Developmental Psychology

How Do Children Construct a Concept of Age?

Kosta Boskovic and David Barner

CITATION
Boskovic, K., & Barner, D. (2026). How do children construct a concept of age? *Developmental Psychology*. Advance online publication. https://dx.doi.org/10.1037/dev0002223

AMERICAN PSYCHOLOGICAL ASSOCIATION

Developmental Psychology

ISSN: 0012-1649

https://doi.org/10.1037/dev0002223

How Do Children Construct a Concept of Age?

Kosta Boskovic and David Barner
Department of Psychology, University of California, San Diego

Acquiring an adultlike understanding of age involves coordinating knowledge across several domains of abstract content (e.g., time, number, biology). In the present study, we explored children's early understanding of age and how it is informed by features including size, facial and bodily morphology, and numerical age. Across two preregistered experiments, we tested 215 three- to five-year-old children on their identification of which of two figures is older. Like previous studies, we found that children often confound age with size. However, we also found that this tendency was eliminated when children were provided with less extreme differences in size, they had access to information regarding numerical age, or when facial and bodily cues to age were more pronounced. We also found that, overall, children's age judgments were related to their mastery of number words, suggesting a role for numeracy in understanding age. These results suggest that young children possess a concept of age, which is differentiated from a concept of size and which draws on multiple converging cues for reasoning about age.

Public Significance Statement

This study suggests that children reason about age in a complex manner by 3 or 4 years old, demonstrating the ability to base age judgments on numerical ages and facial and bodily morphology rather than only on size. We also find that children's age judgment ability is related to their numerical knowledge, suggesting that acquiring a number system may help children understand age.

Keywords: age concepts, number words, number concepts, conceptual development, language development

Supplemental materials: https://doi.org/10.1037/dev0002223.supp

As we grow older, our bodies, minds, and behaviors change, making age an important social cue to predicting how others will act in the world. However, little is known about how children learn to reason about age or make predictions about its correlates. Whereas substantial research has documented how children make judgments of intelligence, social status, and social preferences on the basis of social categories like race and gender (e.g., Alton et al., 2025; Bian et al., 2017, 2018; Bigler et al., 2003; Diesendruck & HaLevi, 2006; Dukler & Liberman, 2022; Elenbaas & Killen, 2016; Mandalaywala et al., 2020; Martin, 1989; Olson et al., 2012; Shutts et al., 2013, 2016; Waxman, 2010), much less is known about children's understanding of age. Further, although children begin to differentiate various age groups in the first year of life (Brooks & Lewis, 1976), some studies have suggested that they may not acquire an adultlike understanding of chronological age until much later in development and that they may confuse it with correlated properties and processes, like size and growth (see Galper et al., 1981). In the present study, we investigated children's understanding of age and how they develop a concept of age, which is distinct from its physical correlates of size or growth.

Although many children are able to report their ages sometime after their second or third birthday (Woolley & Rhoads, 2019), previous studies argue that they do not understand that age denotes the amount of time someone has been alive until much later in development. For example, 4- to 6-year-old children often struggle to understand that aging is a continuous, linear process, and often believe that different people can age at different rates and that adults may cease to age

---

María Inés Susperreguy served as action editor.

Kosta Boskovic https://orcid.org/0009-0005-7733-859X

All data, code, materials, and preregistrations for this research are publicly available on the Open Science Framework at https://osf.io/j6sc8 (Boskovic & Barner, 2026). Part of these findings were presented as a poster at the 2025 Cognitive Science Society Meeting. The authors have no conflicts of interest to disclose.

The authors thank the children, families, schools, and museums who contributed to this work. The authors also thank Sara Alkhouli, Anna Claire Rogness, Marianna Thorne, Ellie Yeung, Elina Zamiri, and other members of the Language and Development Lab at the University of California, San Diego, for assistance in data collection. The authors are grateful for the feedback from Martin Zettersten on this project and earlier versions of this article.

Kosta Boskovic played a lead role in data curation, formal analysis, investigation, and writing–original draft and an equal role in conceptualization, methodology, and writing–review and editing. David Barner played a lead role in funding acquisition, resources, and supervision and an equal role in conceptualization, methodology, and writing–review and editing.

Correspondence concerning this article should be addressed to Kosta Boskovic, Department of Psychology, University of California, San Diego, La Jolla, CA 92093-0109, United States. Email: kboskovic@ucsd.edu

(Burdett & Barrett, 2016; Galper et al., 1981; Piaget, 1971). Children at this age also struggle to relate birth order to age and may know that they are older than their younger siblings but fail to recognize that they were therefore born earlier (Piaget, 1971). Finally, identifying the causal relationships between aging and its co-occurring activities and processes is also difficult for children. For example, many 3- to 4-year-old children attribute a causal role to birthday parties, believing that a child who is turning 3 will remain 2 if they do not have a birthday party (Klavir & Leiser, 2002; Woolley & Rhoads, 2019).

Of particular interest to the present study, some previous studies argue that young children reason about age solely on the basis of physical size. In one early study of this, Looft (1971) presented 3- to 9-year-old children with pairs of drawings of males in four different age groups (infancy, childhood, adolescence, and adulthood) and in two different sizes (3.5 or 5.5 inches tall). For each pair, children were asked which figure was older or if they were the same age. Looft found that 3-year-old children achieved just 40% accuracy and made a substantial number of size errors (i.e., selecting the younger and bigger figure as older). Children became progressively more accurate with age, and 9-year-olds made adultlike responses over 80% of the time. Interestingly, children between 3 and 6 accompanied their responses with spontaneous justifications such as “he’s bigger” or “he has to be big to be older,” further suggesting a belief that size determines age. The finding that young children often base age on size has been replicated when children are presented with photographs of people (Kratochwill & Goldman, 1973), with dolls of varying heights (Kuczaj & Lederberg, 1977), and in a non-Western culture, in Malaysia (Looft et al., 1972).

Consistent with the idea that early age judgments might be based on size, other work finds an early-developing understanding of physical growth among preschoolers. In one study, Rosengren et al. (1991) argued that 3- to 5-year-old children understand that animals, but not artifacts, grow. Children were shown a figure of an animal and told that it represented that animal as a “baby.” Then, children were asked to identify what the animal would look like as an “adult” by selecting between the same figure and a bigger version of the original figure. Similarly, children were shown a figure of an artifact and asked to identify whether the same artifact or a bigger version of that artifact represented how it would look after “a very long time.” Children selected the bigger figure more often for the animals than the artifacts, supporting the conclusion that preschoolers understand that animals, but not artifacts, grow. Further suggesting that children reason about growth early in life, other studies find that 3- to 5-year-olds are more likely to judge that animals and plants can grow relative to artifacts or nonliving natural kinds and that transformations in appearance are more likely to be the result of human intervention rather than natural processes such as growth for artifacts relative to living things (Backscheider et al., 1993; Fouquet et al., 2017; French et al., 2018; Herrmann et al., 2013; Inagaki & Hatano, 1996; Inagaki & Sugiyama, 1988; Jipson & Callanan, 2003; Margett & Witherington, 2011; Zhu & Fang, 2000).

While multiple previous studies argue that young children base age judgments on size—perhaps because they conceptualize aging as physical growth—there are also reasons to believe that past studies may underestimate children’s understanding. For example, past studies that tested children with an age judgment task presented unrealistically large size differences between characters, making size a highly salient property of the stimuli (e.g., 3- vs. 5.25-inch. tall; Kratochwill & Goldman, 1973; Kuczaj & Lederberg, 1977; Looft, 1971; Looft et al., 1972). Given this, it is possible that children in these studies inferred that, given its salience, size must be relevant to the questions being asked, leading them to use it as a basis for their judgments despite understanding that age is not defined by size. Alternatively, another possibility is that children made judgments on the basis of size because other available cues such as facial morphology were either not salient to them or were not sufficiently discriminable. For example, young children may not have adultlike sensitivity to cues such as hair styles, wrinkles, or body proportions (e.g., relative head and limb size). Given this possibility, children may have had little other basis for their judgments other than size, despite understanding that age is not defined by size.

One important nonphysical cue that children might use to evaluate which of two people is older is numerical age. For example, a child who understands that age corresponds to the amount of time a person has been alive and who understands the meanings of number words should be able to judge that a person who is 5 years old is older than a person who is 3 years old. However, among studies that have documented a size bias, only one provided children with numerical ages (Kuczaj & Lederberg, 1977). In this study, 4- to 7-year-old children first completed a relative age judgment task in which numerical ages were not provided and size differences between figures were manipulated. Then, children completed another relative age judgment task in which they were given the numerical ages of the dolls (e.g., “This one is twenty years old and this one is five years old”). Kuczaj and Lederberg found that most children who demonstrated some success in identifying the older figure on the basis of physical cues performed even better when these cues were supplemented with numerical age information. Meanwhile, most children who struggled to identify the older figure on the basis of physical cues did not improve when provided with numerical cues and continued to respond randomly or according to size. Kuczaj and Lederberg concluded that these children considered numerical age to be irrelevant and claimed that children initially interpret “older” and “younger” as meaning “taller” and “shorter.”

A limitation of the study by Kuczaj and Lederberg, however, is that it did not test whether children actually understood the numbers used in their experiment. Thus, it is unclear whether children exhibited a size bias because they truly deemed size to be more important than numerical age or because they did not yet understand the number words being used. The latter is a distinct possibility given past findings that many 4-year-olds struggle to judge the relative magnitudes denoted by number words, especially for numbers larger than 5 or 6 (Davidson et al., 2012; Le Corre, 2014). Although previous studies find that most 3-year-olds have memorized a count list, they also find that many children at this age have learned the meanings of only the first few words in their count list (i.e., up to 3 or 4; Wynn, 1990, 1992). As a result, though they are sometimes able to compare the magnitudes of small numbers, they struggle to do so for numbers greater than 3 or 4 (Condry & Spelke, 2008; Davidson et al., 2012; Le Corre, 2014). For example, when told that there are six fish in one container and 10 fish in another and asked which box contains more fish, many 4-year-olds respond randomly, despite the fact that they can accurately count and label sets up to 10. Even many 5-year-olds appear to not understand that numbers that occur later in the count list denote greater quantities (i.e., the later-greater principle).[1] Given this evidence, it is unclear what information children at this age extract from numerical age expressions involving numbers larger than 3 or 4.

Understanding how children reason about numerical age is important for two reasons. First, given the limitations of past studies, it is possible that preschool-aged children may not actually equate age with size and "older" with "taller," as previous studies have argued (Kratochwill & Goldman, 1973; Kuczaj & Lederberg, 1977; Looft, 1971; Looft et al., 1972). Instead, children may rely on multiple cues when reasoning about age, including not only size but also facial and bodily morphology and numerical age. When these other cues are not salient or interpretable, such as when children do not understand the number words, or when facial features differ only subtly between characters, children may resort to size as a heuristic for judging age given that the two are often correlated. Second, learning numerical expressions may help children discover that age denotes time spent alive. Once a child learns how number words encode magnitudes, they may notice that a person with a higher numerical age (e.g., 10 years old) is referred to as "older" than a person with a lower age (e.g., 5 years old), allowing them to base their relative age judgments on numerical ages when available rather than mere correlates of age such as size. Consistent with this possibility, as children learn the meanings of number words between 3 and 6 years old (Carey, 2004, 2009; Carey & Barner, 2019; Davidson et al., 2012; Fuson, 1988; Gelman & Gallistel, 1978; Le Corre & Carey, 2007; Sarnecka & Carey, 2008; Schaeffer et al., 1974; Wynn, 1990, 1992), they also increasingly reason about age in an adultlike manner, improving in making age judgments between pairs of figures, rank-ordering figures by age, and relating age to birth order (Burke, 1982; Kogan et al., 1961; Kratochwill & Goldman, 1973; Kuczaj & Lederberg, 1977; Looft, 1971; Piaget, 1971; Seefeldt et al., 1977; Taylor et al., 1982).

