adults' memory data, as shown in Figure 5A. As expected, ResMem HR significantly correlated with adult HR for the scene images ($\rho = .54$, $p < .001$). This result means that ResMem HR can be treated as a reliable proxy of adults' memory of the scene images in further analyses. Figure 5B shows results of the same Spearman correlation performed over the 60 animal targets. Results showed that ResMem HR ($M = 0.85$, $SD = 0.07$, $\min = 0.67$, $\max = 0.97$) did not correlate with adult HR of animal images ($\rho = .22$, $p = .90$), meaning that ResMem HR was not a good substitution for adults' memory of animal images. We found the same results when adult HR was corrected for false alarms (scene: $p = .003$; animal: $p > .05$; see Table 2 in the online supplemental materials for statistics). We revisit ResMem's performance with the animal images in the Discussion section. Because ResMem HR could not be treated as a reliable proxy of adults' memory for the animal images, we will mainly focus on results for scene images.

Does ResMem Predict Children's Memory?

The central question of this study is by what age children show adult-like visual memory patterns. To see which children's age group remembered memorable images to adults more than forgettable ones, we Spearman correlated ResMem-predicted memorability with children's memory accuracy within each age group across delay conditions (3-year-olds: $M = 0.52$, $SD = 0.19$; 4-year-olds: $M = 0.62$, $SD = 0.15$; 5-year-olds: $M = 0.76$, $SD = 0.07$) using the 20 scene images from the child data set. Figure 6A–C shows the crucial results: ResMem HR did not predict children's image memory at the age of 3 ($\rho = -.15$, $p = .53$), but significantly predicted children's image memory at the ages of 4 ($\rho = .52$, $p = .02$) and 5 ($\rho = .58$, $p = .007$; FDR-corrected, $q < 0.05$). This means that children develop adult-like sensitivity to image memorability by the age of 4. Fisher's z tests confirmed that 3-year-olds were significantly less predictable by ResMem than 4- and 5-year-olds (3-year-olds vs. 4-year-olds: Fisher's $z = -3.33$, $p = .001$; 3-year-olds vs. 5-year-olds: Fisher's $z = -3.77$, $p < .001$; 4-year-olds vs. 5-year-olds: Fisher's $z = -.4$, $p > .05$; two-tailed; FDR-corrected $q < 0.05$).

To see whether different delay intervals between encoding and memory test phases affected the image memorability effect on children's memory, we Spearman correlated ResMem HR with children's memory within each delay condition across age groups (5-min delay:

Figure 4
Consistency Graphs for Adult HR



Note. (A) Consistency graph for adult HR of the 59 scene images. (B) Consistency graph for adult HR of the 60 animal images. Adult participants were split into random halves in each of the 1,000 iterations. The memorability scores calculated from the first half of the participants (Half 1: dark green/gray line) were ranked from high to low, and those from the second half (Half 2: light green/gray line) were ranked using the image rank of Half 1. The two green/gray lines represent the average of adult HR ranks over the 1,000 iterations. If the two halves of participants are perfectly consistent in memory, the green/gray lines should coincide. In contrast, if the two halves of participants are not consistent in memory, the light green/gray line should coincide with the chance line in black, which represents a random shuffle of image ranks in Half 1. The significant Spearman rank correlation indicates that split halves of participants tend to remember and forget the same images. Error bars represent standard error of the mean. Adult HR = adult hit rates. See the online article for the color version of this figure.

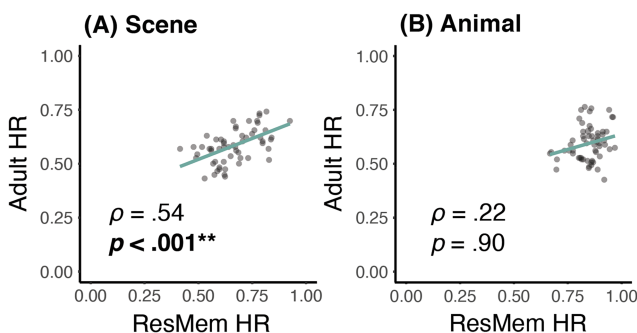
*** $p < .001$.

$M = 0.79$, $SD = 0.09$; 24-h delay: $M = 0.66$, $SD = 0.13$; 1-week delay: $M = 0.47$, $SD = 0.11$) using the same 20 scenes. ResMem HR did not predict children's image memory across ages after a 5-min ($p = .29$, $p = .21$) or a 24-hr delay ($p = -.15$, $p = .52$), but significantly predicted children's image memory after a 1-week delay ($p = .55$, $p = .01$; FDR-corrected, $q < 0.05$), as shown in Figure 6D–F. We confirmed these results by combining 4- and 5-year-olds into one group and running the same Spearman correlation within each delay condition. We found the same results, where ResMem HR

predicted children's memory after a 1-week delay ($p = .65$, $p = .002$; FDR-corrected, $q < 0.05$) but not after shorter delays ($ps > .05$). This means that the memorability effect drives children's image memory accuracy after long delays. Fisher's z tests confirmed a change in predictability of ResMem on children's memory between the 24-hr and the 1-week delays but not between the 5-min and the longer delays (5-min vs. 24-hr: Fisher's $z = 2.06$, $p > .05$; 5-min vs. 1-week: Fisher's $z = -1.48$, $p > .05$; 24-hr vs. 1-week: Fisher's $z = 3.57$, $p = .001$; two-tailed; FDR-corrected $q < 0.05$).

