Behavior Research Methods
https://doi.org/10.3758/s13428-022-02056-3

Maximizing valid eye-tracking data in human and macaque infants by optimizing calibration and adjusting areas of interest

Guangyu Zeng[1] · **Elizabeth A. Simpson**[1] · **Annika Paukner**[2]

Accepted: 24 December 2022
© The Psychonomic Society, Inc. 2023

Abstract
Remote eye tracking with automated corneal reflection provides insights into the emergence and development of cognitive, social, and emotional functions in human infants and non-human primates. However, because most eye-tracking systems were designed for use in human adults, the accuracy of eye-tracking data collected in other populations is unclear, as are potential approaches to minimize measurement error. For instance, data quality may differ across species or ages, which are necessary considerations for comparative and developmental studies. Here we examined how the calibration method and adjustments to areas of interest (AOIs) of the Tobii TX300 changed the mapping of fixations to AOIs in a cross-species longitudinal study. We tested humans ($N = 119$) at 2, 4, 6, 8, and 14 months of age and macaques (*Macaca mulatta*; $N = 21$) at 2 weeks, 3 weeks, and 6 months of age. In all groups, we found improvement in the proportion of AOI hits detected as the number of successful calibration points increased, suggesting calibration approaches with more points may be advantageous. Spatially enlarging and temporally prolonging AOIs increased the number of fixation-AOI mappings, suggesting improvements in capturing infants' gaze behaviors; however, these benefits varied across age groups and species, suggesting different parameters may be ideal, depending on the population studied. In sum, to maximize usable sessions and minimize measurement error, eye-tracking data collection and extraction approaches may need adjustments for the age groups and species studied. Doing so may make it easier to standardize and replicate eye-tracking research findings.

Keywords Visual attention · Infancy · Eye gaze · Measurement · Vision · Orienting · Developmental psychology · Comparative psychology

Introduction

Eye tracking is a popular method to examine the development of cognitive, social, and emotional functions in preverbal and non-verbal populations, including human infants (see Gredebäck et al., 2009; Oakes, 2012 for reviews) and non-human primates (see Hopper et al., 2021; Machado & Nelson, 2011 for reviews). Compared to simple observations and manual coding of gaze behaviors, remote screen-based eye tracking has numerous advantages. For example, it can automatically track more complex gaze patterns (speed and direction of gaze shifts) on more complex stimuli (dynamic, multi-part videos) while also enabling a high spatial and temporal resolution, in addition to being less laborious and more accurate (Oakes, 2012; Wass et al., 2013). However, the eye-tracking data quality (i.e., accuracy, precision, and usability of the gaze signal; Holmqvist et al., 2011) collected from human infants and non-human primates (referred to as "primates" hereafter) remains unclear. In the current study, we targeted the usability aspect of eye-tracking data quality by examining two approaches to improve the capture of meaningful and valid measures of gaze behaviors. One approach focuses on calibration methods. The other approach focuses on data extraction methods.

Infant and primate eye tracking: Opportunities and challenges

Remote eye-tracking methods have been increasingly popular in infant and animal research in the last couple of decades, offering opportunities and challenges. Comparative

✉ Elizabeth A. Simpson
simpsone@miami.edu

[1] Department of Psychology, University of Miami, Coral Gables, FL, USA

[2] Department of Psychology, Nottingham Trent University, Nottingham, UK

eye-tracking studies have reported similarities in social attention development between human and primate infants (Damon et al., 2017; Jakobsen et al., 2016; Maylott et al., 2020; Parr, Murphy, et al., 2016b; Simpson et al., 2017). Eye-tracking technology is also useful in measuring individual differences in infancy, as well as atypicalities in social attention in human infants and primates (Jones & Klin, 2013; Machado et al., 2015). For example, across species, more eye contact is associated with greater sociality (Pons et al., 2019; Ryan et al., 2020), and females were more socially attentive than males (Gluckman & Johnson, 2013; Simpson, Nicolini, et al., 2016b). Human infants who were later diagnosed with autism spectrum disorder showed a decline in looking at eyes from 2 to 6 months, while typically developing infants increased eye looking across those ages (Jones & Klin, 2013). A similar pattern of less attention to other macaques' eyes was found in maternal immune activated rhesus macaque infants (a method to induce autistic traits in primates), compared to the control group (Machado et al., 2015). In sum, across species, eye tracking may help capture species-typical developmental changes, as well as identify individual differences in infancy.

Despite the growing popularity, there are substantive obstacles to address to study these populations to ensure research methods are appropriately capturing infants' and primates' abilities. Collecting reliable eye-tracking data from infants and animals is more difficult than from human adults as infants and animals are less able to understand and follow instructions, and more likely to move during testing, which generates unstable data and reduces data quality (e.g., poorer accuracy and precision, more error and data loss; for a review: Hessels & Hooge, 2019; Wass et al., 2013 in human infants; Hopper et al., 2021 in primates). Compared to older individuals, infants have less oculomotor control and shorter attention spans, making calibration—the process of measuring characteristics of each participant's eyes to improve eye-tracking accuracy—more difficult (Feng, 2011). Thus, we need to be aware of and minimize confounds in data quality (due to age, species, or other group differences) before interpreting findings based on eye-tracking measures (Hessels & Hooge, 2019).

Mapping fixations on areas of interest depends on eye-tracking spatial and temporal accuracy

In eye-tracking studies, the most commonly used data in analyses are fixations. A fixation is a group of raw gaze points that appears on a location within a particular threshold of eye movement parameters, such as velocity, angle, and duration (Rayner, 2009). Fixations are not the direct products of eye tracking but the computational outputs of a series of algorithms, which group raw gaze data together to reduce noise and small fluctuations. Fixations can reflect various attentional processes, such as sustained attention (i.e., holding attention on a target) and selective attention (i.e., allocating attention to specific information), which are indicative of cognitive functions (Aslin, 2007; Liversedge & Findlay, 2000). Extracting meaningful and valid fixations—located in stimulus regions of interest—is a necessary step in eye-tracking analysis. This step is typically accomplished by creating areas of interest (AOIs) of different sizes and shapes, which can be activated and deactivated at specific times, and may move dynamically, to capture fixations aligned with static or moving regions of interest (Dupierrix et al., 2014; Gluckman & Johnson, 2013; Gredebäck et al., 2009; Senju & Csibra, 2008).

Obtaining reliable and valid fixation data relies on detecting real gazes on the stimuli (true-positive gazes) and excluding noise (false-positive gazes), all of which are affected by the spatial accuracy of raw data—the locations of the collected gaze data relative to true gaze locations (Morgante et al., 2012). An accuracy test for a Tobii TX300 eye tracker reported spatial deviations in accuracy: 18-month-old infants ($N = 28$) had an average of 1.31° (range, 0.18–3.85°) and 30-month-old infants ($N = 31$) had an average of 1.29° (range, 0.67–2.33°) (Dalrymple et al., 2018). A large recent study reported median spatial accuracy of the Tobii TX300 for 4- to 7-month-olds ($N = 490$) as 2.7°, for 8- to 12-month-olds ($N = 486$) as 1.6°, and for 3-year-olds ($N = 131$) as approximately 1°, reflecting increasing spatial accuracy with age (De Kloe et al., 2022). Notably, these younger infant spatial accuracies were lower than that reported for human adults on the same Tobii TX300 eye tracker (Dalrymple et al., 2018). Together, these findings point out that gazes on the stimuli may fail to be captured because they were detected just beyond the border of the stimuli, an issue that may be more prominent at younger ages, raising the concern about losing valid gaze data. Additionally, because the eye-tracking system is only estimating the central gaze point, this estimate does not consider the actual area of the viewer's foveated visual field (Akbas & Eckstein, 2017; Groot et al., 1994; O'Shea, 1991). Consequently, a viewer could be focused just outside of the target but still be seeing the target within the foveal visual field. Therefore, researchers should consider ways to collect and analyze eye-tracking data to maximize inclusion of valid fixations while minimizing the chance of capturing noisy data.

Moreover, in a review of primate eye-tracking studies, Tobii eye trackers were the most common among 32 non-invasive eye-tracking studies from 2009 to 2019 (Hopper et al., 2021). Two studies in juvenile and adult chimpanzees, one with a Tobii T60 and another with a Tobii X120, reported preliminary spatial accuracy of 0.15–0.66° deviations in small samples of chimpanzees ($N = 6$ for each study; Hirata et al., 2010; Kano & Tomonaga, 2009), comparable to accuracy reported for human adults. However, it remains

unclear whether this level of accuracy is generalizable to primates of younger ages and other species.