The present study investigated whether young children possess a concept of age that is differentiated from size and that also draws on cues like numerical age and physical morphology. We manipulated these various cues and investigated whether individual differences in children's understanding in related domains, such as their understanding of number words, predict their performance in reasoning about age. In Experiment 1, we asked 3- to 5-year-old children to complete an age judgment task, in which they judged who was older between pairs of images intended to resemble children 3, 5, 7, and 9 years of age.[2] Crucially, size differences between figures were fully discriminable (as determined in norming studies), but much smaller than in previous studies (e.g., Looft, 1971), to avoid creating the inference that size was the most important cue to consider in the task. Moreover, to investigate the importance of numerical knowledge, children in one condition were provided with the numerical ages of the figures presented in the age judgment task, while children in another condition were not. Relatedly, we tested children's comprehension of the numbers used in the age judgment task using the later-greater task (see Davidson et al., 2012; Le Corre, 2014). This allowed us to test whether children who do not base their age judgments on numerical age when provided with it do so because they deem numerical age to be irrelevant or because they do not understand the meanings of number words. Finally, we measured children's ability to temporally order two recent events in their lives in an autobiographical memory task. This task served as a proxy for children's ability to reason about temporal order, an ability that may be involved in representing chronological age (e.g., given that a mature age concept requires understanding that some people's existence precedes others in time).

In Experiment 2, children completed an age judgment task, in which a figure of a child was paired with a figure of an elderly person to test the influence of salient facial morphological differences on children's age judgments. In addition, children in one condition were asked to identify the older figure, whereas children in another condition were asked to identify the grown-up. This manipulation tested whether children who show a size bias might simply not know the meaning of the word "older" rather than having a conceptual misunderstanding of age (e.g., if children confuse the word "older" with "taller," as suggested by Kuczaj & Lederberg, 1977). We reasoned that if children represent age differences but do not understand the word "older," then they may successfully identify grown-ups on the basis of morphological features but nevertheless fail to judge adult figures as older than child figures. Finally, children completed the same later-greater task as in Experiment 1, to further probe the relation of numerical knowledge to an understanding of age.

Experiment 1

In Experiment 1, we asked whether young children possess a concept of age that is differentiated from their concept of size. To explore this, we asked whether children initially base age judgments on height, when size information is perceptible but less salient than in past studies, or are able to base their judgments on more reliable cues such as morphology and numerical age. Moreover, we tested whether changes in children's numerical knowledge and ability to order events from their own lives were related to the development of age judgments. Children first completed an age judgment task to test whether they based age judgments on size or on morphological features of figures. To test how numerical knowledge impacts children's age judgments, we provided the numerical ages of the figures in the relative age judgment task to only half of the children. We also tested all children's numerical knowledge in a later-greater task of magnitudes. Finally, we tested children's ability to order past memories in an autobiographical memory task to understand how this form of temporal reasoning might relate to their development of the age concept.

---

[1] Note that the acquisition of the later-greater principle is not the final stage in children's acquisition of number word meanings, as children only later generalize the successor function, understanding that for every number X, there exists a successor Y such that Y is exactly one greater than X (see Schneider, Pankonin, et al., 2021; Schneider, Sullivan, et al., 2021). This ability may be less related to reasoning about age than the later-greater principle, though, because a child who grasps the later-greater principle should be able to compare the magnitudes of all numerical ages that consist of numbers in their count list.

[2] We chose to ask children to judge the relative age of figures of two different people rather than asking them to identify a younger or older depiction of the same figure, another method used in in other studies (e.g., Rosengren et al., 1991), because children's ability to reason about age from a multiplicity of cues may precede their understanding of the chronological process of aging. That is, a child may reason about age based on a variety of cues such as numerical age and facial morphology, in addition to size, but may not yet understand that one develops from being a baby to a child to a grown-up over time. Such a child may thus succeed on an age judgment task that asks them to compare the ages of two different people at a certain point in time, but not on judging what an older or younger version of a person may look like. Using the former method is thus more likely to be sensitive to children's earliest understanding of age as possibly differentiated from size.

Method

Participants

A preregistration is available at https://osf.io/j6sc8 (Boskovic & Barner, 2026). Participants were 122 native, English-speaking children ($M_{\text{age}} = 4.53$ [3.04, 5.99]; $SD_{\text{age}} = 0.78$; 59 females, 63 males) who were able to count to 9 in a count-to-10 pretest. Participants were recruited through a local museum in San Diego, United States. Residents of San Diego are predominantly White (46.6%) and Hispanic or Latino (29.8%), and the median household income in the region is $102,285 (U.S. Census Bureau, 2024). Forty 3-year-olds, forty 4-year-olds, and forty-two 5-year-olds participated in the study. In each age group, 20 participants were assigned to the numerical age condition of the age judgment task, and 20 participants were assigned to the no numerical age condition. We also inadvertently tested two additional 5-year-olds in the no numerical age condition. In addition to these 122 participants, 19 children participated in the study but were excluded from analyses for the following reasons: failure to remember events in the autobiographical memory task ($n = 6$), failure to complete the study ($n = 4$), experimenter error ($n = 7$), and parental interference ($n = 2$). Our planned sample size of 120 children was determined with an a priori power analysis on G*Power Version 3.1.9.7 (Faul et al., 2009) and is adequate to achieve 80% power for detecting a medium effect, at a significance criterion of $\alpha = .05$ for a linear regression with five predictors (age, congruence of size and age cues, numerical age condition, later-greater task performance, and autobiographical memory task performance). We did not have strong priors regarding the likely size of effects, and therefore we powered our study to detect a medium effect so that we could identify factors that might have a substantial impact on behavior and therefore bear meaningfully on our theoretical questions. This study was approved by an institutional review board at the University of California, San Diego.

Materials and Procedure

All materials, data, and analysis code for the experiment are available at https://osf.io/j6sc8. All participants completed a count-to-10 task, an age judgment task, a later-greater task, and an autobiographical memory task in that order.

Count-to-10. This task served as a prettest to verify whether children were familiar with the number words in the numerical age condition of the age judgment task and the later-greater task. Given that the largest number word in these two tasks was "nine," children were asked to count to 10. Children who failed to count to at least nine did not proceed to complete the study.

Age Judgment Task.

Stimuli. Stimuli in the age judgment task consisted of 36 images intended to resemble children of four age groups: 3, 5, 7, and 9 years of age (e.g., see Figure 1).[3] Half of the figures were male, and half were female. To control for variables such as facial expression, body positioning, and posture, stimuli were created using ChatGPT's image generator DALL-E 3 (OpenAI, 2024). The following is an example of a prompt provided to ChatGPT to generate a stimulus: "Generate an image of a 3-year-old boy." The 36 stimuli were organized into 18 pairs, which included every possible pair of ages. Sex of the figures was controlled such that male-appearing images were only paired with male-appearing images and female-appearing images were only paired with female-appearing images. The figure meant to be older was presented on the left of the pair for half the trials and on the right of the pair for half the trials.

The images were presented in one of two heights in a computer slideshow presentation: 4.45 or 4.7 inch. These particular heights were selected based on norming studies, presented below, which ensured that all images were clearly discriminable in height. When modifying the height of each image, the original height-to-width ratio was retained. The height of the images in each pair was manipulated to create three types of trials (see Figure 1): (1) congruent trials, in which the image representing an older child was taller and the image representing a younger child was shorter; (2) incongruent trials, in which the image representing a younger child was taller and the image representing an older child was shorter; and (3) same-height trials, in which both images were presented in the taller height. Pairs of images were shown on a white background, with the base of the two images at an even level and the two images positioned equidistant to the center of the slide.

Norming Studies. Prior to conducting the main study, we conducted two norming studies. The first sought to ensure that the height difference used in the incongruent and congruent trials of the age judgment task was perceptible to children. Eleven 3-year-old children ($M_{\text{age}} = 3.55$ years, eight females, three males, all primary English speakers) were recruited from a local preschool and presented with 12 pairs of stimuli that were used in the primary experiment. For each pair, the child was asked, "Who is taller?" The first six pairs featured a large height difference that was clearly perceptible and served the purpose of identifying whether children understood the meaning of the word "tall." One child was excluded from the norming study because they did not answer all six of these questions correctly. The next six pairs featured the height difference used in the age judgment task. The 10 remaining children responded with a total accuracy of 95% across these six trials, demonstrating that the height difference is reliably perceptible to children in the target age range.

The second norming study sought to ensure that the ages of children depicted in the stimuli were discernible to participants with mature knowledge of age cues. To do so, we recruited 28 adult participants via the University of California San Diego's undergraduate research participant pool. Participants saw all 18 pairs of stimuli used in the age judgment task and were asked, "Who is older?" for each trial. The overall mean accuracy in selecting the child intended to represent an older age was 80%. Though performance was not quite at ceiling, this demonstrates that the stimuli reasonably represented the relative ages intended according to adult judgments (congruent trials = 83%; incongruent = 82%; same-height trials = 75%).

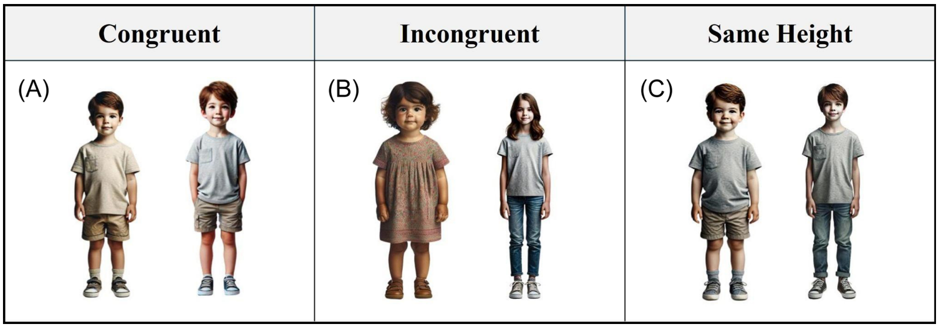
Procedure. Participants were randomly assigned to a numerical age condition or a no numerical age condition. Within each condition, participants were randomly assigned to one of four different versions, varying only in trial order. All participants were told the following at the beginning of the task: "I'm going to show you photographs of children and ask you who is older. Can you do that?" For each trial, participants in the numerical age condition were told

---

[3] We restricted the numerical ages of the figures to numbers within the child's count list because past studies find that preschoolers are unable to make "more" judgments for larger numbers while showing substantial variability in judgments for numbers from 1 to 10 (see Davidson et al., 2012; Le Corre, 2014). This allowed us to test whether children fail to use numerical ages because they deem them irrelevant to judging who is older (e.g., Kuczaj & Lederberg, 1977) or if instead they fail because they do not yet comprehend the numbers used.

Figure 1

Example Trials in the Age Judgment Task



Note. (A) A congruent trial of images representing a 3-year-old boy (left) and a 5-year-old boy (right), (B) an incongruent trial of images representing a 3-year-old girl (left) and a 9-year-old girl (right), and (C) a same-height trial of images representing a 3-year-old boy (left) and a 9-year-old boy (right). The images were generated via ChatGPT. See the online article for the color version of this figure.

the ages of each of the two children and asked, “Who is older?” Participants in the no numerical age condition were not told the ages and were only asked to indicate who was older for each trial.