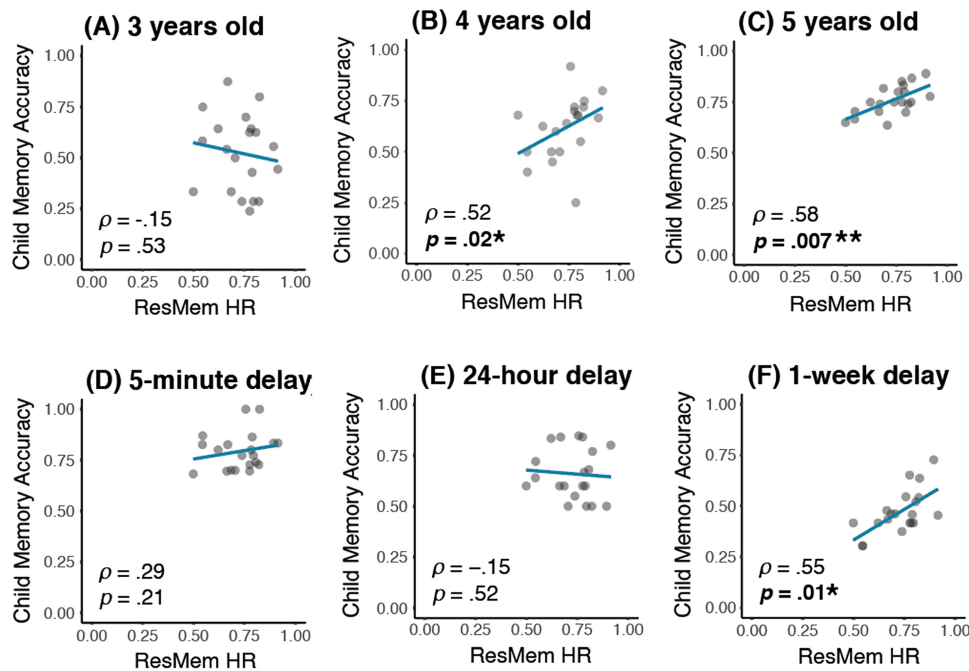
Knowing that both age and delay could influence how well ResMem predicted children's image memory, we wanted to create a statistical model that could predict children's memory using ResMem HR by specifying children's ages and delay intervals. We fitted the 1,096 test trials from the child data set to our base regression model (see Method), which included ResMem HR predictions for scenes, children's age, delay, and random intercepts of subject ID and associated animal ID, and found a significant model fit (adjusted $R^2 = .39$, $p < .001$, AIC = 5,069). As expected, we found that the predictor of ResMem HR for scenes was significant ($\beta_{\text{ResMem HR Scenes}} = .85$, $p = .01$). We also found that both age predictors were significant ($\beta_{\text{Age 4}} = .70$, $p = .02$; $\beta_{\text{Age 5}} = 1.31$, $p < .001$). Both delay predictors were also significant ($\beta_{\text{Delay 2}} = -.94$, $p = .002$; $\beta_{\text{Delay 3}} = -1.95$, $p < .001$). This means that the intrinsic memorability of scenes, children's age, and time delay all had a significant effect on children's memory. While we tested other model structures, this base model led to the best model fit (all models $p < .001$, the base model had the highest adjusted R^2 and the lowest AIC; see Table 1.1 in the online supplemental materials for statistics of all models). However, we found the same results when we instead used continuous rather than categorical children's age as a predictor in the logistic regression models. The base model was still the model with the best fit, and ResMem HR for scenes, continuous age, and delay were still significant predictors (see Table 1.2 in the online supplemental materials for statistics of all models).

Figure 5
Spearman Correlation Graphs of ResMem HR Versus Adult HR



Note. (A) Spearman correlation graph of ResMem HR versus adult HR on the 59 scene images. ResMem HR significantly predicted adult HR on the 59 scene images, meaning that ResMem HR is a good proxy of adults' memory of the scene images. (B) Spearman correlation graph of ResMem HR versus adult HR on the 60 animal images. ResMem HR did not predict adult HR on the 60 animal images, meaning that ResMem HR is not a good proxy of adults' memory of the animal images. ResMem HR = ResMem hit rates; adult HR = adult hit rates. See the online article for the color version of this figure.

** $p < .01$.

Figure 6*Spearman Correlation Graphs of ResMem HR Versus Children's Memory Accuracy*

Note. Top (A–C) Spearman correlation graphs of ResMem HR versus children's memory accuracy in each age group across delay conditions on the 20 scene images from the child memory data set. (A) ResMem did not predict the memory of 3-year-olds across delays. (B–C) ResMem significantly predicted the memory of 4- and 5-year-olds across delays. Bottom (D–F) Spearman correlation graphs of ResMem HR versus children's memory accuracy in each delay condition across age groups on the 20 scene images from the child memory data set. (D–E) ResMem did not predict children's memory across age after a 5-min or a 24-hr delay. (F) ResMem significantly predicted children's memory across age after a 1-week delay. ResMem HR = ResMem hit rates. See the online article for the color version of this figure.