In addition to being affected by the eye tracker's spatial accuracy, the validity of fixation-AOI mappings, and the eye-tracking measures calculated using these fixations, may also be affected by the eye tracker's temporal accuracy (i.e., the timing of the eye movements relative to stimulus events; Morgante et al., 2012). Only a few studies have measured temporal accuracy, and those that have, have only been in human adults (Morgante et al., 2012; Xue et al., 2017). One study reported a 54-ms delay in the temporal accuracy of a T60XL eye tracker (Morgante et al., 2012). The Tobii TX300 eye tracker has an even higher degree of temporal precision: 3.33 ms (De Kloe et al., 2022). However, it is unclear whether such high temporal accuracy can be achieved in infant and animal studies. Therefore, it is important for researchers to carefully account for temporal delays over the time course of their stimulus presentations when calculating eye-tracking measures to operationalize the constructs of interest.

Developmental changes in infants' visual and attentional systems with age

Developmental changes in infants' perceptual and attentional systems may also impact the mapping of fixations onto AOIs. As they develop, human and primate infants' visual acuity and attention improve (Chandna, 1991; Dobson & Teller, 1978; Ordy et al., 1964; Richards, 2004; Teller, 1981; Xiang et al., 2021). When viewing complex visual scenes, human 4- to 14-month-olds' fixations become more systematic and predictable, less driven by low-level salience, and more adult-like (Pomaranski et al., 2021). Human infants' ability to hold their attention on a stimulus also improves from 14 to 26 weeks, suggesting a reduction in head and body movements during eye tracking, a developmental increase in the stability in their fixations, and more stable gaze signal and data loss (Richards, 2004). Moreover, human and primate infants' attention orienting improves rapidly over the first 6 months after birth, enabling faster attention shifting and disengagement, and better visual tracking and responsiveness (Boothe et al., 1982; Johnson et al., 1991; McConnell & Bryson, 2005; Ross-Sheehy et al., 2015), which may improve the temporal mapping between infants' fixations and the stimuli. Moreover, across the first year after birth, macaque infants' visual acuity and motion sensitivity develop to adult-like levels and the noise signal in their visual neural system decreases (Kiorpes, 2015; Ordy et al., 1964). Therefore, human and macaque infants' fixations may be more likely to be captured within the AOI (i.e., better fixation-AOI-mappings) as they get older and develop better visual acuity, faster orienting, more gaze fixations, and increasingly stable gaze. However, these developmental changes vary across primate species and may differ from human developmental changes (Maylott et al., 2020; Teller, 1981). It remains unclear how such differences in visual and attentional systems across ages and species may differently influence the mapping of fixations onto AOIs among different populations. Therefore, a systematic and longitudinal evaluation of eye-tracking designs is needed to improve the ability to obtain reliable and valid eye-tracking measures in human and primate developmental research.

Decisions in Tobii infant calibration

Calibration procedures also affect eye-tracking data quality (i.e., accuracy, precision, data loss), which, in turn, affects fixation-AOI mapping. Yet, calibration procedures remain largely unexplored in infancy, a developmental period in which calibration is particularly challenging. When using an eye-tracking device, a calibration procedure takes place before beginning data collection to estimate the accuracy of the mapping between individual eye characteristics and actual gaze locations captured by the eye tracker (Gredebäck et al., 2009). An experimenter must make choices during the calibration procedure, such as the number of calibration points to attempt and the display durations of the calibration stimuli, each of which influence the subsequent quality of data collected (Carter & Luke, 2020). For example, the order of calibration points can be randomized (e.g., EyeLink; SR Research, 2007) or must proceed in a predetermined order (e.g., Tobii Studio; Tobii Technology, 2016).

While such flexibility may be achieved by using external toolboxes (Niehorster et al., 2020), the commonly used built-in calibration procedure for infants in the Tobii TX300 system is completed by having participants look at the calibration target as it appears in a certain number of predefined locations, presented sequentially, one at a time, in a predetermined order. While calibration procedures for human adult studies are relatively easy (as they have stable attention and can follow instructions), calibration is more challenging for studies of human and primate infants. For instance, calibration accuracy (i.e., average distance between calibration gaze samples and calibration location) and precision (i.e., standard deviation of the distances among repeated gaze samples on the same calibration location) were reported to be greater in human adults and school-age children than in 18- and 30-month-old toddlers (Dalrymple et al., 2018). Therefore, it is crucial to uncover whether specific decisions about calibration approaches can maximize calibration quality in human and primate infants.

One decision is the number of calibration points to use. While a larger number of calibration points is assumed to result in greater spatial accuracy than fewer calibration points (Gredebäck et al., 2009), it is not always feasible to obtain a large number of points, particularly with primates

and young infants who have limited attention spans. Indeed, studies in humans suggest using 5- or 6-point calibrations in infants at 4 months of age and older, and 2-point calibration in infants younger than 4 months, given their short attention spans (Gredebäck et al., 2009). A reduction in the number of calibration points may decrease the necessary total amount of time required for calibration, which decreases the likelihood that an infant becomes fussy, fatigued, or disinterested during the calibration procedure (Aslin & McMurray, 2004; Schlegelmilch & Wertz, 2019). Similar to studies in human infants, the majority of primate eye-tracking studies use only two calibration points because of difficulties maintaining primates' attention throughout a longer calibration procedure (see Hopper et al., 2021 for a review). In sum, the use of fewer calibration points appears common and to be based on the untested assumption that it may have some advantages over approaches with a greater number of calibration points, enabling participants to better maintain their attentiveness during and after calibration.

On the other hand, there may also be advantages to using a larger number of calibration points. When there are more points, they are closer together spatially, so infants must shift their attention shorter distances, and at more acute (i.e., smaller) angles, which are easier for younger infants, given their perceptual-attentional constraints (D'Entremont, 2000; Van Renswoude et al., 2016). For example, compared to the built-in 9-point calibration in the Tobii TX300 system, infants must shift their focus of attention across a longer distance and more obtuse (i.e., wider) angles for the built-in 5-point calibration (see Fig. 1 for details). Given that young

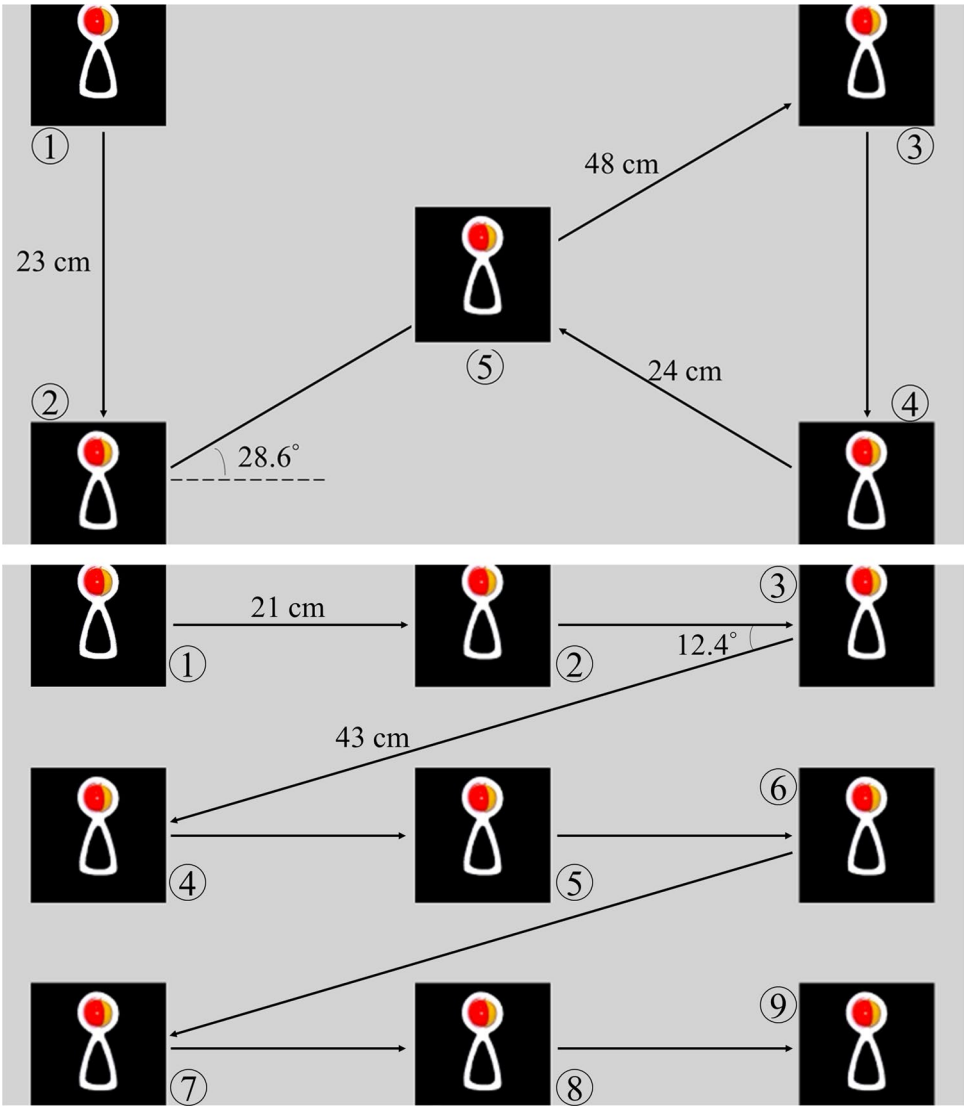

Fig. 1 Calibration screens for 5-point calibration (*top*) and 9-point calibration (*bottom*), displaying the order in which the calibration stimulus (here, a rattle) appeared (reflected in the *circled numbers*), distances between calibration points, and visual angles. In addition, for 2-point calibration in the Tobii TX300 system, infants must shift their focus of attention once across a distance of 48 cm (45.03° visual angle) and an angle of 28.6°