Later-Greater Task. This task was adapted from Le Corre’s (2014) study to test children’s knowledge of the relative magnitudes denoted by the number words included as numerical ages in the age judgment task. The experimenter introduced the child to a stuffed toy:

> Now I’m going to introduce you to my friend, Mr. Monkey. Can you say hi to Mr. Monkey? Mr. Monkey is really hungry and loves to eat fish. He has two boxes with fish in them and he needs your help choosing which box of fish to eat from.

The experimenter then showed the child two open tin boxes that contained small plastic fish and said, “This box has one fish and this box has two fish. Mr. Monkey wants this box because he wants more fish!” Practice trials were then administered in the following order: one versus two, two versus four, and one versus two. Participants were told whether they were correct or incorrect on the practice trials. Children were then presented test trials in which they were told that fish had been placed in closed tin boxes, followed by a verbal description of how many were in each, ensuring that their choices were based on their understanding of the meanings of the number words. Children received each pair of number words that were included in the age judgment task once and were not told whether they were correct or incorrect during the test trials. The order of presentation was pseudo-randomized and counter-balanced. In both the practice trials and test trials, the order in which the larger number word was presented, and the side of the table on which it was placed was also pseudo-randomized. Halfway through the test trials, the children were administered an attention check. Children were told, “Look! Do you remember Mr. Monkey? Do you remember what Mr. Monkey likes to eat? What does he like to eat? That’s right! Mr. Monkey likes eating fish.” They were reminded that the stuffed animal was still hungry and needed their help to choose which box of fish to eat from a few more times.

Autobiographical Memory Task. This task was adapted from Pathman et al.’s (2013) study to test children’s ability to temporally order two past events from their own lives. As each child began the study, their caregiver filled out a questionnaire titled “Recent Events in Your Child’s Life” to provide the experimenter with events to ask the child about in the autobiographical memory task. Caregivers were instructed to provide three events that occurred 1 week ago and three events that occurred 1 month ago ranked by how memorable they judged them to be. Caregivers were asked to select events that were meaningful, memorable, and of interest to the child (e.g., a trip to the zoo) rather than routine events (e.g., eating breakfast). To begin the task, the experimenter asked the child, “Do you remember when you XXX?” referencing an event that happened either a week ago or a month ago, as it was described by the caregiver. Whether the child was first asked about the event a week ago or a month ago was counterbalanced. The experimenter then asked the child, “Can you tell me about it?” Further prompts from the experimenter were used to elicit recall if necessary, such as, “What else did you do?” or “Can you tell me more?” When appropriate, more specific prompts were used to test recall (e.g., asking a child, “What was your costume?” if the event was Halloween trick-or-treating). If the child produced two pieces of information about the event, it was coded as “remembered” and eligible to be used in the task. If the child failed to do so, the event was not considered to be remembered and was deemed ineligible. This process was then repeated for the next event provided by the parent for that time period if the child did not remember the first one. If the child successfully remembered the event, the experimenter moved onto asking about the first event from the other time period.

Once the child demonstrated recall of one event that occurred a week ago and one event that occurred a month ago, the experimenter prompted the child to indicate which event occurred more distantly and which event occurred more recently. To do so, the experimenter told the child, “So, you remember when you did *x* and you remember when you did *y*. One of these was a long time ago and one of these was a short time ago. Which one was a long time ago? Which one was a short time ago?” Half of the children were asked about the event “a long time ago” first, and half of the children were asked about the event “a short time ago” first.

Transparency and Openness

All materials, data, and analysis code for both Experiments 1 and 2 are available at https://osf.io/j6sc8 (Boskovic & Barner, 2026). Data were analyzed using R (Version 4.5.1; R Core Team, 2025) and the R packages lme4 (Bates et al., 2015), car (Fox & Weisberg, 2019), emmeans (Lenth, 2025), ggplot2 (Wickham, 2016), dplyr (Wickham et al., 2022), tidyr (Wickham et al., 2025), and tidyverse (Wickham et al., 2019). The design and analysis of both experiments were preregistered, and we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Results

The primary aim of this study was to understand if children continue to show a size bias when making age judgments when size differences are less extreme and when children are given access to numerical information about age. To test this question, we asked how performance on the age judgment task related to children's judgments on the later-greater and autobiographical memory tasks and also whether performance differed when children were presented with numerical age information.

Before conducting these preregistered analyses, we conducted exploratory analyses to describe children's performances on each of the individual tasks in the study. Model comparisons were performed using likelihood ratio tests, and the best fitting models were selected on the basis of a significant chi-square statistic and reduced Akaike information criterion value. Where the previous model to be compared differed in random effect structure, random effects were added to render the random effect structure identical to the augmented model for the likelihood ratio test, such that the only difference between models was the addition of the fixed effect (see Winter, 2013). In an effort to reduce Type I error (see Barr et al., 2013; Brauer & Curtin, 2018; Schielzeth & Forstmeier, 2009), we reported models that demonstrated singular fit as is rather than pruning random effects. All post hoc pairwise comparisons were adjusted for multiple comparisons using a Bonferroni correction.

Age Judgment Task

Overall, children performed above chance (0.5) on the age judgment task, $M = 0.69$, $SD = 0.21$, $t(121) = 10.03$, $p < .001$, suggesting that many young children are able to judge relative age when size differences are subtle, contrary to prior studies (e.g., Looft, 1971). Post hoc one-sample *t* tests found that 3-year-old children did not differ from chance either when numerical ages were provided, $M = 0.60$, $SD = 0.22$, $t(19) = 1.96$, $p = .064$,[4] or when they were not provided, $M = 0.54$, $SD = 0.17$, $t(19) = 1.02$, $p = .322$. By contrast, 4-year-olds differed from chance both when numerical ages were provided, $M = 0.79$, $SD = 0.18$, $t(19) = 7.35$, $p < .001$, and when ages were not provided, $M = 0.58$, $SD = 0.15$, $t(19) = 2.36$, $p = .029$. Moreover, 5-year-olds performed significantly above chance both when ages were provided, $M = 0.96$, $SD = 0.06$, $t(19) = 31.84$, $p < .001$, and when they were not, $M = 0.69$, $SD = 0.12$, $t(21) = 7.34$, $p < .001$. These data suggest that by around age 4, children are able to make age judgments at a rate better than chance (Figure 2).[5]

Later-Greater Task

Overall, children performed better than chance (0.5) on the later-greater task, $M = 0.80$, $SD = 0.27$, $t(121) = 12.48$, $p < .001$. Post hoc one-sample *t* tests found that 3-year-olds, $M = 0.63$, $SD = 0.26$, $t(39) = 3.25$, $p = .002$; 4-year-olds, $M = 0.82$, $SD = 0.27$, $t(39) = 7.52$, $p < .001$; and 5-year-olds, $M = 0.95$, $SD = 0.17$, $t(41) = 16.87$, $p < .001$, all performed significantly above chance. Exploratory analyses and figures investigating how children's performance differed across specific pairs of numbers can be found in the Supplemental Materials.

Autobiographical Memory Task

Overall, 61% of children responded correctly on the autobiographical memory task when asked which event was a long time ago and which event was a short time ago. A child's response on this task was coded as correct if they correctly identified both which event was a long time ago and which event was a short time ago. A child's response was coded as incorrect if they responded incorrectly for either one or both of the events. Only 55% of 3-year-olds and 50% of 4-year-olds provided correct responses on the task, whereas 76% of 5-year-olds did so, suggesting that there is notable improvement in children's ability to temporally order past memories between 3 and 5 years of age.

Primary Analyses

To test which factors are related to children's understanding of age, we constructed several generalized linear mixed models predicting children's accuracy on the age judgment task. The models described below are summarized in Supplemental Table 2. While the main text describes the model structure and comparisons, the table reports the full coefficient estimates and model fit statistics (Akaike information criterion and log likelihood) for each model. Our base model included only age (years, continuous, *z* scored) as a fixed effect, with participant and item as random intercepts. This model revealed that children's performance in the age judgment task improved with age ($\beta = 0.74$, $p < .001$).

Next, we analyzed the impact of size on children's age judgments. We added a term for congruence of the relative ages and sizes of the figures in the age judgment task (congruent/incongruent/same height) to our base model, as well as a by-participant slope for congruence. Relative to the simpler model with only age as a fixed factor, this augmented model significantly improved fit to the data, $\chi^2(2) = 25.79$, $p < .001$, suggesting that children were influenced by size when making age judgments, unlike adults, whose performance was similar across trial types in the norming study ($M_{\text{Congruent}} = 0.83$; $M_{\text{Incongruent}} = 0.82$; $M_{\text{Same Height}} = 0.75$). Post hoc pairwise comparisons showed that children were more accurate in the congruent condition than in the same height ($OR = 2.96$, $SE = 0.94$, $z = 3.43$, $p = .002$) and incongruent conditions ($OR = 8.73$, $SE = 3.18$, $z = 5.95$, $p < .001$). Moreover, they performed better in the same height condition than in the incongruent condition ($OR = 2.95$, $SE = 0.83$, $z = 3.85$, $p < .001$). Exploratory *t* tests comparing children's performance to chance for each pairing of numerical age condition and congruence level revealed that while they performed above chance for all other pairings of condition and congruence (*p*s < .05), children performed below chance when size and age cues were incongruent and numerical ages were not

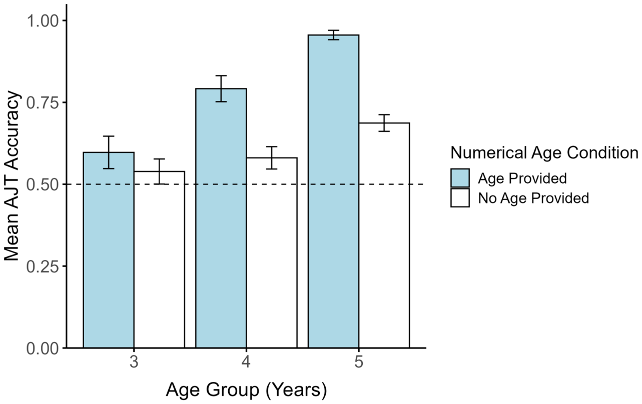
---

[4] Note that adult performance in the norming study (see the Method section) was particularly poor on the incongruent trial consisting of images representing 7- and 9-year-old boys ($M_{\text{acc}} = 0.5$) compared with the other trials ($M_{\text{acc}} = 0.67+$ on each trial). Given this, we repeated all analyses without the data from this trial and found that except in one instance, all results remained the same. When this trial was removed from the analysis, 3-year-olds performed above chance when numerical ages were provided, $M = 0.61$, $SD = 0.23$, $t(19) = 2.12$, $p = .047$.

[5] See Supplemental Figure 2 for a plot of children's performance on the age judgment task for each pair of numerical ages of the figures.

Figure 2

Children's Performance in the AJT in Each Numerical Age Condition Across Age Groups



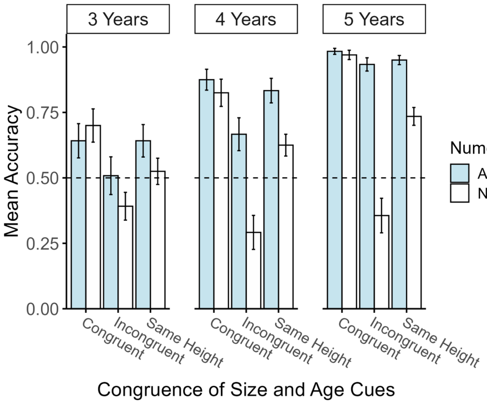
Note. The dotted line represents chance accuracy. Error bars represent standard error of the mean. AJT = age judgment task. See the online article for the color version of this figure.

provided, $M = 0.35$, $SD = 0.28$, $t(61) = -4.3$, $p < .001$, suggesting that many children identify a younger child as older if they are taller in the absence of numerical age cues (Figure 3).