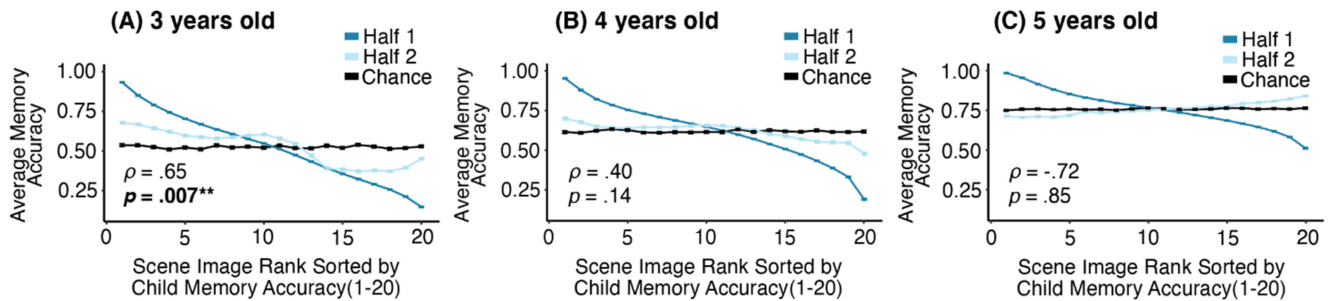
* $p < .05$. ** $p < .01$.

Possible Explanations for the Change in Children's Memory

One potential hypothesis to account for the fact that ResMem could predict the scene memory of 4- and 5-year-olds but not that of 3-year-olds is that 3-year-olds' memory might be noisier and less consistent across participants compared to older children's, given that their memory performance is generally worse (Saragosa-Harris et al., 2021). To test this hypothesis, we ran a split-half consistency analysis for each age group using the 20 scene images. Similar to the adult consistency analyses mentioned earlier, we split children within each age group into random halves and ran a Spearman correlation between the 20 data points from each half over 1,000 iterations, as shown in Figure 7. To our surprise, 3-year-olds were consistent in the scene images they remembered and forgot (averaged Spearman–Brown corrected $\rho = .65$, $p = .007$), whereas 4- (averaged Spearman–Brown corrected $\rho = .40$, $p = .14$) and 5-year-olds (averaged Spearman–Brown corrected $\rho = -.72$, $p = .85$) were not consistent in memory. We acknowledge that the nonsignificant results for the two older groups might be due to low power considering the low number of participants within each age group (forty-five 3-year-olds, forty-five 4-year-olds, and forty-seven 5-year-olds) as well as the low number of images (20; see Results SR5 in the online

supplemental materials for more results on this point). Nevertheless, the reliable memory pattern at the age of 3 ruled out the hypothesis that young children might have noisier, inconsistent memory patterns. Given that ResMem could not predict the scene memory of 3-year-olds, these results suggest that 3-year-olds might consistently use a common strategy that ResMem does not capture. We tested two candidate strategies used by 3-year-olds—a reliance on (a) highly familiar scenes (see Experiment SE1 in the online supplemental materials), or (b) bright/contrasting color (see Results SR8 in the online supplemental materials), and found neither predicted 3-year-old memory. See more discussion on potential alternate explanations in the General Discussion. We also ran split-half analyses for children's memory within each delay condition but found no consistency for any delay (all p s $> .05$; see Results SR6 in the online supplemental materials for statistics and interpretations).

An additional explanation is that children might begin to develop adult-like encoding strategies around the age of 4. Given that 3-year-olds had consistent memory patterns that could not be predicted by ResMem, whereas 4- and 5-year-olds had inconsistent memory patterns that could be predicted by ResMem, we conjectured that the inconsistent memory pattern at ages 4 and 5 might indicate a mix of encoding strategies. Specifically, some 4- and 5-year-olds might

Figure 7*Consistency Graphs for Children's Memory Accuracy Within Age*

Note. Within-age consistency graphs for children's memory accuracy of the 20 scene images from the child memory data set. Child participants within each age group ($N \approx 45$) were split into random halves in each of the 1,000 iterations. The 20 data points (image memory accuracy scores) from the first half of the children (Half 1: dark blue/gray line) were ranked from high to low accuracy, and the 20 data points from the second half (Half 2: light blue/gray line) were ranked using the image rank of Half 1. The two blue/gray lines represent the average accuracy at each rank over the 1,000 iterations. If the two halves of children are perfectly consistent in memory, the blue/gray lines should coincide. In contrast, if the two halves of participants are not consistent in memory, the light blue/gray line should coincide with the chance line in black, which represents a random shuffle of image ranks in Half 1. (A) 3-year-olds consistently remembered and forgot certain scene images. (B–C) 4- and 5-year-olds did not show consistent memory patterns of scene images within each age group. Error bars represent the standard error of the mean. See the online article for the color version of this figure.

** $p < .01$.

have developed adult-like strategies, while others remained using the same strategies as 3-year-olds. Because age is continuous and development takes place gradually, older 4-year-olds may show more similar memory patterns to adults than younger 4-year-olds. If this is the case, then we would expect to see ResMem better predicts the memory of older 4-year-olds than that of younger 4-year-olds. To test this hypothesis, we split 4-year-olds into a younger subgroup (age = [4, 4.5], $N = 18$, $M = 0.55$, $SD = 0.20$) and an older subgroup (age = [4.5, 5], $N = 27$, $M = 0.69$, $SD = 0.17$), and Spearman correlated ResMem HR with children's memory accuracy within each subgroup using the 20 scene images from the child data set. As expected, ResMem HR did not predict the memory of the younger 4-year-old subgroup ($p = .34$, $p = .14$), but significantly predicted that of the older subgroup ($p = .48$, $p = .03$; Figure 8). Although a Fisher's z test does not confirm a significant change in ResMem's predictability between the younger and older age subgroups (Fisher's $z = -0.51$, $p > .05$, two-tailed), the p increases from .34 for the younger 4-year-olds to .48 for the older 4-year-olds in correlation to ResMem, indicating an increase in consistency between ResMem predictions and child memory with age. This result could not be accounted for by overall improvement in memory performance (Results SR7 in the online supplemental materials). These results suggest that the emergence of adults' memory patterns in early childhood is likely accompanied by the development of adult-like visual encoding strategies.