infants have a difficult time shifting their visual attention to stimuli across wider areas, including those further in their periphery (D'Entremont, 2000; Kulke et al., 2015), and have a horizontal bias, making it easier for them to shift their gaze horizontally than vertically (Van Renswoude et al., 2016), a 9-point calibration may be advantageous compared to a 5-point calibration when using the Tobii built-in calibration procedures because it requires them to shift their attention across shorter distances at less obtuse angles.

Shorter distances may also be advantageous for calibrating primates. For example, some smaller-bodied species of primate (e.g., squirrel monkeys and marmosets) need to shift their heads rather than just their eyes to visually scan these wider distances, compared to larger-bodied species, such as humans and chimpanzees (Heiney & Blazquez, 2011; Mitchell & Leopold, 2015). Indeed, some primate studies have had success using 9-point calibrations (e.g., gorillas, chimpanzees; Hopper et al., 2021). In sum, there is a need to systematically test whether one calibration approach is more advantageous than another in maximizing the amount and quality of usable data collected, and whether the calibration approach should vary depending on the study population.

Using AOI size and duration to improve fixation-AOI mapping

Researchers must also make a number of choices including the sizes and durations of AOIs to maximize the mapping of fixations onto AOIs. There are trade-offs to consider when creating AOIs. On the one hand, creating AOIs that perfectly align spatially and temporally with the borders of stimuli—often used in human adult studies—may seem ideal as they minimize the capture of fixations that would be inaccurately classified as being located on the stimulus (i.e., false positives; Vehlen et al., 2022) and enable the use of densely organized stimuli (e.g., arrays of 64 images; Simpson et al., 2019) without concern about overlapping AOIs (e.g., Hessels et al., 2016). However, not all fixations detected fall perfectly within the spatial and temporal borders of the stimuli (Dalrymple et al., 2018; McConnell & Bryson, 2005). Therefore, on the other hand, an AOI that perfectly aligns with the borders of stimuli may increase the risk of excluding meaningful fixations. Larger and longer AOIs located further apart from one another may capture more true fixations (Orquin et al., 2016). For example, enlarging AOI sizes relative to stimuli sizes may address the issues of spatial deviations in eye-tracking data, capturing additional valid fixations and reducing data loss (Dalrymple et al., 2018; Hirata et al., 2010; Kano & Tomonaga, 2009; Morgante et al., 2012). A study in human adults found that enlarging AOIs to 1.5° of visual angle around the stimulus border helps maximize the inclusion of true and valid fixations to the stimulus (Orquin et al., 2016). Larger AOIs may also serve as a robust solution when eye-tracking data are less accurate (Holmqvist et al., 2011; Vehlen et al., 2022), such as with infant eye tracking (Hessels et al., 2016). However, an AOI that is too large or too long may elevate the risk of including more noise and errors. Moreover, compared to stimulus-sized AOIs (that align with stimulus borders), larger AOIs that expand beyond stimulus borders also require a greater distance between stimuli, which may make their application only appropriate in sparsely organized stimuli (e.g., relative looking to two side-by-side images; Orquin et al., 2016). Therefore, it is crucial to consider how to balance the needs of maximizing valid fixation inclusion and minimizing noise and errors.

The quality of eye-tracking data may also vary with age during early infancy. For example, one study reported both spatial deviations and data loss decreased from 5 to 10 months of age in human infants using a Tobii TX300 eye tracker (Hessels & Hooge, 2019). Fixations remaining at a location after a stimulus disappears may be meaningful for measuring infants' attention and information processing, which researchers should carefully consider when designing developmental eye-tracking studies (McConnell & Bryson, 2005). The ideal methods for fixation-AOI mapping may vary with age, which highlights the need to examine the effects of various AOI parameters at different ages during infancy.

In sum, given the poorer eye-tracking data quality in infants compared to adults (Hessels & Hooge, 2019), their rapidly developing visual and attentional systems in the first year after birth (Brémond-Gignac et al., 2011; Kiorpes, 2015; Richards, 2004, 2010), and the unique challenges to eye-tracking studies in human and primate infants, there is a need to systematically examine participant age and species when deciding which spatial and temporal parameters to use for AOIs to balance the proportion of true and false positive fixations. Filling such gaps in our knowledge may make it easier to standardize and replicate eye-tracking research findings.

Current study

The current study aimed to provide a tentative initial set of guidelines for calibration procedures and for determining the sizes and durations of AOIs, to optimize fixation-AOI mapping in human and primate infant eye-tracking research studies across the first year after birth. We chose rhesus macaque monkeys because of the large number of eye-tracking studies in infants of this species (e.g., Mendelson et al., 1982; Muschinski et al., 2016; Parr, Brooks, et al., 2016a; Parr, Murphy, et al., 2016b; Paukner et al., 2014, 2018; Wang et al., 2020), as well as the fact that they share with humans many qualities related to their perceptual, cognitive, and social development, making them a popular model species for humans (Nelson et al., in press; Ryan et al., 2019).

In addition, compared to humans, macaque monkeys have more advanced visual acuity at birth (Ordy et al., 1964) and develop approximately four times faster (Boothe et al., 1982), enabling earlier and faster longitudinal eye-tracking studies than are possible in humans (Parr, Brooks, et al., 2016a).

Here, we tracked human and macaque infants' fixations on a rotating disk with stripes that appeared to move around the screen using a Tobii TX300 eye tracker, a popular system among developmental scientists (De Kloe et al., 2022). We longitudinally followed human infants at the age of 2, 4, 6, 8, and 14 months and rhesus macaque (*Macaca mulatta*) infants at the age of 2 weeks, 3 weeks, and 6 months. We selected these ages to cover a wide span of "early infancy" in both species. We explored whether the total number of registered calibration points (i.e., calibration points with fixations)—theorized to be an index of calibration quality (Wilkinson & Mitchell, 2014)—was associated with a greater number of valid fixation-AOI mappings. We examined how enlarging and prolonging the AOIs around the disk changed the fixation mappings onto the AOI. We also examined how the effects of AOI enlargement and prolongation changed developmentally within each species.

Methods

Participants

Human infants

A total of 119 infants participated in the current study (41.18% female). Among parents, 55% identified as Hispanic or Latino. Infants were racially diverse: 61% White, 18% Black, or African American, 14% multiracial, and 7% unknown/unreported (for details, see Table S4). Infants were tested longitudinally at 2 months ($N = 79$, $M_{age} = 8.98$ weeks, SD =0.93), 4 months ($N = 88$; $M_{age} = 17.97$ weeks, SD = 1.03), 6 months ($N = 83$, $M_{age} = 26.53$ weeks, SD = 1.52), 8 months ($N = 38$; $M_{age} = 35.15$ weeks, SD = 0.92), and 14 months of age ($N = 24$; $M_{age} = 60.39$ weeks, SD = 1.59). See Table S4 for detailed demographics. Infants were recruited from Miami, Florida and tested at the University of Miami. Infants were healthy, full-term (≥37 weeks gestation), and had normal or corrected-to-normal vision. We obtained caregivers' informed consent for infants' participation. Families were compensated $50 for each visit.