The previous models demonstrate that children's age judgments were impacted by size. To assess whether numerical age influenced children's performance on the age judgment task, we added a term for numerical age condition (age provided/age not provided) to our earlier model that contained child age and congruence as fixed effects, as well as a by-item slope for condition. This augmented model significantly improved fit to the data, $\chi^2(1) = 19.13$, $p < .001$, suggesting that numerical age information influenced children's age judgments (Figure 2). Post hoc pairwise comparisons found that children were more accurate when they had access to numerical ages than when they did not ($OR = 4.09$, $SE = 1.11$, $z = 5.21$, $p < .001$). Next, we asked how children's ability to make use of numerical age cues changed with age (Figure 2). We found that an Age × Condition Interaction improved model fit, $\chi^2(2) = 23.93$, $p < .001$, relative to a simpler model containing just age and condition as fixed effects. Post hoc pairwise comparisons found that 3-year-olds' accuracy did not differ between conditions ($OR = 1.38$, $SE = 0.41$, $z = 1.08$, $p = .280$), whereas performance was better in the numerical age condition for both 4-year-olds ($OR = 3.72$, $SE = 1.15$, $z = 4.23$, $p < .001$) and 5-year-olds ($OR = 14.94$, $SE = 5.72$, $z = 7.06$, $p < .001$), suggesting that access to numerical age cues becomes useful for children in making age judgments around 4 years of age, allowing them to use numerical age as the basis of their judgments rather than a less reliable cue to age such as size. Finally, we asked whether having access to numerical ages reduced children's size bias by examining how the congruence of relative ages and size interacted with the availability of numerical age (Figure 3). We found that a Congruence × Condition interaction model improved model fit, $\chi^2(2) = 35.20$, $p < .001$, relative to a simpler model containing just congruence and condition as fixed effects. Post hoc pairwise comparisons found that while there was no significant difference in accuracy between the two conditions when the size and ages of the figures were congruent ($OR = 1.25$, $SE = 0.37$, $z = 0.75$, $p = .450$), children's age judgments were more accurate when they had access to numerical ages than when they did not when the cues were incongruent ($OR = 6.82$, $SE = 1.85$, $z = 7.07$, $p < .001$) and when the two figures were the same height ($OR = 3.53$, $SE = 0.99$, $z = 4.51$, $p < .001$). This suggests that although children can be misled by size in their age judgments, this effect is reduced when they are able to make use of numerical age cues.

Figure 3

Children's Performance in the Age Judgment Task for Each Age Group in Each Numerical Age Condition for Different Levels of Congruence of the Relative Ages and Sizes of Figures



Note. The dotted line represents chance accuracy. Error bars represent standard error of the mean. See the online article for the color version of this figure.

The results thus far indicate that access to numerical age information helped children make more accurate age judgments. Left open by these analyses is whether children who failed to benefit from numerical age information did so because of their still-developing understanding of age or instead because they had not yet learned the meanings of the number words involved in the task. To explore this, we constructed a model that included an interaction between numerical age condition and by-trial later-greater accuracy and found that the Condition × Later-Greater Accuracy interaction model improved model fit, $\chi^2(1) = 4.23$, $p = .040$, relative to a simpler model that included only condition and later-greater accuracy as fixed effects. Post hoc pairwise comparisons revealed that when children had access to the numerical age information, later-greater performance for the same pair of numbers as the numerical ages presented significantly predicted their accuracy on the age judgment task ($OR = 0.51$, $SE = 0.12$, $z = -2.98$, $p = .003$). By contrast, when children did not have access to the numerical age information, their accuracy was not predicted by this by-trial later-greater performance ($OR = 0.98$, $SE = 0.22$, $z = -0.10$, $p = .921$). This suggests that children who failed to use numerical information likely did so because they did not comprehend the numbers used and not because of a specific failure to reason temporally (though this knowledge may nevertheless also be absent, as we discuss below).

One limitation of this analysis of later-greater knowledge is that performance on specific numerical comparisons (e.g., 3 vs. 5) may not be a valid measure of the importance of numerical knowledge for all children in the study given that half of the children did not receive explicit numerical age cues.[6] For example, it is possible that children

---

[6] We found that adding a term for accuracy on the later-greater task on the trial with the same pair of numbers as the ages of the figures on that trial of the age judgment task were meant to represent did not significantly improve model fit compared with the earlier model with age, congruence, and numerical age condition as fixed effects, $\chi^2(1) = 1.52$, $p = .217$.

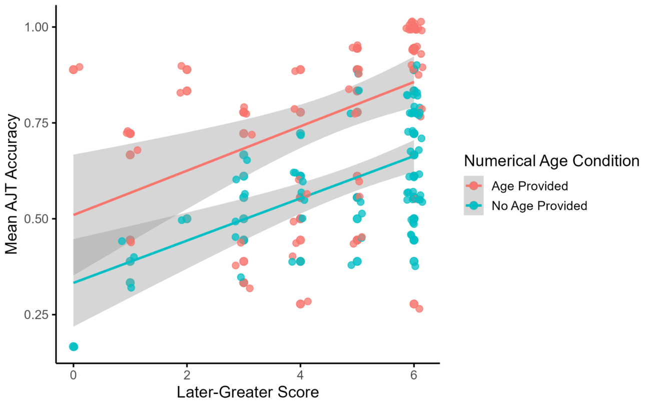
who struggle to compare two particular magnitudes, like 3 versus 5, are able to compare other numbers, and have a more general preliminary understanding that later numbers denote greater quantities. To test this possibility, we constructed an exploratory model that added a term for accuracy on the later-greater Task as a total score (rather than by trial) to the model with age, congruence, and numerical age condition as fixed effects. This model did significantly improve fit to the data, $\chi^2(1) = 11.10$, $p < .001$, suggesting that, generally, more advanced numerical knowledge may lead children to a more accurate concept of age and ability to make relative age judgments (Figure 4).

Thus far, we have found that size information, numerical knowledge, and access to specific numerical age information all impact children's age judgments beginning as early as age 3. Left open by the analyses presented thus far, however, is whether children at specific ages were able to make reliable age judgments when both size and numerical age information were completely absent (Figure 5). Post hoc one-sample t tests found that 3-year-old children did not differ from chance on these trials, $M = 0.53$, $SD = 0.22$, $t(19) = 0.5$, $p = .62$. By contrast, 4-year-olds, $M = 0.63$, $SD = 0.19$, $t(19) = 3$, $p = .007$, and 5-year-olds, $M = 0.73$, $SD = 0.16$, $t(19) = 6.89$, $p < .001$, performed above chance. This suggests that by age 4, children have not only acquired the ability to make use of numerical age cues to make age judgments but have also learned to infer age from facial and bodily morphological cues. Further, the mean accuracy of 5-year-olds on these same-height trials (0.73) was similar to that of adults in the norming study (0.75), suggesting children may approach adultlike competence in inferring age from morphological cues by 5 years of age.

Our final analyses focused on the relation between children's ability to temporally order recent memories and understanding of age. We found that adding a term for accuracy on the autobiographical memory task to the earlier model containing age, congruence, numerical age condition, and by-trial accuracy on the later-greater task did not significantly improve model fit, $\chi^2(1) = 0.44$, $p = .506$, suggesting that children's developing understanding of age may not be related to their ability to temporally order past memories.

Figure 4

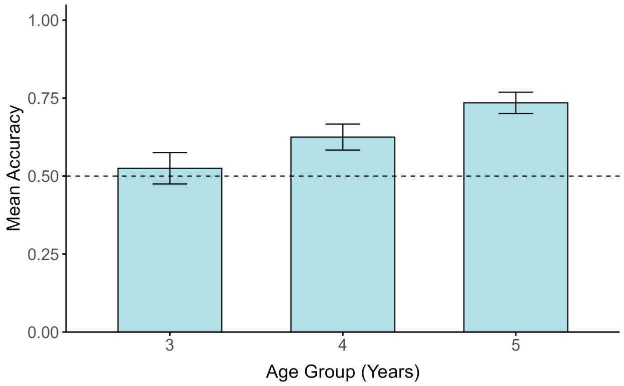
Children's Performance on the AJT as a Function of the Number of Trials They Responded Correctly on the Later-Greater Task (Split by Numerical Age Condition)



Note. AJT = age judgment task. See the online article for the color version of this figure.

Figure 5

Children's Performance on the Age Judgment Task When Numerical Age Cues Are Not Provided on Trials in Which the Two Figures Are the Same Height



Note. The dotted line represents chance accuracy. Error bars represent standard error of the mean. See the online article for the color version of this figure.

Discussion

Experiment 1 found three main results. First, we found that children judged relative age above chance by 4 years old when size differences were subtle or completely absent. This suggests that young children's poor age judgment ability demonstrated in previous studies (e.g., Looft, 1971) may have been partially an artifact of the large size differences employed (which may have suggested to children that size was the relevant dimension to base their responses on). Although children performed above chance in their age judgments in this study, we did find an effect of size/age congruence as in previous studies, such that children's age judgments were most accurate when age was congruent with size (i.e., the older figure was also taller), and least accurate when the cues were incongruent (i.e., the younger figure was taller). Finally, we found that children's accuracy was influenced by both their understanding of number words and the explicit availability of numerical ages. By age 4, children's age judgments were more accurate if they were told the numerical ages of the figures. Further, overall performance on the later-greater task, a test of number knowledge, predicted children's performance on the age judgment task. In addition, children's performance on a given later-greater number pair (e.g., 3 vs. 5) specifically predicted their age judgments when those same numbers were provided as the figures' ages. These results suggest that while children may resort to size in making age judgments when they lack other cues to age, once they have access to numerical age information, they make more accurate relative age judgments that focus less on physical size.

Overall, Experiment 1 found that 4- and 5-year-old children often make accurate age judgments, whether using numerical age cues, or facial and bodily morphology, and can do this even when size information is completely absent. However, many children (especially 3-year-olds) continued to show substantial sensitivity to size information, unlike adults. One interpretation of this finding is that very young children do not initially differentiate between age and size, as argued by previous studies (e.g., Looft, 1971). An alternative, however, is that even 3-year-olds do not conceptualize age solely in terms of physical size but may instead lack the ability to extract perceptual cues to age that are more easily detected by older children. For example, in Experiment 1, some children may have responded according to size because they could not yet differentiate age on the basis of facial or bodily morphology and did not yet possess the number knowledge required to make use of numerical age cues, when available. We explored this question in Experiment 2. Another factor that may explain errors in age judgment is that very young children may not understand the meaning of the word "older." Children may not know what "older" means or may actually view "older" as a synonym for "bigger." In either case, children may not understand the test question and resort to responding based on physical size as a result, despite possessing some understanding of age, which has yet to be mapped to the word "older." We also tested this question in Experiment 2.

Experiment 2

In Experiment 2, 3- and 4-year-old children again completed a count-to-10 pretest, an age judgment task and a later-greater task. However, the age judgment task was modified in several ways. To understand whether young children truly equate age with size or simply could not differentiate age on the basis of the morphological cues we provided in Experiment 1, in Experiment 2 we presented children with pairs of images contrasting a child with an elderly person, providing them with more salient morphological cues to age. Numerical ages were no longer provided in any condition because the numbers needed to label the ages of elderly people are too large to be known by 3- and 4-year-old children. Instead, we included the later-greater task as a measure of whether general numerical knowledge might play a role in understanding age. Finally, to directly probe whether children's understanding of the word "older" might explain their difficulty in the age judgment task, children were randomly assigned to one of two conditions. In the first, they were asked to identify the older figure, and in the second, they were asked to identify the "grown-up," a term children use already at 2 years old (Edwards & Lewis, 1979).