Does Adults' Memory Predict Children's Memory?

Knowing that ResMem predicted children's memory in older age groups and after a long delay, we expected adults' memory to predict children's memory, because ResMem HR was a reliable proxy of adults' memory. We ran Spearman correlations between adult HR and children's memory accuracy within each age group and delay condition using the 20 scene images from the child data set. None of the correlations yielded significant results (all $ps > .05$, see Table 3 in the online supplemental materials for detailed statistics), highlighting

that ResMem outperformed adults' memory data in predicting children's memory. Adult HR corrected for false alarms also could not predict children's accuracy in any age or delay condition (see Table 4 in the online supplemental materials). A potential explanation is that ResMem may be a cleaner measure of adult memorability compared to adult behavior data (see more discussion on this point in the Discussion subsection *associative and item-based measures of memorability*). As a result, we opted to use ResMem predictions rather than adult memory data as a predictor in the logistic regression models for this study.

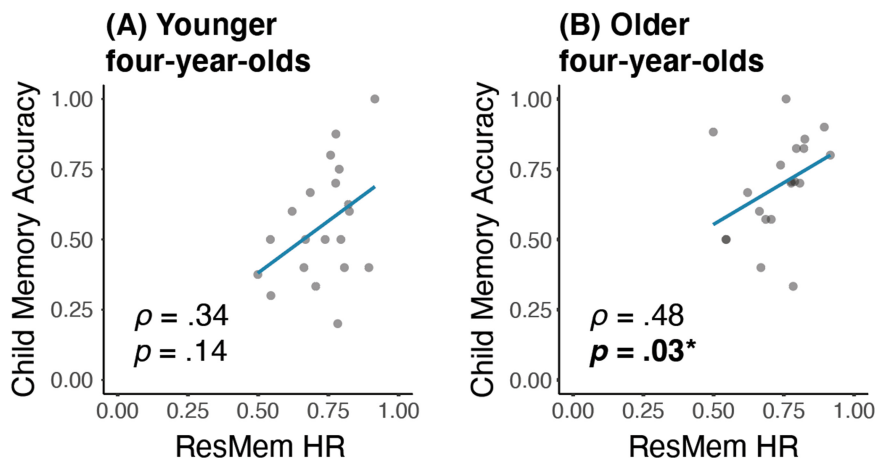
To see whether ResMem HR and adult HR predicted children's memory of animal images, we ran the same Spearman correlations within each age group and delay condition using the eight animal images. None of the correlations yielded significant results (all $ps > .05$), suggesting that neither ResMem nor adults' memory could predict children's memory of animal images. Such nonsignificant results might be because there were only eight data points in each correlation due to a low number of animal images. Children also had to remember a combination of the animal cues and the scene targets in order to perform the associative memory task. Hence, another potential explanation is that children's memory is less driven by the visual features of the cue items than the target items during this type of task. See more discussion on this point in General Discussion section.

Discussion

Here, we discover that children develop adult-like visual memory patterns by the age of 4, and ResMem predictions of adult memorability were able to predict children's memory for scene images, particularly after a 1-week delay. Unexpectedly, children at the age of 3 showed consistent memory patterns within their age group, whereas older children stop showing consistent memory patterns within each age group, reflecting a period of visual memory development of becoming more adult-like. We have also created a novel logistic regression model that could be used to predict children's scene memory accuracy using ResMem predictions, age, and delay.

Figure 8

Spearman Correlation Graphs of ResMem HR Versus Children's Memory Accuracy at the Age of 4



Note. Spearman correlation graphs of ResMem HR versus children's memory accuracy at the age of 4 on the 20 scene images from the child memory data set. The 4-year-olds ($N = 45$) were separated into a younger and an older group by month. (A) ResMem did not predict the memory of the younger 4-year-old group (4–4.5 years). (B) ResMem significantly predicted the memory of the older 4-year-old group (4.5–5 years). ResMem HR = ResMem hit rates. See the online article for the color version of this figure.

* $p < .05$.

Note that although children showed adult-like visual memory by the age of 4, it does not mean that children younger than 4 remembered little. We found that 3-year-olds did not show adult-like memory patterns but consistently remembered certain scenes more than others within their age group. In contrast, 4- and 5-year-olds exhibit adult-like memory patterns despite not showing consistent memory within each age group. There are a few possibilities for the consistency in 3-year-olds' memory. Although 3-year-olds do not seem to rely on image familiarity, esthetics, or color during encoding (see Experiment SE1 and Results SR8 in the online supplemental materials), it is possible that 3-year-olds rely on other image aspects, such as image complexity, or used some other simple response strategy for encoding. It is also possible that 3-year-olds are able to pick up on this image memorability that adults are sensitive to at encoding, but they fail to further encode memorable images due to a premature medial temporal lobe or they have difficulty retrieving memorable images (Bjorklund et al., 2008). Nonetheless, 3-year-olds reliably remember certain images, and 4- and 5-year-olds start to develop adult-like visual memory strategies. We discuss potential mechanisms underlying such development in the next subsection (*potential mechanisms for the development of adult-like memory patterns*).