Macaque infants

Subjects were 21 infant rhesus macaques (*Macaca mulatta*; 13 females and 8 males) and were tested longitudinally at the age of 2 weeks (11-15 days, $M_{age} = 12.83$, SD = 1.27; $N = 12$), 3 weeks (21-25 days, $M_{age} = 22.80$, SD = 1.26; $N = 15$), and 6 months (150-199 days, $M_{age} = 177.35$, SD = 15.18; $N = 26$). Animals were housed at the Laboratory of Comparative Ethology, National Institutes of Health Animal Center, Poolesville, Maryland. All infants were separated from their mothers on the day they were born (typically by 8am), and were reared in a nursery facility for ongoing, unrelated research studies. All infants were given inanimate cloth-covered surrogates, along with daily enrichment such as loose fleece squares, plastic toys, forage balls, and climbing chains, and were socialized for a minimum of 2 h per day. Infants received LabDiet High Protein Monkey Diet (#5054) and daily food enrichment consisting of fruit, seeds, and nuts. Water was available ad libitum. See Simpson, Miller, et al. (2016a) for more details on rearing practices.

Video stimulus

The video stimulus (1280 × 720 pixels) was identical for human and macaque infants at all ages (see Video 1). The video stimulus is also available at https://osf.io/p9mwk/?view_only=a0800300342b44f883c95145d45b411c. The video consisted of a series of high-contrast white disks with orthogonal stripes, including one black stripe and one brightly colored stripe (blue, green, or yellow), which appeared one at a time on a black background. Each disk appeared for 2 s, then disappeared, with 1 s between each presentation (black screen only). The disks appeared at six predetermined locations, always in the same order (center, top left, bottom left, top right, bottom right, center), accompanied by rotations and various sound effects. Disks were 90 pixels in height (3.42° visual angle) and 98 pixels wide (3.73° visual angle). The center of each stimulus disk appeared at each of the six locations (*x* and *y* coordinates relative to the top left corner of 0,0 pixels) in the order: middle (641, 320); top left (320, 181); bottom left (320, 541); top right (959, 181); bottom right (959, 541); middle (641, 320). AOIs were created around each disk location (Fig. 2). In total, the video was 18 s long.

Procedure

Human infants

Eye movements were recorded via corneal reflection using a Tobii TX300 eye tracker, a remote 58.4 cm monitor (51 cm in width × 28 cm in height) with integrated eye-tracking technology with the resolution set at 1280 × 720 pixels and a sampling rate of 300 Hz. While most screen-based studies with infants use a dark testing room to limit distractions (Holmqvist et al., 2022), this was not possible because the eye-tracking system requires some illumination in the room to track gaze fixations (see Tobii Technology, 2017

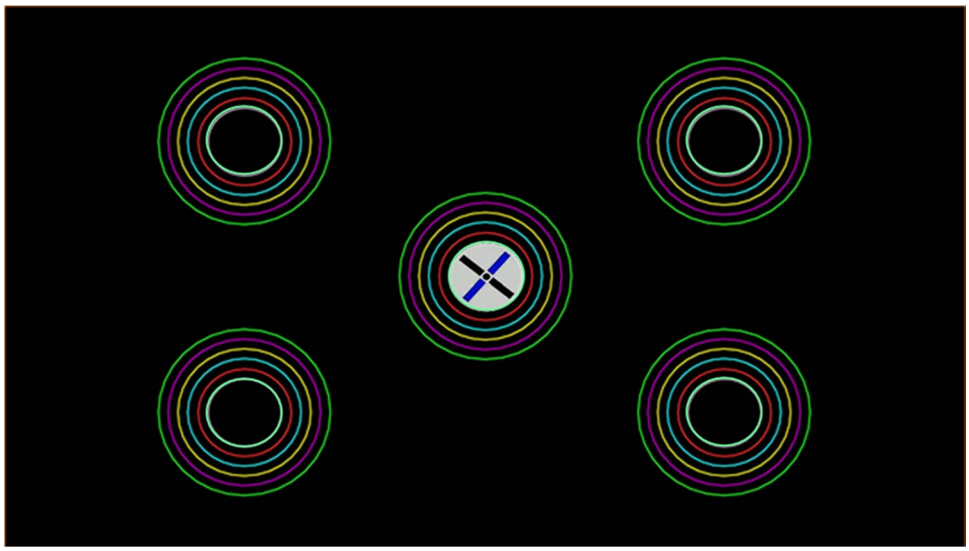

Fig. 2 Illustration of areas of interest (AOIs) on the video stimulus with various spatial enlargements (from innermost circle to outermost circle: 0°, 1°, 2°, 3°, 4°, and 5°) and AOI duration (temporal prolongation time: 0 ms, 200 ms, 400 ms, 600 ms, 800 ms, 100 ms). The smallest circle at each location matched the Disk Stimulus perfectly. A still image example of the Disk Stimulus is shown in the center AOI

for further details on how various room illuminations influence accuracy and precision in adults, which note better performance with greater lighting). We, therefore, decided to balance these trade-offs and test infants in a room where windows/direct sunlight was blocked with an illumination of 202 lux that was achieved with overhead lights. This lighting level is common among screen-based eye-tracking studies with infants and seems not to lower the eye-tracking data quality in young infants from the ideal eye-tracking illumination condition for adults (Katus et al., 2019; Tobii Technology, 2017).

Testing took place when the infants were awake, alert, and calm. Infants were seated in their parent's lap approximately 60 cm in front of the screen (Fig. 3A). Infants were calibrated using either a 5-point calibration (77 sessions) or a 9-point calibration (235 sessions) using Tobii Studio's preset locations, which presented a rattle cartoon that appeared at one location at a time (see Fig. 1 for calibration locations). Both eyes were calibrated simultaneously. The experimenter determined when the infant fixated at each calibration point (Hessels et al., 2015; Nyström et al., 2013). A calibration point (for each screen location and each eye) was registered when the infant fixated on it; individual calibration points that were not registered were repeated until we obtained an acceptable calibration (for calibration outcomes at each age, see Table 1). Infants varied in the

Fig. 3 Side-view of the experimental testing setup for A human infant on a caregiver's lap and B macaque infant held by an experimenter (from Maylott et al., 2020)

Table 1 Sample sizes, means, standard deviations, and range of the total number of registered calibration points in 5-point and 9-point calibration for human infants

Age (months)	5-Points per eye (total of 10)				
	N	Mean	SD	Range	Not calibrated
2	30	6.63	1.64	[3, 10]	6
4	26	7.96	1.75	[5, 10]	0
6	21	8.62	1.16	[6, 10]	0
	9-Points per eye (total of 18)				
	N	Mean	SD	Range	Not calibrated
2	49	12.98	3.40	[5, 18]	5
4	62	13.92	2.92	[7, 18]	2
6	62	15.02	2.49	[10, 18]	0
8	38	15.55	2.44	[10, 18]	0
14	24	15.67	3.02	[7, 18]	0

N = number of infants in each calibration. SD = standard deviation. No infants at 8 and 14 months were tested using the 5-point calibration. We started the current study with a 5-point calibration as suggested by previous studies (Gredebäck et al., 2009) and decided to transition fully to a 9-point method on the recommendation of a colleague, given that we, like our colleague, observed young infants (e.g., 2-month-olds) seemed to have an easier time shifting their fixations in the shorter distances in the 9-point calibration compared to the longer distances required of the 5-point calibration, which anecdotally appear to result in a faster and higher-quality calibration

duration of time required to obtain a calibration, ranging from 1 to 10 min. Some infants were calibrated successfully on the first attempt, while others required repeated attempts. Typically, the 2-month-olds took longer and more attempts to calibrate, and as infants grew older, calibration became easier and faster. Following the calibration, we showed the 18-s video stimulus.

Several infants could not be calibrated within 10 min or before they showed signs of being bored or fussy in some testing sessions (13 sessions); for this subset of sessions, a calibration from an infant of the same age was used instead (2 months old: six sessions of 5-point and five sessions of 9-point; 4 months old: two sessions of 9-point). Among these cases, three sessions (2 months old: two sessions; 4 months old: one session) were excluded from the subsequent analyses due to no fixations on the screen (see Fig. S2 for details). We detected no difference in the results with and without the data from these sessions using others' age-matched calibration profile (see Supplemental Materials), so we report the results with all available data.