If children do not initially differentiate age from size, they should continue to exhibit a size bias in their age judgments despite the presence of more salient morphological cues to age. Moreover, if children exhibit a delay in mapping their understanding of age to the word "older," then those who are asked to identify the grown-up should perform better than children who are asked to identify who is older.

Method

Participants

A preregistration is available at https://osf.io/j6sc8. Participants were 93 native, English-speaking children ($M_{\text{age}} = 4.09$ [3.03, 4.99]; $SD_{\text{age}} = 0.56$; 48 females, 45 males) who were able to count to 9 in the count-to-10 pretest. Children were recruited through a local museum in San Diego, United States. Residents of San Diego are predominantly White (46.6%) and Hispanic or Latino (29.8%), and the median household income in the region is $102,285 (U.S. Census Bureau, 2024). Forty-six 3-year-olds and forty-seven 4-year-olds participated in the study. In each age group, 23 participants were assigned to the older condition of the age judgment task, and 23 participants were assigned to the grown-up condition. We also inadvertently tested one additional 4-year-old in the older condition. In addition to these 93 participants, three children participated in the study but were excluded from analyses for the following reasons: failure to complete the study ($n = 1$) and parental interference ($n = 2$). Our planned sample size of 92 children was determined with an a priori power analysis on G*Power Version 3.1.9.7 (Faul et al., 2009) and is adequate to achieve 80% power for detecting a medium effect, at a significance criterion of $\alpha = .05$ for a linear regression with four predictors (age, congruence of size and age cues, lexical item condition, and later-greater task performance).

Materials and Procedure

All materials, data, and analysis code for the experiment are available at https://osf.io/j6sc8/. All participants completed a count-to-10 task, an age judgment task, and a later-greater task in that order.

Count-to-10. This task served as a pretest that children were familiar with the number words in the later-greater task. Given that the largest number word in these two tasks was "nine," children were asked to count to 10. Children who did not count to at least nine did not proceed to complete the study.

Age Judgment Task.

Stimuli. Stimuli in the age judgment task consisted of 36 images, 18 of which were intended to resemble children in middle childhood and 18 of which were intended to resemble elderly people. Half of the figures were male, and half were female. Stimuli were created using ChatGPT's image generator DALL-E 3 (OpenAI, 2024). The 36 stimuli were organized into 18 pairs. Each pair consisted of one child figure and one elderly figure. Sex of the figures was controlled such that male images were only paired with male images and female images were only paired with female images. The figure meant to be older was presented on the left of the pair for half the trials and on the right of the pair for half the trials.

As in Experiment 1, the images were presented in one of two heights in a computer slideshow presentation: 4.45 or 4.7 in. with the original height-to-width ratio retained. The height of the images in each pair was manipulated to create three types of trials (see Figure 6): (1) congruent trials, in which the image representing an elderly person was taller and the image representing a child was shorter; (2) incongruent trials, in which the image representing a child was taller and the image representing an elderly person was shorter; and (3) same-height trials, in which both images were presented in the taller height.

Procedure. Participants were randomly assigned to an older or grown-up condition. Within each condition, participants were randomly assigned to one of four different versions, varying only in trial order. All participants were told the following at the beginning of the task: "I'm going to show you photographs of people and ask you who is older/a grown up. Can you do that?" For each trial, participants in the older condition were asked "Who is older?" and participants in the grown-up condition were asked "Who is a grown up?"

Later-Greater Task. The procedure of the later-greater task was identical to that in Experiment 1.

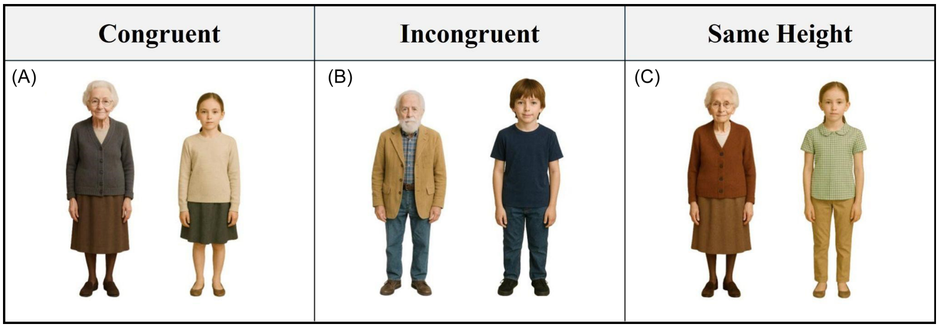
Results

Age Judgment Task

Overall, children performed above chance (0.5) on the age judgment task, $M = 0.69$, $SD = 0.32$, $t(92) = 5.69$, $p < .001$. Post hoc

Figure 6

Example Trials in the Age Judgment Task



Note. (A) A congruent trial of images representing an elderly woman (left) and a female child (right), (B) an incongruent trial of images representing an elderly man (left) and a male child (right), and (C) a same-height trial of images representing an elderly woman (left) and a female child (right). The images were generated via ChatGPT. See the online article for the color version of this figure.

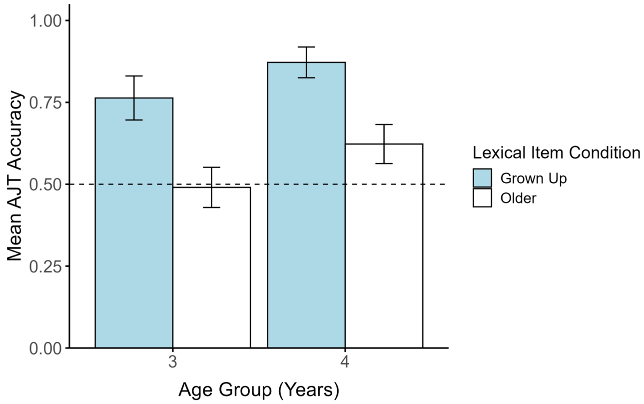
one-sample *t* tests found that 3-year-old children performed above chance in the grown-up condition, $M = 0.76$, $SD = 0.32$, $t(22) = 3.92$, $p < .001$, but did not differ from chance in the older condition, $M = 0.49$, $SD = 0.29$, $t(22) = -0.16$, $p = .88$. Likewise, 4-year-olds also performed above chance in the grown-up condition, $M = 0.87$, $SD = 0.23$, $t(22) = 7.93$, $p < .001$, but did not differ from chance in the older condition, $M = 0.62$, $SD = 0.29$, $t(23) = 2.06$, $p = .05$.[7] These data suggest that children represent differences between age groups already at 3 years old but may take longer to learn the meaning of "older" (Figure 7).

Later-Greater Task

Children performed better than expected by chance (0.5) on the later-greater task, $M = 0.73$, $SD = 0.24$, $t(92) = 9.46$, $p < .001$. Post hoc one-sample *t* tests found that 3-year-olds, $M = 0.64$, $SD = 0.26$, $t(45) = 3.79$, $p < .001$, and 4-year-olds, $M = 0.82$, $SD = 0.17$, $t(46) = 12.39$, $p < .001$, performed significantly above chance. These findings replicate those of Experiment 1. Exploratory analyses investigating how children's performance differed across pairs of numbers can be found in the Supplemental Materials.

Figure 7

Children's Performance in the AJT in Each Lexical Item Condition Across Age Groups



Note. The dotted line represents chance accuracy. Error bars represent standard error of the mean. AJT = age judgment task. See the online article for the color version of this figure.

Primary Analyses

To test which factors influence children's understanding of age, we constructed several generalized linear mixed models predicting children's accuracy on the age judgment task using the lme4 (Bates et al., 2015) R package. The models described below are summarized in Supplemental Table 4. While the main text describes the model structure and comparisons, the table reports the full coefficient estimates and model fit statistics (Akaike information criterion and log likelihood) for each model. Our base model included only age (years, continuous, *z* scored) as a fixed effect, with participant and item as random intercepts. This model revealed that children's performance in the age judgment task improved with age ($\beta = 0.66$, $p = .04$).

Next, we analyzed the impact of size on children's age judgments. We added a term for congruence of the relative ages and sizes of the figures in the age judgment task (congruent/incongruent/same height) to our base model, as well as a by-participant slope for congruence. Relative to the simpler model with only age as a fixed factor, this augmented model did not improve fit to the data, $\chi^2(2) = 0.57$, $p = .75$, suggesting that children were not influenced by size when making age judgments when morphological cues were salient (see Figure 8).[8]

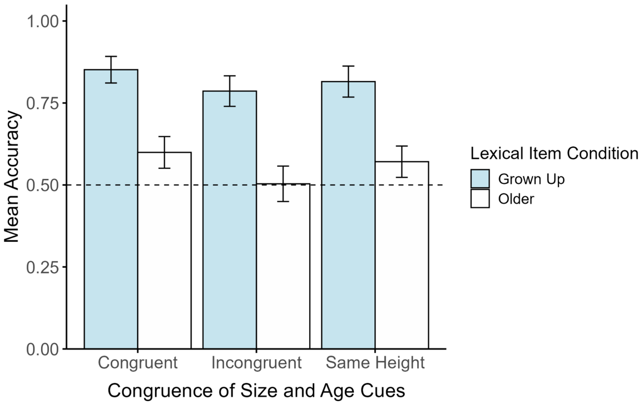
---

[7] Four-year-old children performed above chance on the age judgment task in Experiment 1, but not in Experiment 2 when asked to identify the older figure, even though their mean accuracies were highly similar across the two studies (Experiment 1: $M = 0.59$; Experiment 2: $M = 0.62$). Thus, the difference in statistical significance between experiments is unlikely to reflect a meaningful difference in accuracy. Instead, it is more likely driven by greater variability in children's responses in Experiment 2 ($SD = 0.29$) compared with Experiment 1 ($SD = 0.15$).

[8] A model without a by-participant slope for congruence constructed during a check of robustness of effects in which random effects were pruned to reduce singular fit revealed a significant effect of congruence, such that performance was worse on incongruent trials compared with congruent trials ($\beta = -0.68$, $p = .003$). This effect is not interpreted in the article due to the increased probability of Type I error when random slopes are removed (see Barr et al., 2013; Brauer & Curtin, 2018; Schielzeth & Forstmeier, 2009).

Figure 8

Children's Performance in the Age Judgment Task in Each Lexical Item Condition for Different Levels of Congruence of the Relative Ages and Sizes of Figures



Note. The dotted line represents chance accuracy. Error bars represent standard error of the mean. See the online article for the color version of this figure.

To assess whether the lexical item in the test question influenced children's performance on the age judgment task, we added a term for condition (older/grown-up) to our earlier model, which contained child age and congruence as fixed effects, as well as a by-item slope for condition. This augmented model significantly improved fit to the data, $\chi^2(1) = 20.9$, $p < .001$, suggesting that the lexical item influenced children's age judgments (Figure 7). Post hoc pairwise comparisons found that children were more accurate when they were asked to identify the grown-up rather than who is older ($OR = 16.6$, $SE = 10.7$, $z = 4.4$, $p < .001$). Next, we asked whether the influence of size on children's age judgments differed between lexical item conditions. We found that a Congruence × Condition interaction did not improve model fit, $\chi^2(2) = 0.74$, $p = .69$, relative to a simpler model containing just congruence and condition as fixed effects, suggesting that there was no difference in the influence of size between conditions.