There are a few potential explanations for why ResMem could predict children's memory after a long, 1-week delay but not after shorter delays. One possible explanation is that the information used by ResMem to predict children's memory better matches the information remaining in children's memory after a long delay. Specifically, given the decay of visual information in memory over time observed in adults (Lampinen et al., 2001; Potter, 2012), it is possible that richer semantic features of an image—those shown to be most predictive of memorability (Kramer et al., 2023)—is what persists in children's memory after a delay. Although the

memorability effect remains stable with different delays among adults (Isola et al., 2014; Lin et al., 2021), it is unclear whether it pertains across time among children. It is possible that there might have been a ceiling effect in children's memory after shorter delays, especially after the 5-min delay, and other factors such as differences in attention might have been the main sources of variance during shorter delays. This would have made it difficult for ResMem HR to predict children's memory performance.

Potential Mechanisms for the Development of Adult-Like Memory Patterns

We raised two competing hypotheses of the emergence of humans' sensitivity to the memorability effect. One hypothesis is that infants develop such sensitivity through neural maturation or visual information accumulation (the developmental account). Through development, infants' visual memory emerges from nonadult-like to adult-like between ages 3 and 4. Another hypothesis is that from very early on without any postnatal developments, infants possess sensitivity to the memorability effect and show adult-like visual memory patterns. Results from the current study support the developmental account. Children's memory patterns develop to be more adult-like and become predictive by ResMem, a proxy of adults' memory, by age 4. Another piece of evidence is that older 4-year-olds showed more adult-like memory patterns compared to younger 4-year-olds. This implies that the memory of older children was more adult-like compared to their younger counterparts both within and across age groups.

One possible mechanism behind the developmental account lies in neural changes in early childhood. We have noted that a decline in neurogenesis in the hippocampus improves its functional competence in holding long-term memory and leads to the fading of

infantile amnesia by the age of 4 (Akers et al., 2014; Bauer et al., 2011; Dennis et al., 2016; Schneider & Pressley, 2013). Results from this study suggest a temporal coincidence between the disappearance of infantile amnesia and the emergence of adult-like visual sensitivity. Previous research suggests a role of the medial temporal lobe and hippocampus in sensitivity to memorability by prioritizing memorable perceptual information during encoding (Bainbridge & Rissman, 2018; Bainbridge et al., 2017; Xie et al., 2020). We thus conjecture that the maturity of the hippocampus through a neurogenesis decline may have also supported adult-like visual sensitivity at around age 4. Additionally, a neural change in plasticity across development (Scher, 2008; Scher & Loparo, 2009) may also explain the emergence of adult-like visual memory. Children up to the age of eight can recognize faces from all races equally well, whereas older children and adults recognize own-race faces better than other-race faces (Goodman et al., 2007). With a decline in neural plasticity, the fluidity in visual perception declines, which shapes the perceptual expertise in one category of face stimuli. Similarly, young children may have high fluidity and are sensitive to diverse visual statistics during image encoding, whereas adults are sensitive to certain visual statistics that constitute the memorability effect, shaped by a decline in neural plasticity. It is possible that children have an established sense of certain visual statistics by age 4 and thus tend to remember the same images as adults.

Another potential mechanism underlying the developmental account may be the accumulation of cognitive experience. It is possible that by the age of 4, children have accumulated enough visual samples from the world to support adult-like visual processing or an adult-like model of visual memory. Around the age of 4, children start to attend kindergarten and engage in more diverse experiences, which may also result in children acquiring more adult-like visual experiences. In addition, the accumulation of language inputs leads to a rapid improvement in young children's language skills, which may allow children to extract high-level information and encode abstract representations of scene images by 4 years of age. Abstract, semantic information may be the type of features that predominantly drive consistencies in memorability (Kramer et al., 2023; Needell & Bainbridge, 2022).

Associative and Item-Based Measures of Memorability

In this study, we compared children's memory obtained from an associative memory task with adult memory and neural network predictions obtained from an item-based recognition memory task. Child participants encoded animal-scene image associations and were tested on scene recognition with animal cues after delays. The adult experiment did not have a separate memory test phase; instead, the continuous recognition paradigm showed repeated targets in a timed sequence and asked adult participants to indicate target recognition. The ResMem DNN was trained on data with different images from the same continuous recognition task paradigm. Despite these task differences, ResMem is remarkably able to predict children's memory for targets in an associative memory task. In fact, Xie et al. (2020) showed that in an associative word recall task, the memorability of the target word better predicted participants' recall memory performance than that of the associated word cue. Indeed, the logistic regression results in this study show that the memorability of the target scenes rather than animal cues predicts children's memory accuracy. This provides evidence that the memorability of the target items being recalled drives memory success in an associative memory task.

It remains a question why ResMem does not predict children's and adults' memories of animal images. One possible explanation is that ResMem was trained using natural images, whereas the animal images used in this study might have become unnatural after being cropped and placed in the middle of a black background. We used ResMem to predict the memorability of unnatural animal images because ResMem has been shown to significantly predict adult memorability of food images with black backgrounds and adult memorability of artwork that includes abstract pieces (Davis & Bainbridge, 2023). However, ResMem might not be well suited to discern the memorability differences of these unnatural animal images, as suggested by the smaller standard deviation of ResMem HR for animals ($SD = 0.07$), compared to scenes ($SD = 0.11$). This resonates with the ceiling effect in ResMem predictions for animals (min = 0.67, max = 0.97), suggesting ResMem's incapability in capturing memorability differences between unnatural animal images.