Macaque infants

We recorded eye movements via corneal reflection using a Tobii TX300 eye tracker with the resolution set at 1280 × 720 and a sampling rate of 60 Hz. Infants were tested in a room where windows were blocked (no sunlight), and

Table 2 Sample sizes, means, standard deviations, and range of the total number of registered calibration points in 5-point calibration for macaque infants

Age	5-Points per eye (total of 10)			
	N	Mean	SD	Range
2 weeks	12	5.75	2.05	[3, 9]
3 weeks	15	6.60	1.55	[4, 9]
6 months	26	9.65	0.56	[8, 10]

N = number of infants in each calibration. SD = standard deviation. We completed data collection for macaque infants before (2014–2016) we started collection for human infants (2016–2019). Therefore, we used 5-point calibration for all macaque infants in line with previous primate eye-tracking studies (e.g., Kano et al., 2012; Paukner et al., 2013)

illumination of 250 lux was achieved by one overhead light (approximately 4 feet behind subject) and one additional light to the right of subjects. One experimenter stood in front of the eye tracker at a distance of approximately 60 cm from the screen and held each infant in her hands/arms wrapped in soft fleece fabric (Fig. 3B). Each infant was calibrated using a 5-point calibration procedure to Tobii Studio's preset locations; individual calibration points that were not registered were repeated until an acceptable calibration was obtained (Table 2). Both eyes were calibrated simultaneously. Infants varied in the duration of time required to obtain a calibration, ranging from 1 to 3 min. Some infants were calibrated successfully on the first attempt, while others required repeated attempts. Typically, the 2- and 3-week-old infants were more distracted and more difficult to calibrate than the 6-month-olds. However, if they could not be calibrated, we were able to attempt calibration at another time, later that day or the following day, until a usable calibration was obtained. Therefore, all macaque infants were calibrated successfully. Following the calibration, the 18-s video stimulus was shown.

Measures

Proportion of AOI hits

We drew AOIs over the target disk and in concentric circles of 1, 2, 3, 4, and 5° of visual angle larger than the disk (AOI size; see Fig. 2). These AOI sizes were designed to match the range of spatial deviations of infant eye-tracking data (Dalrymple et al., 2018; De Kloe et al., 2022; Morgante et al., 2012). The AOIs were activated when the disk appeared at that location and inactivated at 0, 200, 400, 600, 800, and 1000 ms after the disk disappeared (AOI duration).

We used the I-VT fixation filter in Tobii Studio software (Tobii Technology, Danderyd, Sweden), which defined fixations by a velocity threshold of 30°/s. Moreover, the I-VT filter discards short fixations with a minimum duration of 100 ms and merges adjacent fixations with a maximum time gap of 75 ms and a maximum angle of 0.5° (Olsen & Matos, 2012). We choose to use the I-VT filter because it is easy to use, one of the most common, and is robust to noisy data from infants, with the options to handle brief gaps in gaze signals, loss of one eye, and short fixations (Wass et al., 2013). We extracted the number of samples that were classified as fixations and located within the AOIs at each spatial and temporal manipulation (i.e., AOI hits), as well as the number of samples that were classified as fixations and located anywhere else on the screen during each AOI activation. We calculated the proportion of AOI hits by computing the number of fixation samples mapped onto the AOI divided by the number of fixation samples on the screen for each combination of spatial and temporal manipulation of the AOI. Therefore, there were a total of 36 proportions of AOI hits: 6 AOI sizes (0, 1, 2, 3, 4, 5° of visual angles over the disk) × 6 AOI durations (0, 200, 400, 600, 800, 1000 ms after disk disappearance). We used proportions instead of the raw fixation frequency to measure fixation-AOI mapping because we wanted to measure infants' fixations to the AOI out of their total fixation to the entire screen more generally. Infants may appear to look outside of AOIs for various reasons (e.g., off-task, measurement error), so these off-target looks need to be taken into account when considering on-target hits. This approach also enabled us to compare across AOIs of various sizes, with larger AOIs being more likely to capture looks by chance alone.

Registered calibration points

Tobii Studio provided calibration feedback through a pop-up window (see Fig. S1) that reported the number of calibration points registered for each eye. We counted the number of registered calibration points for each test session as an index of calibration quality (Wilkinson & Mitchell, 2014). The 5-point and 9-point calibrations provided a maximum of 10 (5 for each eye) or 18 (9 for each eye) registered points, respectively.

Data exclusion

We excluded 16 test sessions from human infants without any fixations on the screen (2 months: eight sessions; 4 months: six sessions; 6 months: one session; 8 months: 0 sessions; 14 months: one session) because no reliable data were provided, due to technical problems ($N = 1$), inattentiveness (i.e., no looking; $N = 8$), and fussiness/sleepiness (i.e., crying and/or eyes closed; $N = 7$). The final sample included 116 human infants (285 sessions in total) in the calibration analysis and 117 human infants (295 sessions in total) in the AOI analysis. See Fig. S2 for detailed exclusion procedures.

No data were excluded from macaque infants given that macaque infants for whom we could not obtain usable data—due to sleepiness, fussiness, inattentiveness, or other factors—were retested until usable data were obtained.

Data analysis

All statistical analyses were conducted using R (version 4.0.2) through RStudio (version 1.3.1073). We conducted multilevel linear mixed effects modeling to account for the nested structure of our data—multiple AOIs (level 1) were nested within multiple ages/visits (level 2), which were nested within individual infants (level 3). For model construction procedures for all analyses, we started with a baseline model including only a random intercept at the infant-level. Then we entered fixed effects and random variance into the models stepwise and selected the best-fitting models using likelihood ratio tests for model comparisons. All linear mixed effects models were conducted with the R packages "lme4" for model estimation (Batess et al., 2015) and "lmerTest" for significance tests of fixed-effects of the best-fitting models (Kuznetsova et al., 2017). Statistically significant interactions were examined with one-way repeated measures ANOVAs and pairwise t comparisons with Bonferroni corrections.

The R markdown for replicating data analyses and the data files for both species are available in Supplementary Materials and are also available at https://osf.io/p9mwk/?view_only=a0800300342b44f883c95145d45b411c.

Results

Human infant data

Effect of calibration

We first examined whether the calibration method (5-point vs. 9-point) influenced the proportion of AOI hits (outcome variable) that we detected, averaging across all AOI sizes and durations. We focused only on the age groups who were calibrated using both approaches (2-, 4-, and 6-month-olds). Potential fixed effects of calibration methods and age, as well as random variance at the age-level, calibration-methods-level, and infant-level, were added into the model stepwise. We added the random variance at the age-level and calibration-method-level to examine possible effects sourced

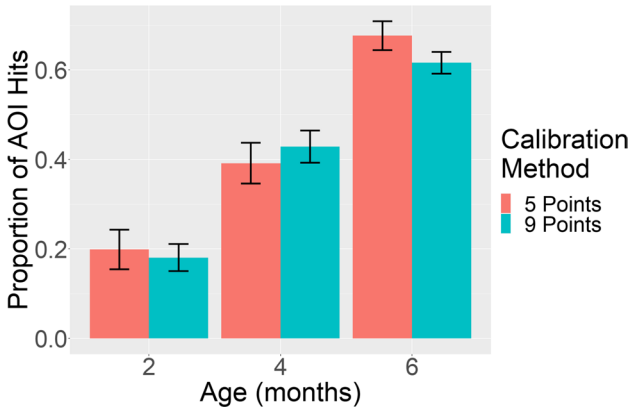

Fig. 4 Means and standard errors of the proportions of AOI hits averaging across AOI sizes and durations for 5-point (*red*) and 9-point (*blue*) calibration in 2-, 4-, and 6-month-old human infants

from uneven group sizes between calibration groups and among age groups (Milliren et al., 2018). We treated calibration methods and age as categorical variables. Moreover, the age factor was coded with repeated contrasts (2 vs. 4 months, 4 vs. 6 months).

The best-fitting model (m2; see Table S5 for model comparisons) included only a fixed main effect of age and a random intercept at the infant-level. That is, averaging across both calibration methods, infants had a higher proportion of AOI hits as they got older, $F(2, 159) = 83.28$, $p < 0.001$, $\eta_p^2 = 0.45$ (Fig. 4). Specifically, post hoc pairwise comparisons of the age effect showed that the proportion of AOI hits increased from 2 to 4 months of age, $t(169) = 6.60$, $p < 0.001$, $d = 1.02$, as well as from 4 to 6 months of age, $t(144) = 6.67$, $p < 0.001$, $d = 1.11$. However, the best-fitting model revealed no difference between 5- and 9-point calibration methods on the proportion of AOI hits averaging across all ages and we detected no interaction between age and calibration method.

We then expanded our analysis to all age groups (2–14 months old) and examined the relationship between the total number of registered calibration points and the proportion of AOI hits, and explored how this relationship changed with age using two multilevel regressions, one for each calibration method (5-point and 9-point calibrations). For both calibration methods, potential fixed effects of total number of registered calibration points (continuous) and age (categorical coded with repeated contrasts), as well as random variance at the age-level and infant-level, were added into the models stepwise.