Finally, we assessed whether children's numerical knowledge was related to their ability to make age judgments given that a positive relationship was found in Experiment 1. To do so, we added a term for later-greater task accuracy (sum score) to our earlier model, which contained age, congruence, and condition as fixed effects. This augmented model significantly improved fit to the data, $\chi^2(1) = 8.84$, $p = .003$, suggesting that more advanced numerical knowledge is related to more accurate age judgments (Figure 9). We then asked whether the influence of numerical knowledge differed between lexical item conditions. We found that a Congruence × Later-Greater Task Accuracy model did not improve model fit compared with a model with just congruence and later-greater task accuracy as fixed terms, $\chi^2(1) = 0.78$, $p = .38$, suggesting that this is not the case.

Discussion

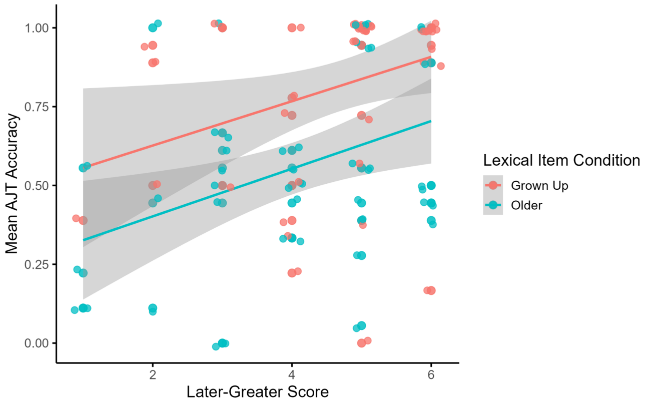
Experiment 2 found four main results. First, 3-year-old children were able to make accurate age judgments when provided with salient morphological cues to age and asked to identify who is a grown-up. Second, unlike in Experiment 1, children did not exhibit a size bias in their age judgments when provided with salient morphological cues. This suggests that children may not equate age with size but that they only use size as a basis for age judgments when they lack other salient cues to age, as in Experiment 1, or when size is made artificially salient, as in prior studies (e.g., Looft, 1971). Third, children performed better in the grown-up condition than the older condition. This suggests that children learn the word older later than grown-up, though it is unlikely that this is because children initially interpret older as a synonym for bigger, as there was no significant interaction between lexical item and congruence of size and age cues. This delayed acquisition of older may also explain children's poor performance and frequent size responses in previous studies (e.g., Looft, 1971), as children may not have understood the test question. Finally, children's ability to make age judgments was predicted by their numerical knowledge on the later-greater task, providing further evidence that acquiring a number system may play a role in children's development of the concept of age and that this knowledge informs age judgments even when numbers are not explicitly mentioned.

General Discussion

In two experiments, we investigated preschoolers' understanding of age and found evidence that children as young as 3 years old use multiple distinct factors to make age judgments and are not limited to conceptualizing age in terms of physical size. Three main results support this conclusion. First, we found that although children exhibited a size bias in Experiment 1 when judging the relative ages of children, in Experiment 2, we found they did not have such a bias when comparing children to elderly adults. This suggests that when facial and bodily cues to age are sufficiently strong, children as young as 3 prefer to use these cues over size to support age judgments. Second, children's age judgments in both Experiments 1 and 2 were predicted by their ability to order and compare number words, and this knowledge of number words had the strongest effect when children were explicitly provided with numerical ages. This suggests that acquiring a number system may play a key role in

Figure 9

Children's Performance on the AJT as a Function of the Number of Trials They Responded Correctly on the Later-Greater Task (Split by Lexical Item Condition)



Note. AJT = age judgment task. See the online article for the color version of this figure.

children's development of the concept of age, perhaps by leading them to conceptualize age in terms of relative magnitudes defined over temporal units, like months and years (e.g., "three years old" vs. "four years old"). Finally, Experiment 2 found that although many 3- and 4-year-old children failed to make age judgments when asked who was older, children at these ages performed better when asked to identify the grown-up, suggesting that some errors reported in past studies may be due to children's still-developing comprehension of the word "older" rather than to their understanding of age, per se. Thus, overall, we found that children use a constellation of cues including facial/bodily morphology, numerical age, and sometimes size to infer the relative ages of people and that this ability improves over development as children acquire numerical knowledge. These findings suggest a richer conceptualization of age than reported in previous studies that draws on multiple sources of information from early in development rather than just size alone.

The finding that children did not exhibit a size bias in their age judgments in Experiment 2 contrasts with Experiment 1 and previous studies that reported a size bias (Kratochwill & Goldman, 1973; Kuczaj & Lederberg, 1977; Looft, 1971; Looft et al., 1972). Two factors may explain this difference. First, our results are compatible with the hypothesis that children use size to judge age when they lack access to other salient cues. Children likely recognize the strong correlation between size and age and frequently use size as a proxy for age. For example, in Experiment 1, children who responded based on physical size may have done so because the age differences between the children in each pair of stimuli were too small (i.e., between 3 and 9 years of age), resulting in smaller differences in facial or bodily cues to age that are perhaps not discriminable to young children. Second, as noted in the introduction, past studies may have found a size bias in part because of the extreme differences in size that were presented to children. This large and highly salient difference may have created a pragmatic pressure, in which children inferred that because size differences were so large, they must be relevant to the judgments at hand. This might explain why a size bias was found even in studies that presented children with more extreme age differences (e.g., when figures representing an infant and an elderly person were paired; see Looft, 1971).

While our study demonstrates that young children possess a concept of age that is differentiated from a concept of size, our findings remain compatible with different degrees of understanding. In particular, although accurate performance on the age judgment task is compatible with an understanding of age as time spent alive, it is also consistent with more limited knowledge states. For example, recall that in Experiment 2 we found that 3-year-olds successfully identified the grown-up but failed when asked to identify who was older. One possibility is that their interpretation of grown-up reflects an understanding that people who look older have been alive longer and that acquisition of the word "older" is delayed relative to this early emerging understanding of age as time spent alive. However, another possibility is that 3-year-olds categorize people with gray hair, wrinkles, and so forth as grown-ups, without understanding that they have been alive longer than younger people, explaining their worse performance on judging who is older. Similarly, the finding that children can use numerical age information to make accurate judgments by 4 years of age is also compatible with different levels of understanding. Although children may understand that numerical age corresponds to time spent alive, it is also possible that they employ a simpler heuristic that the person labeled with the larger number is referred to as older, without understanding that this relates to time.

Relatedly, another open question is how children use the various cues that are available to them to acquire such a chronological understanding of age. Our work and evidence from past studies (e.g., Looft, 1971) together show that children recruit several cues to judge age, including facial/bodily morphology, size, and numerical age. Moreover, in Experiment 2 we found that prior to making adultlike judgments of who is older, children categorize individuals into different age groups, like child, grown-up, and so forth (see French et al., 2018; Rosengren et al., 1991, for further evidence that young children have formed age group categories mapped onto labels). One possibility, consistent with constructivist accounts of conceptual change (e.g., Carey, 1985), is that children begin by categorizing people into different groups and that these groups serve as placeholder structures, which lay the terrain for building a concept of age. For example, 3-year-olds may form categories like baby, child, and grown-up by associating labels with different groups of people but may lack the understanding that people transition from being babies to children to grown-ups over time.[9] To make sense of these categories, children may draw on their burgeoning understanding of number words. For example, they might notice that different age groups are referred to by different numerical ages and that grown-ups tend to have greater ages than children but also that ages are roughly correlated with size cues. Although this information alone would not support the inference that age is chronological, it might set the terrain for this discovery. This is because around the same time that children map larger numbers to magnitudes, they also learn the meanings of duration words like "day," "month," and "year," albeit gradually and over the course of several years (Grant & Suddendorf, 2011; Shatz et al., 2010; Tillman & Barner, 2015). Given this, it is possible that children might discover that age denotes quantities of time by learning the meanings of number words and that year denotes some amount of time, and thus that a 10-year-old must be older than a 5-year-old because they differ with respect to a quantity of years.

Another factor children might draw upon when learning about age is their ability to represent approximate temporal magnitudes (see Brannon et al., 2008; Wearden & Lejeune, 2008, for discussions of this ability). For example, children might implicitly measure the amount of time different people have been alive and make age judgments on this basis. However, although past studies find that children can discriminate temporal magnitudes for very brief stimuli (Brannon et al., 2007; Droit-Volet & Wearden, 2001, 2002; Droit-Volet et al., 2001; Levin, 1977, 1979; Levin & Gilat, 1983; Levin et al., 1978, 1980; Odic, 2018; Qu et al., 2021; VanMarle & Wynn, 2006), little is known about their ability to represent larger quantities of time. Adding to this, past studies find that children are very slow to associate duration words like "day," "month," and "year" with approximate magnitudes and likely do not do so until around age 6 or 7 when they learn their definitions (e.g., that 1 hr = 60 min; Tillman & Barner, 2015). A final challenge to using this

---

[9] We thank an anonymous reviewer for noting that although adults possess a chronological concept of age such that they represent the progression through time of being a baby to becoming an adult, they may also reason about age categorically, segmenting people across the lifespan into different groups. This is supported by evidence that adults organize their autobiographical memory and conceptualization of their future into time periods corresponding to distinct "chapters" of their life, with temporal landmarks marking boundaries between them (Fuhrman & Wyer, 1988; Peetz & Wilson, 2013, 2014; Shum, 1998; Skowronski et al., 2007; Thomsen, 2009).

strategy is that, absent linguistic measures of approximate duration, children would be limited to measuring age for people who are born after them (because all people born before them have spent equal time alive, from their perspective). Given these constraints, although it is possible that children are able to represent and compare larger temporal magnitudes, it seems unlikely that such intuitions are a primary source of evidence that they use to make judgments of relative age.

While it remains uncertain what role temporal reasoning plays in early age judgments, our findings make clear that young children recruit cues other than size in reasoning about age. Nevertheless, some puzzles based on seemingly conflicting findings remain. For example, if children in fact only rely on size to judge age when other, more informative, cues are not available, how should we make sense of past reports that children sometimes justify size-based judgments by saying, for example, "he's bigger so he has to be older" (Looft, 1971)? One possibility is that such justifications reflect the fact that children understand that size is a strong predictor of age rather than suggesting that they believe that the bigger person actually *has* to be older. Consistent with this, children do not exhibit adultlike understanding of the modal force of expressions like "have to" until 5 or 6 years of age, often interpreting "have to" as a possibility modal (Cournane et al., 2026; Leahy & Zalnieriunas, 2021). Thus, children's use of "have to" when describing the relationship between age and size may reflect their still-developing understanding of modal expressions rather than a belief that age is actually defined by size.

Another puzzle that remains is why children's ordering of autobiographical memories was unrelated to their age judgments. This result is perhaps surprising because an adultlike ability to order events in time seems important to acquiring a chronological understanding of age and aging. However, there are multiple reasons why we might expect children's ordering of autobiographical memories and age judgments to be unrelated. First, as previously mentioned, changes in children's age judgments might not have reflected a shift in conceptualizing age in terms of time but rather the application of a heuristic that the person with the bigger number is older. Alternatively, it is possible that even if children who succeeded at judging relative age conceptualized age as time spent alive, their ability to compare ages may not be related to their ability to order memories because they do not reason about people's ages by ordering their births but rather by making judgments based on the magnitudes of people's ages. That is, a child who knows that age refers to time spent alive and succeeds on age judgment tasks by comparing the magnitude of the temporal durations denoted by people's ages might not necessarily be adept at sequencing events in the past, including people's births and their own memories. Another possibility, though, is that such a relation between these two abilities exists but was simply not captured in our study. For example, some 3- and 4-year-old children may have performed poorly on the autobiographical memory task because it required not only understanding how to order events but also the ability to remember which event occurred more recently, as well as an understanding of the phrases "a long time ago" and a "short time ago" given that children are still learning gradable adjectives such as long and short at this age (e.g., Barner & Snedeker, 2008; Bartlett, 1976; Clark, 1970; Ehri, 1976; Maratsos, 1973; Nelson & Benedict, 1974; Ryalls, 2000; Smith et al., 1986; Townsend, 1976). Moreover, although we asked parents to identify events that occurred a month ago or a week ago, we could not independently verify the temporal precision of their choices, also potentially introducing error. For such reasons, the autobiographical memory task may not be a good proxy for children's intuitive understanding of time. This raises the possibility that other tasks might be, offering an avenue of future research.