One surprising result that requires further exploration was that experimentally measured adults' memory was not predictive of children's memory. This is counterintuitive because ResMem could predict both adult and children's memories, making us initially expect adults' memory to also correlate with children's memory. However, adults may have used more elaborate strategies or relied on extra visual information that ResMem does not capture, making adults' memory a noisier model than ResMem in predicting children's memory. We hypothesize that the lack of predictability between adults' and children's memory data in the current study may have resulted from task-dependent effects present in the adult data. Indeed, recent evidence shows that ResMem may provide a clean measure of image memorability independent of task effects. While adults' memory relies on both image-intrinsic properties and image-extrinsic factors, ResMem predicts adult memorability without relying on image-extrinsic factors (Zhao et al., 2023). This suggests that ResMem prediction is less affected by various sources of noise, such as differences in task paradigms (measuring human memorability using a continuous recognition task vs. an associative memory task), compared to adults' memory data. Another piece of evidence that ResMem provides less noisy predictions of memorability compared to adults' online data is that ResMem can predict people's memory in the real world (an in-person art museum visit), but online memory test data cannot (Davis & Bainbridge, 2023). With this, we can view ResMem as a useful model of children's memorability and make predictions about children's memories using the ResMem DNN without having to collect additional adult data. To further explore the lack of correlation between children's and adults' memories, a future step will be to measure children's and adults' memories using the same associative memory task or continuous recognition task with a larger image set. This is the first step to tease apart whether the lack of correlation between children's and adults' memories was due to task or strategy differences between the age groups.

One limitation of this study is that only 20 scene images were used to compare children and adults' memories due to the constraint that the original children's study (Saragosa-Harris et al., 2021) only included 20 scene images. Although it is true that previous memorability research often used large-scale image data sets, prior work has shown that memorability measures are stable across different task parameters and image contexts (Bainbridge, 2020; Bylinskii et al., 2015; Xie et al., 2020). In addition, we found that the 20 scene images can be sufficient in testing the memorability effect among adults, such

that adults showed consistent memory when tested split-half consistency using only the 20 scene images or the 20 scene plus the 8 animal images (see Figure 2 in the online supplemental materials). Nonetheless, future studies could use a larger image set to examine and compare the memories of children and adults.

Another limitation is that adult HR was not corrected for false alarms because adult memorability scores were compared with noncorrected memorability data computed from ResMem and the child data set. ResMem predicts image hit rates that are not corrected for false alarms. In the child data set (Saragosa-Harris et al., 2021), children were asked to select one out of four scenes that were associated with an animal in each test trial. The four scenes consist of a correct scene, two lure scenes that have appeared in the encoding phase for other animals, and a novel foil scene that has never appeared before. Thus, a correct scene image in one test trial may be used as a lure in another test trial. In later trials, the correct and lure scenes may have appeared in previous trials, making it easier for children to identify the novel foil scene and avoid making false alarms. Because the scene images in different test trials were not independent of one another, we could not calculate traditional signal detection measures like FAR for the images because their memory could be impacted by the surrounding old (correct or lure) images. With these constraints, we utilized accuracy rather than FA-corrected HR rates to compare children's memory with memorability computed from adults' data and ResMem. Note that the results of these comparisons still replicate when tested using false alarm correction in a limited way without taking the previously mentioned constraints into consideration (see Results SR1 and SR2 and Figure 1 and Tables 2, 4, and 5 in the online supplemental materials for results on FAR and FA-corrected HR). Again, future studies could collect children and adults' image memory data using the same experimental paradigm that allows for the calculation of false-alarm-corrected hit rates.

Concluding Remarks

This article reveals the developmental trajectory of when children start to show consistencies in their visual memory. We provide evidence that children develop adult-like susceptibility to image memorability by the age of 4 through experience. Such susceptibility was the most salient when tested after a long, 1-week delay. Results suggest that the memorability effect emerges in early childhood through neural development or cognitive experience accumulation. These findings pave the way for exciting applications that could engage young children in more intuitive and memorable visual learning experiences. Teachers and caregivers could leverage the power of the ResMem DNN to select more memorable visual learning tools for children. Because young children at different ages seemed to use different visual memory strategies, education materials targeting young children could cater to different preferred visual strategies at different ages. We have outlined a few future directions of behavioral research to further investigate the development of visual memory. Remarkably, despite the individual differences that occur within and across age groups, by the age of 4, children already remember and forget the same images as adults.