The best-fitting models for both 5-point and 9-point calibration (m2 for both 5-point and 9-point; see Table S6 for model comparisons) included a significant fixed effect of total number of registered calibration points, which positively predicted the proportion of AOI hits averaging across ages (5-point: $b = 0.05$, $SE = 0.01$, $\beta = 0.64$, $t(68) = 3.68$, $p < 0.001$; 9-point: $b = 0.02$, $SE = 0.01$, $\beta = 0.30$, $t(216.7) = 3.02$, $p = 0.003$; Fig. 5A). There was also a main effect of age on the proportion of AOI hits (5-point: $F(2, 68) = 20.31$, $p < 0.001$, $\eta_p^2 = 0.37$; 9-point: $F(4, 181) = 31.44$, $p < 0.001$, $\eta_p^2 = 0.37$), suggesting that infants had a higher proportion of AOI hits as they got older (Fig. 5B). However, the best-fitting models revealed no interaction between the total number of registered calibration points and age. In sum, it appears that more registered calibration points are associated with a higher proportion of AOI hits.

Next, we explored which calibration method was associated with more registered calibration points. We sequentially added potential fixed effects of calibration methods (categorical: 5-point vs. 9-point) and age (categorical coded

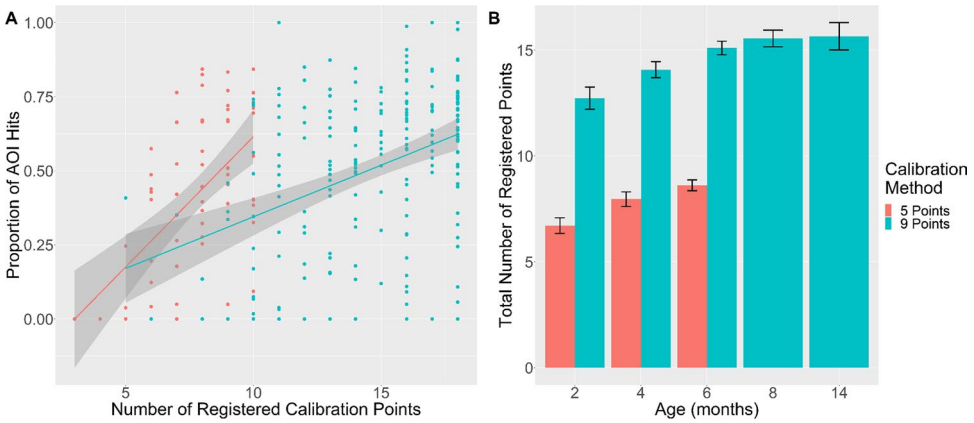

Fig. 5 A Scatter plot displaying the correlation between the number of registered calibration points and the proportion of AOI hits detected in the areas of interest, averaging across age, AOI sizes, and AOI durations, for infants using different calibration methods (*red*: 5-point calibration; *blue*: 9-point calibration). The *lines* indicate the regression line and the *shaded area* surrounding indicates standard error of the regression line. B Means and standard errors of the total number of registered calibration points for 5-point (*red*) and 9-point (*blue*) calibration in human infants at each age. The 5-point calibration was used only at 2, 4, and 6 months

with repeated contrasts), as well as random variance at the age-level, calibration-method-level, and infant-level, into the baseline model.

The best-fitting model (m2; see Table S7 for model comparisons) included fixed main effects of age and calibration methods, as well as a random intercept at the infant-level. We found that, averaging across ages, infants successfully registered more calibration points when using 9-point calibration procedure (M = 14.52, SD = 2.96) than when using 5-point calibration procedure (M = 7.78, SD = 1.74), $F(1, 190) = 268.71$, $p < 0.001$, $\eta_p^2 = 0.50$. This calibration method difference did not appear to change with age as the best-fitting model did not support an interaction between age calibration methods and age (Fig. 5B). Moreover, infants successfully registered more calibration points with age, $F(4, 243) = 10.71$, $p < 0.001$, $\eta_p^2 = 0.14$. Specifically, post-hoc pairwise comparisons of the age effect showed that the total number of registered points increased from 2 to 4 months of age, $t(243) = 2.99$, $p = 0.003$, $d = 0.38$, as well as from 4 to 6 months of age, $t(222) = 2.40$, $p = 0.016$, $d = 0.32$, but did not change from 6 to 8 months of age, $t(234) = 0.98$, $p = 0.324$, $d = 0.13$, nor from 8 to 14 months of age, $t(228) = 0.24$, $p = 0.809$, $d = 0.03$. In sum, 9-point calibrations registered more calibration points than 5-point calibrations, suggesting the former may confer an advantage.

Together, these results suggest that, while we detected no difference in fixation-AOI mapping between the 5- and 9-point calibrations, the eye tracker better captured valid fixation samples on the AOIs for infants who successfully registered more calibration points regardless of age, and since the 9-point calibrations registered more calibration points than 5-point calibrations, a 9-point calibration procedure may be advantageous in maximizing the number of registered calibration points thereby improving fixation-AOI mapping.

Effect of AOI size enlargement and duration prolongation

We examined whether spatial enlargement and temporal prolongation of the AOI improved the proportion of AOI hits (outcome variable). Potential fixed effects of AOI size (categorical: 0, 1, 2, 3, 4, 5° enlargement of the original AOI), AOI duration (categorical: 0, 200, 400, 600, 800, 1000 ms AOI prolongation after the disk disappeared), and age (categorical: 2, 4, 6, 8, 14 months of age), as well as random variance at the age-level and infant-level, were added into the model stepwise. We also added calibration methods (5-point vs. 9-point) as a control variable to account for potential differences due to calibration methods. The factors of AOI size, AOI duration, and age were coded with repeated contrasts (AOI size: 0 vs. 1°, 1 vs. 2°, 2 vs. 3°, 3 vs. 4°, 4 vs. 5° enlargement; AOI duration: 0 vs. 200 ms, 200 vs. 400 ms, 400 vs. 600 ms, 600 vs. 800 ms, 800 vs. 1000 ms; age: 2 vs. 4 months, 4 vs. 6 months, 6 vs. 8 months, 8 vs. 14 months).

The best-fitting model (m7; see Table S8 for model comparisons) for human infants revealed a main effect of age, $F(4, 61) = 75.90$, $p < 0.001$, $\eta_p^2 = 0.03$. Hence, there was a greater proportion of AOI hits as infants aged. We also found main effects of AOI size, $F(5, 10241) = 1117.26$, $p < 0.001$, $\eta_p^2 = 0.35$, and AOI duration, $F(5, 10241) = 166.59$, $p < 0.001$, $\eta_p^2 = 0.07$. Both spatial enlargement of AOIs and temporal prolongation of AOIs improved the proportion of AOI hits. However, the best-fitting model did not include an AOI size × duration interaction, nor an AOI size × AOI duration × age interaction. We did, however, detect an AOI size × age interaction, $F(20, 10241) = 23.51$, $p < 0.001$, $\eta_p^2 = 0.04$, as well as an AOI duration × age interaction, $F(20, 10241) = 7.47$, $p < 0.001$, $\eta_p^2 = 0.01$. We explore each of these interactions in the following sections.

AOI size effects at each age To explore the statistically significant AOI size × age interaction effect, we conducted a follow-up one-way ANOVA to test for the main effect of AOI size at each age. The AOI size main effect was statistically significant at each of the ages, ps < 0.001 (Table 3), suggesting that, regardless of age, increasing AOI size increased the proportion of AOI hits.

We evaluated the AOI size effect with multiple post-hoc pairwise comparisons between consecutive levels of spatial enlargement of AOI (repeated contrast coding) separately at each age (Table 4). As shown in Figs. 6 and 7, at 2 months, each degree of spatial enlargement increased the proportion of AOI hits. At 4 months, spatial enlargement of the AOI up to 4° increased the proportion of AOI hits. At 6 months, each degree of spatial enlargement increased the proportion of AOI hits. At 8 months, spatial enlargement up to 4° increased the proportion of AOI hits. At 14 months, spatial enlargement up to 2° improved the proportion of AOI hits. Furthermore, as the infants aged, their fixations became increasingly concentrated around the target disk.

AOI duration effect at each age We explored the statistically significant AOI duration × age interaction effect with five follow-up one-way ANOVAs, one at each age, which all revealed main effects of AOI duration, ps < 0.001 (Table 3).

We evaluated the temporal effect with multiple post hoc pairwise comparisons between consecutive levels of AOI duration (temporal prolongation of AOI; repeated contrast coding) separately at each age (Table 5). As shown in Figs. 7 and 8, at 2 months, averaging across all spatial enlargements, temporal prolongation of the AOI after the disk disappearance did not appear to increase the proportion of AOI hits. At 4 months, temporal prolongation up to 800 ms after the disk disappeared increased the proportion of AOI hits. At 6 months, temporal prolongation up to 400 ms after the disk disappeared increased the proportion of AOI hits. At 8 and 14 months, temporal prolongation did not appear to increase the proportion of AOI hits.