The present study features several other limitations that should also be addressed in future work. One limitation relates to the certainty with which we can conclude that children's ability to judge relative age is related to their acquisition of a number system. It is possible that the observed relationship between children's performance on the later-greater task and the age judgment task may instead be explained by additional variables such as the child's domain-general capacities (e.g., executive functioning) or level of attentiveness to the experimental tasks.[10] However, the specificity of the relation between children's performance on the later-greater task and the age judgment task (i.e., that children's age judgment performance in Experiment 1 when numerical ages were provided was predicted by children's ability to judge the relative magnitude of the particular pair of numbers used in the numerical age expressions) suggests that the relation between these variables was not merely the result of a more general underlying factor. Next, our sample was limited to children from one Western culture, limiting the generalizability of our findings to cultures whose rituals and linguistic expressions related to aging differ from those of Western cultures. For example, the Amondawa, an Amazonian tribe whose numerical system does not extend beyond the number four, do not assign numerical ages to people. Instead, the Amondawa have words for age groups akin to "baby," "man," "old man," and so forth and change a person's proper name when a person transitions to a different life stage or role in the community, or when a new person in their family is born, with the newborn sometimes taking the name previously held by the youngest family member (Sinha et al., 2011). Moreover, members of some cultures view time as cyclical rather than linear (see Whitrow, 1988, for a discussion of the history of such cross-cultural differences), inviting the question of how this view might impact people's conception of age and its development. Finally, we did not record whether participants had siblings in the present study. This is potentially important because children with older or younger siblings may have more direct familiarity with the fact that people who are born later are younger, while people born earlier are older. Future studies should explore how such differences in cultural, linguistic, and family backgrounds influence reasoning about age, and its development.

In summary, the present study found that children as young as 3 use a constellation of cues to make age judgments, including physical morphology, and that age judgments improve as children learn number words, especially when provided with explicit numerical cues to age. Meanwhile, size information influenced age judgments chiefly when other stronger cues were not available. Future studies should continue to explore whether early emerging age judgments are rooted in a chronological understanding of age, or reflect categorical binning of individuals as, for example, young or old, or heuristics that, for example, people with larger numbers in their ages are older.

---

[10] We thank an anonymous reviewer for this suggestion.

References

Alton, J., Cimpian, A., & Butler, L. P. (2025). The role of gender labels and gendered appearances in children's social inferences. *Cognition*, *265*, Article 106281. https://doi.org/10.1016/j.cognition.2025.106281

Backscheider, A. G., Shatz, M., & Gelman, S. A. (1993). Preschoolers' ability to distinguish living kinds as a function of regrowth. *Child Development*, *64*(4), 1242–1257. https://doi.org/10.2307/1131337

Barner, D., & Snedeker, J. (2008). Compositionality and statistics in adjective acquisition: 4-year-olds interpret tall and short based on the size distributions of novel noun referents. *Child Development*, *79*(3), 594–608. https://doi.org/10.1111/j.1467-8624.2008.01145.x

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bartlett, E. J. (1976). Sizing things up: The acquisition of the meaning of dimensional adjectives. *Journal of Child Language*, *3*(2), 205–219. https://doi.org/10.1017/S0305000900001458

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bian, L., Leslie, S.-J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*(6323), 389–391. https://doi.org/10.1126/science.aah6524

Bian, L., Leslie, S.-J., & Cimpian, A. (2018). Evidence of bias against girls and women in contexts that emphasize intellectual ability. *American Psychologist*, *73*(9), 1139–1153. https://doi.org/10.1037/amp0000427

Bigler, R. S., Averhart, C. J., & Liben, L. S. (2003). Race and the workforce: Occupational status, aspirations, and stereotyping among African American children. *Developmental Psychology*, *39*(3), 572–580. https://doi.org/10.1037/0012-1649.39.3.572

Boskovic, K., & Barner, D. (2026). *How do children construct a concept of age?* Open Science Framework. https://osf.io/j6sc8/

Brannon, E. M., Libertus, M. E., Meck, W. H., & Woldorff, M. G. (2008). Electrophysiological measures of time processing in infant and adult brains: Weber's law holds. *Journal of Cognitive Neuroscience*, *20*(2), 193–203. https://doi.org/10.1162/jocn.2008.20016

Brannon, E. M., Suanda, S., & Libertus, K. (2007). Temporal discrimination increases in precision over development and parallels the development of numerosity discrimination. *Developmental Science*, *10*(6), 770–777. https://doi.org/10.1111/j.1467-7687.2007.00635.x

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389–411. https://doi.org/10.1037/met0000159

Brooks, J., & Lewis, M. (1976). Infants' responses to strangers: Midget, adult, and child. *Child Development*, *47*(2), 323–332. https://doi.org/10.2307/1128785

Burdett, E. R. R., & Barrett, J. L. (2016). The circle of life: A cross-cultural comparison of children's attribution of life-cycle traits. *British Journal of Developmental Psychology*, *34*(2), 276–290. https://doi.org/10.1111/bjdp.12131

Burke, J. L. (1982). Young children's attitudes and perceptions of older adults. *The International Journal of Aging and Human Development*, *14*(3), 205–222. https://doi.org/10.2190/4J7N-RG79-HJQR-FLDN

Carey, S. (1985). *Conceptual change in childhood*. Massachusetts Institute of Technology Press.

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*, *133*(1), 59–68. https://doi.org/10.1162/001152604772746701

Carey, S. (2009). *The origin of concepts*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195367638.001.0001

Carey, S., & Barner, D. (2019). Ontogenetic origins of human integer representations. *Trends in Cognitive Sciences*, *23*(10), 823–835. https://doi.org/10.1016/j.tics.2019.07.004

Clark, H. H. (1970). The primitive nature of children's relational concepts. In J. R. Haveys (Ed.), *Cognition and the development of language* (pp. 269–278). Wiley.

Condry, K. F., & Spelke, E. S. (2008). The development of language and abstract concepts: The case of natural number. *Journal of Experimental Psychology: General*, *137*(1), 22–38. https://doi.org/10.1037/0096-3445.137.1.22

Cournane, A., Dieuleveut, A., Repetti-Ludlow, C., & Hacquard, V. (2026). Word learning challenges explain nonadult possibility language comprehension in preschoolers. *Developmental Psychology*, *62*(5), 1048–1066. https://doi.org/10.1037/dev0002061

Davidson, K., Eng, K., & Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition*, *123*(1), 162–173. https://doi.org/10.1016/j.cognition.2011.12.013

Diesendruck, G., & HaLevi, H. (2006). The role of language, appearance, and culture in children's social category-based induction. *Child Development*, *77*(3), 539–553. https://doi.org/10.1111/j.1467-8624.2006.00889.x

Droit-Volet, S., Clément, A., & Wearden, J. (2001). Temporal generalization in 3- to 8-year-old children. *Journal of Experimental Child Psychology*, *80*(3), 271–288. https://doi.org/10.1006/jecp.2001.2629

Droit-Volet, S., & Wearden, J. H. (2001). Temporal bisection in children. *Journal of Experimental Child Psychology*, *80*(2), 142–159. https://doi.org/10.1006/jecp.2001.2631

Droit-Volet, S., & Wearden, J. H. (2002). Speeding up an internal clock in children? Effects of visual flicker on subjective duration. *The Quarterly Journal of Experimental Psychology Section B*, *55*(3), 193–211. https://doi.org/10.1080/02724990143000252

Dukler, N., & Liberman, Z. (2022). Children use race to infer who is "in charge." *Journal of Experimental Child Psychology*, *221*, Article 105447. https://doi.org/10.1016/j.jecp.2022.105447

Edwards, C. P., & Lewis, M. (1979). Young children's concepts of social relations. In M. Lewis & L. A. Rosenblum (Eds.), *The child and its family* (pp. 245–266). Springer. https://doi.org/10.1007/978-1-4684-3435-4_13

Ehri, L. C. (1976). Comprehension and production of adjectives and seriation. *Journal of Child Language*, *3*(3), 369–384. https://doi.org/10.1017/S0305000900007248

Elenbaas, L., & Killen, M. (2016). Age-related changes in children's associations of economic resources and race. *Frontiers in Psychology*, *7*, Article 884. https://doi.org/10.3389/fpsyg.2016.00884

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fouquet, N., Megalakaki, O., & Labrell, F. (2017). Children's understanding of animal, plant, and artifact properties between 3 and 6 years. *Infant and Child Development*, *26*(6), Article e2032. https://doi.org/10.1002/icd.2032

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage Publications. https://www.john-fox.ca/Companion/

French, J. A., Menendez, D., Herrmann, P. A., Evans, E. M., & Rosengren, K. S. (2018). Cognitive constraints influence an understanding of life-cycle change. *Journal of Experimental Child Psychology*, *173*, 205–221. https://doi.org/10.1016/j.jecp.2018.03.018

Fuhrman, R. W., & Wyer, R. S. (1988). Event memory: Temporal-order judgments of personal life experiences. *Journal of Personality and Social Psychology*, *54*(3), 365–384. https://doi.org/10.1037/0022-3514.54.3.365

Fuson, K. C. (1988). *Children's counting and concepts of number*. Springer-Verlag.

Galper, A., Jantz, R. K., Seefeldt, C., & Serock, K. (1981). The child's concept of age and aging. *International Journal of Aging and Human Development*, *12*(2), 149–157. https://doi.org/10.2190/RWC8-3E80-M79U-K97C

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Harvard University Press.