References

- Akers, K. G., Martinez-Canabal, A., Restivo, L., Yiu, A. P., De Cristofaro, A., Hsiang, H.-L., Wheeler, A. L., Guskjolen, A., Niibori, Y., Shoji, H., Ohira, K., Richards, B. A., Miyakawa, T., Josselyn, S. A., & Frankland, P. W. (2014). Hippocampal neurogenesis regulates forgetting during adulthood and infancy. *Science*, 344(6184), 598–602. <https://doi.org/10.1126/science.1248903>
- Alberini, C. M., & Travaglia, A. (2017). Infantile amnesia: A critical period of learning to learn and remember. *Journal of Neuroscience*, 37(24), 5783–5795. <https://doi.org/10.1523/JNEUROSCI.0324-17.2017>
- Almog, G., Naeini, S. A., Hu, Y., Duerden, E., & Mohsenzadeh, Y. (2021). *Memoir dataset: Quantifying image memorability in adolescents*. <https://doi.org/10.31234/osf.io/kd29q>
- Bainbridge, W. A. (2020). The resiliency of image memorability: A predictor of memory separate from attention and priming. *Neuropsychologia*, 141, Article 107408. <https://doi.org/10.1016/j.neuropsychologia.2020.107408>
- Bainbridge, W. A., Dilks, D. D., & Oliva, A. (2017). Memorability: A stimulus-driven perceptual neural signature distinctive from memory. *NeuroImage*, 149, 141–152. <https://doi.org/10.1016/j.neuroimage.2017.01.063>
- Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General*, 142(4), 1323–1334. <https://doi.org/10.1037/a0033872>
- Bainbridge, W. A., & Oliva, A. (2015). A toolbox and sample object perception data for equalization of natural images. *Data in Brief*, 5, 846–851. <https://doi.org/10.1016/j.dib.2015.10.030>
- Bainbridge, W. A., & Rissman, J. (2018). Dissociating neural markers of stimulus memorability and subjective recognition during episodic retrieval. *Scientific Reports*, 8(1), Article 8679. <https://doi.org/10.1038/s41598-018-26467-5>
- Bauer, P. J., Larkina, M., & Deocampo, J. (2011). Early memory development. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 153–179). Wiley-Blackwell.
- Bjorklund, D. F., Dukes, C., & Brown, A. R. D. (2008). The development of memory strategies. In M. L. Courage & N. Cowan (Eds.), *The development of memory in infancy and childhood* (2nd ed., pp. 157–188). Psychology Press.
- Bylinskii, Z., Isola, P., Bainbridge, C., Torralba, A., & Oliva, A. (2015). Intrinsic and extrinsic effects on image memorability. *Vision Research*, 116, 165–178. <https://doi.org/10.1016/j.visres.2015.03.005>
- Cohendet, R., Yadati, K., Duong, N. Q. K., & Demarty, C.-H. (2018). Annotating, understanding, and predicting long-term video memorability. In K. Aizawa, M. Lew, & S. Satoh (Eds.), *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval* (pp. 178–186). ACM. <https://doi.org/10.1145/3206025.3206056>
- Davis, T. M., & Bainbridge, W. A. (2023). Memory for artwork is predictable. *Proceedings of the National Academy of Sciences*, 120(28), Article e2302389120. <https://doi.org/10.1073/pnas.2302389120>
- Dennis, C. V., Suh, L. S., Rodriguez, M. L., Kril, J. J., & Sutherland, G. T. (2016). Human adult neurogenesis across the ages: An immunohistochemical study. *Neuropathology and Applied Neurobiology*, 42(7), 621–638. <https://doi.org/10.1111/nan.12337>
- Farah, M. J., Rabinowitz, C., Quinn, G. E., & Liu, G. T. (2000). Early commitment of neural substrates for face recognition. *Cognitive Neuropsychology*, 17(1–3), 117–123. <https://doi.org/10.1080/026432900380526>
- Goetschalckx, L., Andonian, A., Oliva, A., & Isola, P. (2019, October 7–November 2). *GANalyze: Toward visual definitions of cognitive image properties* [Conference session]. Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, South Korea (pp. 5744–5753).
- Goodman, G. S., Sayfan, L., Lee, J. S., Sandhei, M., Walle-Olsen, A., Magnussen, S., Pezdek, K., & Arredondo, P. (2007). The development of memory for own- and other-race faces. *Journal of Experimental Child Psychology*, 98(4), 233–242. <https://doi.org/10.1016/j.jecp.2007.08.004>
- Hayne, H. (2004). Infant memory development: Implications for childhood amnesia. *Developmental Review*, 24(1), 33–73. <https://doi.org/10.1016/j.dr.2003.09.007>