Table 3 Post hoc ANOVAs of AOI size and AOI duration effects on the proportion of AOI hits at each age in human infants

Age	Effect	df_B	df_W	F	p	η_p^2
2 months	AOI size	5	2451.00	173.30	< 0.001*	0.26
	AOI duration	5	2451.14	17.74	< 0.001*	0.03
4 months	AOI size	5	2823.93	252.86	< 0.001*	0.31
	AOI duration	5	2824.09	101.81	< 0.001*	0.15
6 months	AOI size	5	2800.78	574.49	< 0.001*	0.51
	AOI duration	5	2800.92	74.16	< 0.001*	0.12
8 months	AOI size	5	1320.00	179.16	< 0.001*	0.40
	AOI duration	5	1320.00	24.82	< 0.001*	0.09
14 months	AOI size	5	795.00	196.46	< 0.001*	0.55
	AOI duration	5	795.00	14.36	< 0.001*	0.08

Experimental manipulations to areas of interest (AOI) for human infants: AOI size (spatial enlargement degree: 0°, 1°, 2°, 3°, 4°, and 5°) and AOI duration (temporal prolongation time: 0 ms, 200 ms, 400 ms, 600 ms, 800 ms, 1000 ms). Critical $\boldsymbol{\alpha}$ level was corrected with Bonferroni correction, adjusted $\boldsymbol{\alpha} = 0.05/5 = 0.01$. df_B = between-group degrees of freedom. df_W = within-group degrees of freedom. η_p^2 = partial eta squared. * $p < \boldsymbol{\alpha}_{adj}$ (0.01)

Table 4 Descriptive statistics and post hoc pairwise comparisons of AOI size effects on the proportion of AOI hits at each age in human infants averaging across all AOI durations

Age	AOI Size	Mean	SD	Comparison	df	t	p	d
2 months	0°	0.09	0.13					
	1°	0.15	0.19	1° vs. 0°	2456.00	10.20	< 0.001	0.41*
	2°	0.18	0.21	2° vs. 1°	2456.00	5.05	< 0.001	0.20*
	3°	0.20	0.23	3° vs. 2°	2456.00	3.76	< 0.001	0.15*
	4°	0.22	0.24	4° vs. 3°	2456.00	3.34	0.001	0.13*
	5°	0.23	0.24	5° vs. 4°	2456.00	2.72	0.007	0.11*
4 months	0°	0.29	0.25					
	1°	0.36	0.27	1° vs. 0°	2828.92	9.24	< 0.001	0.35*
	2°	0.41	0.28	2° vs. 1°	2828.92	5.63	< 0.001	0.21*
	3°	0.46	0.29	3° vs. 2°	2828.92	6.44	< 0.001	0.24*
	4°	0.49	0.29	4° vs. 3°	2828.92	3.74	< 0.001	0.14*
	5°	0.51	0.29	5° vs. 4°	2828.92	2.08	0.037	0.08
6 months	0°	0.44	0.24					
	1°	0.58	0.23	1° vs. 0°	2805.77	19.06	< 0.001	0.72*
	2°	0.65	0.20	2° vs. 1°	2805.77	10.32	< 0.001	0.39*
	3°	0.68	0.19	3° vs. 2°	2805.77	5.56	< 0.001	0.21*
	4°	0.72	0.18	4° vs. 3°	2805.77	4.89	< 0.001	0.18*
	5°	0.74	0.17	5° vs. 4°	2805.77	2.90	0.004	0.11*
8 months	0°	0.52	0.29					
	1°	0.62	0.25	1° vs. 0°	1325.00	9.80	< 0.001	0.54*
	2°	0.67	0.24	2° vs. 1°	1325.00	5.24	< 0.001	0.29*
	3°	0.71	0.23	3° vs. 2°	1325.00	3.95	< 0.001	0.22*
	4°	0.76	0.19	4° vs. 3°	1325.00	4.05	< 0.001	0.22*
	5°	0.76	0.18	5° vs. 4°	1325.00	0.83	0.405	0.05
14 months	0°	0.46	0.24					
	1°	0.60	0.23	1° vs. 0°	800.00	11.19	< 0.001	0.79*
	2°	0.71	0.16	2° vs. 1°	800.00	8.58	< 0.001	0.61*
	3°	0.74	0.14	3° vs. 2°	800.00	2.05	0.040	0.14
	4°	0.75	0.14	4° vs. 3°	800.00	1.16	0.246	0.08
	5°	0.78	0.13	5° vs. 4°	800.00	2.04	0.041	0.14

SD = standard deviation. * $p < \boldsymbol{\alpha}$ adj (0.01)

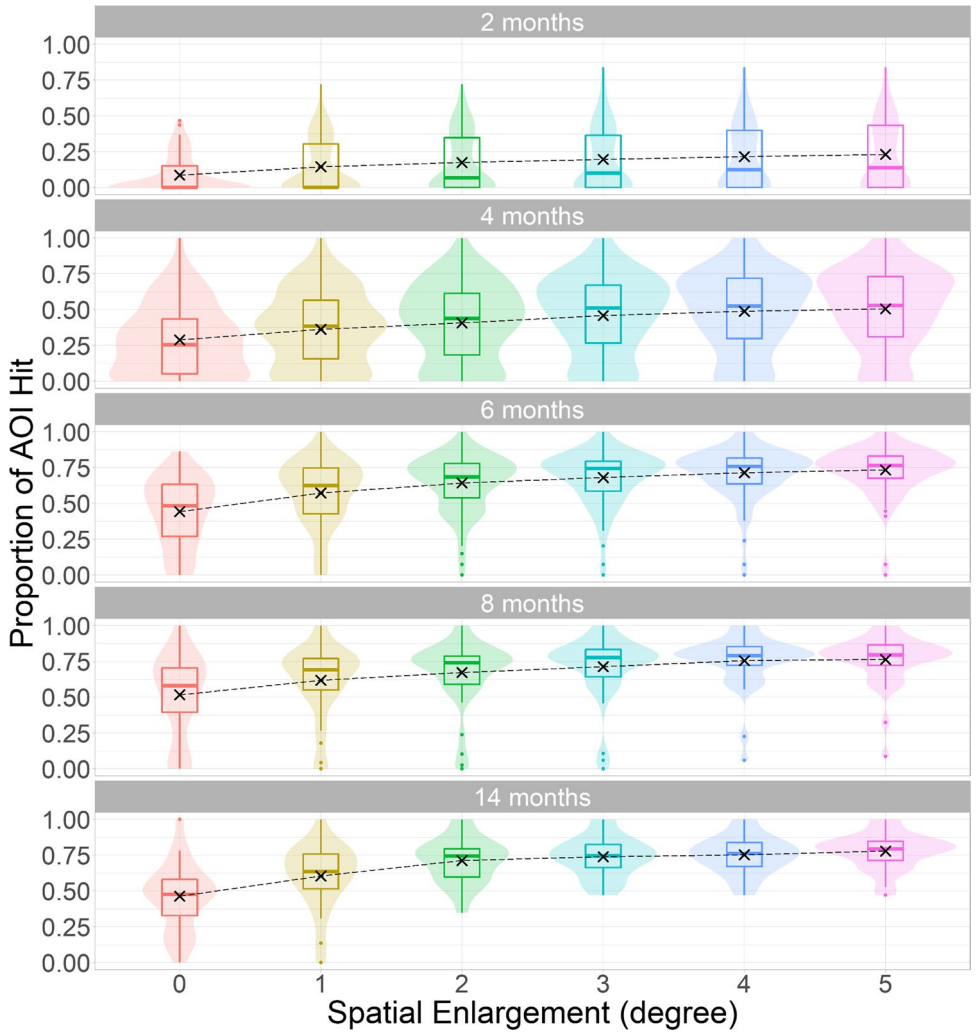

Fig. 6 Effect of AOI size (spatial enlargement of AOI) on the proportion of AOI hits, averaging across AOI duration (temporal prolongation of AOI) at each age, in human infants, from 2 months of age (*top*) to 14 months of age (*bottom*). Boxplots: *Horizontal lines* within the boxplots indicate the medians. The *hinges* of the boxplots show the first (*bottom*) and third (*top*) quartiles. The *whiskers* extend up to 1.5 × interquartile range (IQR; distance between top and bottom hinges), above and below the hinges. The *violin plots* show the distribution of the AOI hits. The black "*X*" indicates the means

Macaque infant data

Effect of calibration

We examined the relationship between the total number of registered calibration points and the proportion of AOI hits in macaque infants. Potential fixed effects of total number of registered calibration points (continuous) and age (categorical coded with repeated contrasts), as well as random variance at the age-level and infant-level, were added into the models stepwise.