Grant, J. B., & Suddendorf, T. (2011). Production of temporal terms by 3-, 4-, and 5-year-old children. *Early Childhood Research Quarterly*, *26*(1), 87–95. https://doi.org/10.1016/j.ecresq.2010.05.002

Herrmann, P. A., French, J. A., DeHart, G. B., & Rosengren, K. S. (2013). Essentialist reasoning and knowledge effects on biological reasoning in young children. *Merrill-Palmer Quarterly*, *59*(2), 198–220. https://doi.org/10.1353/mpq.2013.0008

Inagaki, K., & Hatano, G. (1996). Young children's recognition of commonalities between animals and plants. *Child Development*, *67*(6), 2823–2840. https://doi.org/10.2307/1131754

Inagaki, K., & Sugiyama, K. (1988). Attributing human characteristics: Developmental changes in over- and underattribution. *Cognitive Development*, *3*(1), 55–70. https://doi.org/10.1016/0885-2014(88)90030-5

Jipson, J. L., & Callanan, M. A. (2003). Mother–child conversation and children's understanding of biological and nonbiological changes in size. *Child Development*, *74*(2), 629–644. https://doi.org/10.1111/1467-8624.7402020

Klavir, R., & Leiser, D. (2002). When astronomy, biology, and culture converge: Children's conceptions about birthdays. *The Journal of Genetic Psychology*, *163*(2), 239–253. https://doi.org/10.1080/00221320209598681

Kogan, N., Stevens, J., & Shelton, F. C. (1961). Age differences: A developmental study of discriminability and affective response. *Journal of Abnormal and Social Psychology*, *62*(2), 221–230. https://doi.org/10.1037/h0046242

Kratochwill, T. R., & Goldman, J. A. (1973). Developmental changes in children's judgments of age. *Developmental Psychology*, *9*(3), 358–362. https://doi.org/10.1037/h0034919

Kuczaj, S. A., & Lederberg, A. R. (1977). Height, age, and function: Differing influences on children's comprehension of 'younger' and 'older'. *Journal of Child Language*, *4*(3), 395–416. https://doi.org/10.1017/S0305000900001768

Le Corre, M. (2014). Children acquire the later-greater principle after the cardinal principle. *British Journal of Developmental Psychology*, *32*(2), 163–177. https://doi.org/10.1111/bjdp.12029

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*(2), 395–438. https://doi.org/10.1016/j.cognition.2006.10.005

Leahy, B., & Zalnieriunas, E. (2021). Might and might not: Children's conceptual development and the acquisition of modal verbs. *Semantics and Linguistic Theory* (pp. 426–445). https://doi.org/10.3765/salt.v31i0.5082

Lenth, R. (2025). *Emmeans* (Version 1.11.1-00001). https://rvlenth.github.io/emmeans/

Levin, I. (1977). The development of time concepts in young children: Reasoning about duration. *Child Development*, *48*(2), 435–444. https://doi.org/10.2307/1128636

Levin, I. (1979). Interference of time-related and unrelated cues with duration comparisons of young children: Analysis of Piaget's formulation of the relation of time and speed. *Child Development*, *50*(2), 469–477. https://doi.org/10.2307/1129425

Levin, I., & Gilat, I. (1983). A developmental analysis of early time concepts: The equivalence and additivity of the effect of interfering cues on duration comparisons of young children. *Child Development*, *54*(1), 78–83. https://doi.org/10.2307/1129863

Levin, I., Gilat, I., & Zelniker, T. (1980). The role of cue salience in the development of time concepts: Duration comparisons in young children. *Developmental Psychology*, *16*(6), 661–671. https://doi.org/10.1037/0012-1649.16.6.661

Levin, I., Israeli, E., & Darom, E. (1978). The development of time concepts in young children: The relations between duration and succession. *Child Development*, *49*(3), 755–764. https://doi.org/10.2307/1128245

Looft, W. R. (1971). Children's judgments of age. *Child Development*, *42*(4), 1282–1284. https://doi.org/10.2307/1127812

Looft, W. R., Rayman, J. R., & Rayman, B. B. (1972). Children's judgments of age in Sarawak. *The Journal of Social Psychology*, *86*(2), 181–185. https://doi.org/10.1080/00224545.1972.9918616

Mandalaywala, T. M., Tai, C., & Rhodes, M. (2020). Children's use of race and gender as cues to social status. *PLOS ONE*, *15*(6), Article e0234398. https://doi.org/10.1371/journal.pone.0234398

Maratsos, M. P. (1973). Decrease in the understanding of the word "big" in preschool children. *Child Development*, *44*(4), 747–752. https://doi.org/10.2307/1127719

Margett, T. E., & Witherington, D. C. (2011). The nature of preschoolers' concept of living and artificial objects. *Child Development*, *82*(6), 2067–2082. https://doi.org/10.1111/j.1467-8624.2011.01661.x

Martin, C. L. (1989). Children's use of gender-related information in making social judgments. *Developmental Psychology*, *25*(1), 80–88. https://doi.org/10.1037/0012-1649.25.1.80

Nelson, K., & Benedict, H. (1974). The comprehension of relative, absolute, and contrastive adjectives by young children. *Journal of Psycholinguistic Research*, *3*(4), 333–342. https://doi.org/10.1007/BF01068168

Odic, D. (2018). Children's intuitive sense of number develops independently of their perception of area, density, length, and time. *Developmental Science*, *21*(2), Article e12533. https://doi.org/10.1111/desc.12533

Olson, K. R., Shutts, K., Kinzler, K. D., & Weisman, K. G. (2012). Children associate racial groups with wealth: Evidence from South Africa. *Child Development*, *83*(6), 1884–1899. https://doi.org/10.1111/j.1467-8624.2012.01819.x

OpenAI. (2024). *DALL-E* (Version 3) [Artificial intelligence system].

Pathman, T., Larkina, M., Burch, M., & Bauer, P. J. (2013). Young children's memory for the times of personal past events. *Journal of Cognition and Development*, *14*(1), 120–140. https://doi.org/10.1080/15248372.2011.641185

Peetz, J., & Wilson, A. E. (2013). The post-birthday world: Consequences of temporal landmarks for temporal self-appraisal and motivation. *Journal of Personality and Social Psychology*, *104*(2), 249–266. https://doi.org/10.1037/a0030477

Peetz, J., & Wilson, A. E. (2014). Marking time: Selective use of temporal landmarks as barriers between current and future selves. *Personality and Social Psychology Bulletin*, *40*(1), 44–56. https://doi.org/10.1177/0146167213501559

Piaget, J. (1971). *The child's conception of time*. Ballantine Books.

Qu, F., Shi, X., Zhang, A., & Gu, C. (2021). Development of young children's time perception: Effect of age and emotional localization. *Frontiers in Psychology*, *12*. Article 688165. https://doi.org/10.3389/fpsyg.2021.688165

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rosengren, K. S., Gelman, S. A., Kalish, C. W., & McCormick, M. (1991). As time goes by: Children's early understanding of growth in animals. *Child Development*, *62*(6), 1302–1320. https://doi.org/10.2307/1130808

Ryalls, B. O. (2000). Dimensional adjectives: Factors affecting children's ability to compare objects using novel words. *Journal of Experimental Child Psychology*, *76*(1), 26–49. https://doi.org/10.1006/jecp.1999.2537

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, *108*(3), 662–674. https://doi.org/10.1016/j.cognition.2008.05.007

Schaeffer, B., Eggleston, V. H., & Scott, J. L. (1974). Number development in young children. *Cognitive Psychology*, *6*(3), 357–379. https://doi.org/10.1016/0010-0285(74)90017-6

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, *20*(2), 416–420. https://doi.org/10.1093/beheco/arn145

Schneider, R. M., Pankonin, A., Schachner, A., & Barner, D. (2021). Starting small: Exploring the origins of successor function knowledge. *Developmental Science*, *24*(4), Article e13091. https://doi.org/10.1111/desc.13091

Schneider, R. M., Sullivan, J., Guo, K., & Barner, D. (2021). What counts? Sources of knowledge in children's acquisition of the successor function. *Child Development*, *92*(4), e476–e492. https://doi.org/10.1111/cdev.13524

Seefeldt, C., Jantz, R. K., Galper, A., & Serock, K. (1977). Using pictures to explore children's attitudes toward the elderly. *The Gerontologist*, *17*(6), 506–512. https://doi.org/10.1093/geront/17.6.506

Shatz, M., Tare, M., Nguyen, S. P., & Young, T. (2010). Acquiring non-object terms: The case for time words. *Journal of Cognition and Development*, *11*(1), 16–36. https://doi.org/10.1080/15248370903453568

Shum, M. S. (1998). The role of temporal landmarks in autobiographical memory processes. *Psychological Bulletin*, *124*(3), 423–442. https://doi.org/10.1037/0033-2909.124.3.423

Shutts, K., Brey, E. L., Dornbusch, L. A., Slywotzky, N., & Olson, K. R. (2016). Children use wealth cues to evaluate others. *PLOS ONE*, *11*(3), Article e0149360. https://doi.org/10.1371/journal.pone.0149360

Shutts, K., Roben, C. K. P., & Spelke, E. S. (2013). Children's use of social categories in thinking about people and social relationships. *Journal of Cognition and Development*, *14*(1), 35–62. https://doi.org/10.1080/15248372.2011.638686

Sinha, C., Sinha, V. D. S., Zinken, J., & Sampaio, W. (2011). When time is not space: The social and linguistic construction of time intervals and temporal event relations in an Amazonian culture. *Language and Cognition*, *3*(1), 137–169. https://doi.org/10.1515/langcog.2011.006

Skowronski, J. J., Ritchie, T. D., Walker, W. R., Betz, A. L., Sedikides, C., Bethencourt, L. A., & Martin, A. L. (2007). Ordering our world: The quest for traces of temporal organization in autobiographical memory. *Journal of Experimental Social Psychology*, *43*(5), 850–856. https://doi.org/10.1016/j.jesp.2006.10.001

Smith, L. B., Cooney, N. J., & McCord, C. (1986). What is "high"? The development of reference points for "high" and "low." *Child Development*, *57*(3), 583–602. https://doi.org/10.2307/1130338

Taylor, E., Steele, C., & Roberto, K. (1982). Preschool children's discrimination of age. *Perceptual and Motor Skills*, *54*(2), 539–542. https://doi.org/10.2466/pms.1982.54.2.539

Thomsen, D. K. (2009). There is more to life stories than memories. *Memory*, *17*(4), 445–457. https://doi.org/10.1080/09658210902740878

Tillman, K. A., & Barner, D. (2015). Learning the language of time: Children's acquisition of duration words. *Cognitive Psychology*, *78*, 57–77. https://doi.org/10.1016/j.cogpsych.2015.03.001

Townsend, D. J. (1976). Do children interpret 'marked' comparative adjectives as their opposites? *Journal of Child Language*, *3*(3), 385–396. https://doi.org/10.1017/S030500090000725X

U.S. Census Bureau. (2024). *American Community Survey (ACS) and Puerto Rico Community Survey (PCRS). 5 year estimates: 2019-2023*. https://www.census.gov/quickfacts/fact/table/sandiegocountycalifornia

VanMarle, K., & Wynn, K. (2006). Six-month-old infants use analog magnitudes to represent duration. *Developmental Science*, *9*(5), F41–F49. https://doi.org/10.1111/j.1467-7687.2006.00508.x

Waxman, S. R. (2010). Names will never hurt me? Naming and the development of racial and gender categories in preschool-aged children. *European Journal of Social Psychology*, *40*(4), 593–610. https://doi.org/10.1002/ejsp.732

Wearden, J. H., & Lejeune, H. (2008). Scalar properties in human timing: Conformity and violations. *Quarterly Journal of Experimental Psychology*, *61*(4), 569–587. https://doi.org/10.1080/17470210701282576

Whitrow, G. J. (1988). *Time in history: Views of time from prehistory to the present day*. Oxford University Press.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), Article 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., & Müller, K. (2022). *dplyr: A grammar of data manipulation*. https://dplyr.tidyverse.org

Wickham, H., Vaughan, D., & Girlich, M. (2025). *tidyr: Tidy messy data* (R package Version 1.3.1.9000) [Computer software]. https://github.com/tidyverse/tidyr

Winter, B. (2013). *Linear models and linear mixed effects models in R*. arXiv. https://doi.org/10.48550/arXiv.1308.5499

Woolley, J. D., & Rhoads, A. M. (2019). Now I'm 3: Young children's concepts of age, aging, and birthdays. *Imagination, Cognition and Personality*, *38*(3), 268–289. https://doi.org/10.1177/0276236617748129

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155–193. https://doi.org/10.1016/0010-0277(90)90003-3

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*(2), 220–251. https://doi.org/10.1016/0010-0285(92)90008-p

Zhu, L., & Fang, F. (2000). Development of Chinese preschoolers' understanding of biological phenomena: Growth and aliveness. *International Journal of Behavioral Development*, *24*(1), 105–110. https://doi.org/10.1080/016502500383539

Received November 26, 2025
Revision received March 15, 2026
Accepted April 27, 2026 ■