- Hovhannisyan, M., Clarke, A., Geib, B. R., Cicchinelli, R., Monge, Z., Worth, T., Szymanski, A., Cabeza, R., & Davis, S. W. (2021). The visual and semantic features that predict object memory: Concept property norms for 1,000 object images. *Memory & Cognition*, 49(4), 712–731. <https://doi.org/10.3758/s13421-020-01130-5>
- Isola, P., Xiao, J., Parikh, D., Torralba, A., & Oliva, A. (2014). What makes a photograph memorable? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7), 1469–1482. <https://doi.org/10.1109/TPAMI.2013.200>
- Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? *CVPR, 2011*, 145–152. <https://doi.org/10.1109/CVPR.2011.5995721>
- Jaegle, A., Mehrpour, V., Mohsenzadeh, Y., Meyer, T., Oliva, A., & Rust, N. (2019). Population response magnitude variation in inferotemporal cortex predicts image memorability. *ELife*, 8, Article e47596. <https://doi.org/10.7554/eLife.47596>
- Josselyn, S. A., & Frankland, P. W. (2012). Infantile amnesia: A neurogenic hypothesis. *Learning & Memory*, 19(9), 423–433. <https://doi.org/10.1101/lm.021311.110>
- Kamps, F. S., Hendrix, C. L., Brennan, P. A., & Dilks, D. D. (2020). Connectivity at the origins of domain specificity in the cortical face and place networks. *Proceedings of the National Academy of Sciences*, 117(11), 6163–6169. <https://doi.org/10.1073/pnas.1911359117>
- Kessen, W., Salapatek, P., & Haith, M. (1972). The visual response of the human newborn to linear contour. *Journal of Experimental Child Psychology*, 13(1), 9–20. [https://doi.org/10.1016/0022-0965\(72\)90003-3](https://doi.org/10.1016/0022-0965(72)90003-3)
- Khaligh-Razavi, S. M., Bainbridge, W. A., Pantazis, D., & Oliva, A. (2016). From what we perceive to what we remember: Characterizing representational dynamics of visual memorability. *BioRxiv*, 049700. <https://doi.org/10.1101/049700>
- Khosla, A., Raju, A. S., Torralba, A., & Oliva, A. (2015, December 7–13). *Understanding and predicting image memorability at a large scale* [Conference session]. 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile (pp. 2390–2398). <https://doi.org/10.1109/ICCV.2015.275>
- Kramer, M. A., Hebart, M. N., Baker, C. I., & Bainbridge, W. A. (2023). The features underlying the memorability of objects. *Science Advances*, 9(17), eadd2981.
- Lampinen, J. M., Copeland, S. M., & Neuschatz, J. S. (2001). Recollections of things schematic: Room schemas revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(5), 1211–1222. <https://doi.org/10.1037/0278-7393.27.5.1211>
- Lin, Q., Yousif, S. R., Chun, M. M., & Scholl, B. J. (2021). Visual memorability in the absence of semantic content. *Cognition*, 212, Article 104714. <https://doi.org/10.1016/j.cognition.2021.104714>
- Lorsbach, T. C., & Reimer, J. F. (2005). Feature binding in children and young adults. *The Journal of Genetic Psychology*, 166(3), 313–328. <https://doi.org/10.3200/GNTP.166.3.313-328>
- Lukavský, J., & Děchtěrenko, F. (2017). Visual properties and memorising scenes: Effects of image-space sparseness and uniformity. *Attention, Perception, & Psychophysics*, 79(7), 2044–2054. <https://doi.org/10.3758/s13414-017-1375-9>
- Mohsenzadeh, Y., Mullin, C., Oliva, A., & Pantazis, D. (2019). The perceptual neural trace of memorable unseen scenes. *Scientific Reports*, 9(1), Article 6033. <https://doi.org/10.1038/s41598-019-42429-x>
- Needell, C. D., & Bainbridge, W. A. (2022). Embracing new techniques in deep learning for estimating image memorability. *Computational Brain & Behavior*, 5(2), 168–184. <https://doi.org/10.1007/s42113-022-00126-5>
- Ongchoco, J. D. K., Chun, M. M., & Bainbridge, W. A. (2023). What moves us? The intrinsic memorability of dance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 49(6), 889–899. <https://doi.org/10.1037/xlm0001168>
- Potter, M. C. (2012). Recognition and memory for briefly presented scenes. *Frontiers in Psychology*, 3, Article 32. <https://doi.org/10.3389/fpsyg.2012.00032>
- Saragosa-Harris, N. M., Cohen, A. O., Shen, X., Sardar, H., Alberini, C. M., & Hartley, C. A. (2021). Associative memory persistence in 3- to 5-year-olds. *Developmental Science*, 24(5), Article e13105. <https://doi.org/10.1111/desc.13105>
- Scher, M. S. (2008). Ontogeny of EEG-sleep from neonatal through infancy periods. *Sleep Medicine*, 9(6), 615–636. <https://doi.org/10.1016/j.sleep.2007.08.014>
- Scher, M. S., & Loparo, K. A. (2009). Neonatal EEG/sleep state analyses: A complex phenotype of developmental neural plasticity. *Developmental Neuroscience*, 31(4), 259–275. <https://doi.org/10.1159/000216537>
- Schneider, W., & Pressley, M. (2013). *Memory development between two and twenty*. Psychology Press.
- Travaglia, A., Bisaz, R., Sweet, E. S., Blitzer, R. D., & Alberini, C. M. (2016). Infantile amnesia reflects a developmental critical period for hippocampal learning. *Nature Neuroscience*, 19(9), 1225–1233. <https://doi.org/10.1038/nn.4348>
- Tzourio-Mazoyer, N., De Schonen, S., Crivello, F., Reutter, B., Aujard, Y., & Mazoyer, B. (2002). Neural correlates of woman face processing by 2-month-old infants. *NeuroImage*, 15(2), 454–461. <https://doi.org/10.1006/nimg.2001.0979>
- Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010, June 13–18). *SUN database: Large-scale scene recognition from abbey to zoo* [Conference session]. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, United States (pp. 3485–3492). <https://doi.org/10.1109/CVPR.2010.5539970>
- Xie, W., Bainbridge, W. A., Inati, S. K., Baker, C. I., & Zaghloul, K. A. (2020). Memorability of words in arbitrary verbal associations modulates memory retrieval in the anterior temporal lobe. *Nature Human Behaviour*, 4(9), 937–948. <https://doi.org/10.1038/s41562-020-0901-2>
- Zhao, C., Kim, J., Tang, T. H., Saito, J. M., & Fukuda, K. (2023). *Deep neural network decodes aspects of stimulus-intrinsic memorability inaccessible to humans*. PsyArXiv. <https://doi.org/10.31234/osf.io/urz5s>
- Zimmermann, T. D., & Meier, B. (2006). The rise and decline of prospective memory performance across the lifespan. *Quarterly Journal of Experimental Psychology*, 59(12), 2040–2046. <https://doi.org/10.1080/17470210600917835>

Received December 20, 2022

Revision received August 4, 2023

Accepted September 28, 2023 ■