The best-fitting model (m3; see Table S9 for model comparisons) included fixed main effects of age, registered points, as well as the registered-point × age interaction. The best-fitting model also included a random intercept at the infant-level. The main effect of age ($F(2, 53) = 27.35$, $p < 0.001$, $\eta_p^2 = 0.51$) suggested that the proportion of AOI hits increased as macaque infants got older. Specifically, post hoc pairwise comparisons of the age effect showed that the proportion of AOI hits increased from 3 weeks to 6 months of age, $t(47) = 4.86$, $p < 0.001$, $d = 1.42$, but not from 2 to 3 weeks of age, $t(47) = 1.65$, $p = 0.086$, $d = 0.48$. Moreover, the main effect of registered points ($b = 0.11$, $SE = 0.03$, $\beta = 1.21$, $t(53) = 3.73$, $p < 0.001$) suggested that, there was a statistically significant positive effect of the total number of registered calibration points on the proportion of AOI hits detected averaging across ages. Furthermore, the registered-point × age interaction ($F(2, 53) = 5.28$, $p = 0.008$, $\eta_p^2 = 0.17$) revealed that the effect of registered points on proportion of AOI hits was more prominent as the macaque infants

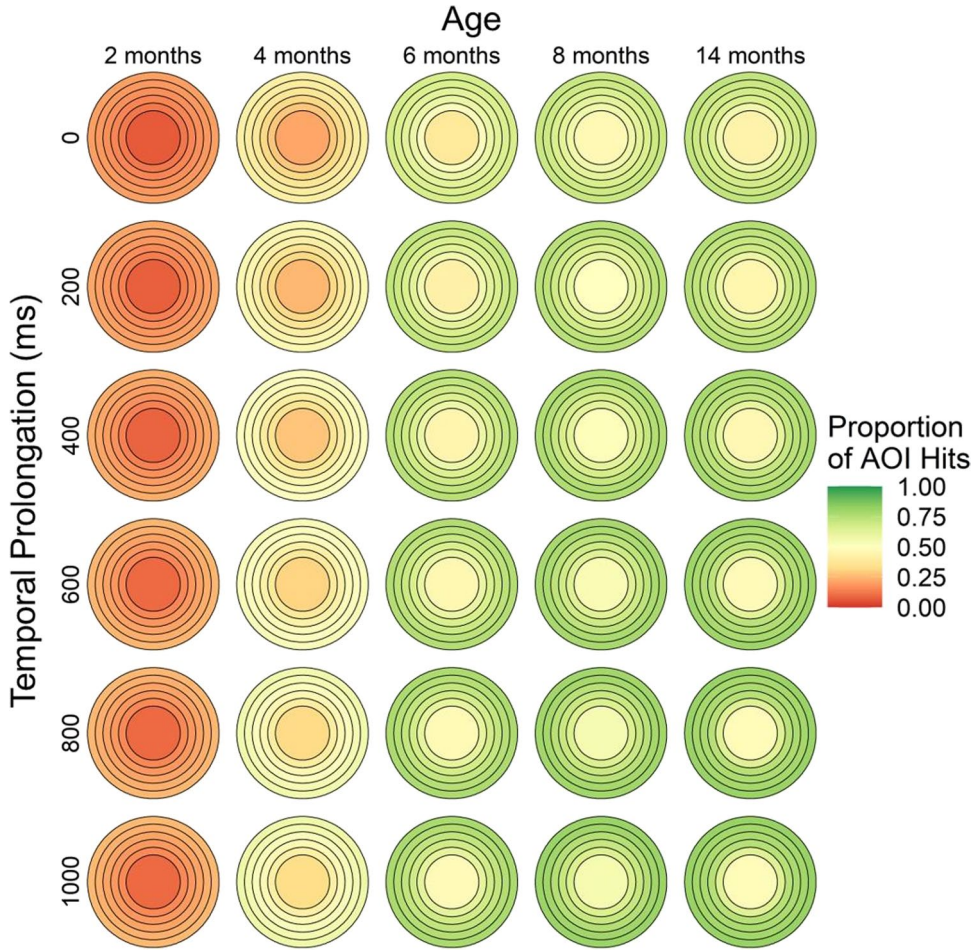

Fig. 7 Effect of AOI size (spatial enlargement of AOI), reflected by the *concentric circles* (from innermost circle to outermost: 0°, 1°, 2°, 3°, 4°, and 5°) on the proportion of AOI hits out of the total number of hits on the screen, for 0-ms (*top*) to 1000-ms temporal prolongation (*bottom*), at each age in human infants, from 2 months of age (*leftmost*) to 14 months of age (*rightmost*), averaging across the 6 target disk locations. *Color shading* represents the cumulative proportion of AOI hits in an AOI with a corresponding spatial enlargement and temporal prolongation (*dark red* = 0, *light yellow* = 0.5, *dark green* = 1.0). The outer circles contain the inner circles, so if the proportion of AOI hits increases as the AOIs grow larger, this change reflects the larger AOIs capturing a greater proportion of AOI hits

aged (Fig. 9A). Specifically, the positive association between registered points and proportion of AOI hits became stronger from 3 weeks to 6 months of age, $b = 0.26$, $SE = 0.09$, $\beta = 0.59$, $t(53) = 2.98$, $p = 0.004$, but did not change from 2 to 3 weeks, $b = 0.01$, $SE = 0.05$, $\beta = 0.03$, $t(53) = 0.22$, $p = 0.824$.

In addition, we examined the association between age and the total number of registered points in macaque infants. We found that, as the macaque infants aged, they successfully registered more calibration points, $F(2, 39) = 55.11$, $p < 0.001$, $\eta_p^2 = 0.72$ (Fig. 9B). Specifically, the total number of registered points increased from 3 weeks to 6 months of age, $t(38) = 7.88$, $p < 0.001$, $d = 2.55$, but not from 2 to 3 weeks of age, $t(34) = 1.70$, $p = 0.089$, $d = 0.58$. Therefore, our eye tracker could detect a higher proportion of AOI hits for macaque infants with more points calibrated and this effect became stronger with age.

Effect of AOI size and duration prolongation

We examined whether spatial enlargement and temporal prolongation of AOIs improved the proportion of AOI hits (outcome variable) in macaque infants. Potential fixed effects of AOI size (categorical: 0, 1, 2, 3, 4, 5° enlargement of the original AOI; coded with repeated contrasts), AOI duration (categorical: 0, 200, 400, 600, 800, 1000 ms AOI prolongation after the disk disappears; coded with repeated contrasts), and age (categorical: 2 weeks, 3 weeks, 6 months of age; coded with repeated contrasts), as well as random variance at the age-level and infant-level, were added into the model stepwise.

The best-fitting model (m5; see Table S10 for model comparisons) revealed a main effect of age, $F(2, 16) = 20.38$, $p < 0.001$, $\eta_p^2 = 0.02$. There was a higher proportion of AOI hits as infants aged. There was also a main effect of AOI size, $F(5, 1860) = 83.17$, $p < 0.001$, $\eta_p^2 = 0.18$,

Table 5 Descriptive statistics and post hoc pairwise comparisons of AOI duration effect on the proportion of AOI hits at each age in human infants averaging across all AOI sizes

Age	AOI duration	Mean	SD	Comparison	*df*	*t*	*p*	*d*
2 months	0 ms	0.15	0.20					
	200 ms	0.16	0.21	200 vs. 0	2456.00	1.94	0.052	0.08
	400 ms	0.17	0.21	400 vs. 200	2456.00	1.59	0.112	0.06
	600 ms	0.19	0.22	600 vs. 400	2456.00	1.89	0.058	0.08
	800 ms	0.19	0.22	800 vs. 600	2456.35	0.73	0.466	0.03
	1000 ms	0.19	0.23	1000 vs. 800	2456.00	0.26	0.798	0.01
4 months	0 ms	0.34	0.28					
	200 ms	0.38	0.28	200 vs. 0	2828.91	4.58	< 0.001	0.17*
	400 ms	0.41	0.28	400 vs. 200	2828.91	3.90	< 0.001	0.15*
	600 ms	0.44	0.29	600 vs. 400	2829.29	2.73	0.006	0.10*
	800 ms	0.47	0.29	800 vs. 600	2828.91	3.18	0.001	0.12*
	1000 ms	0.48	0.28	1000 vs. 800	2828.91	1.13	0.258	0.04
6 months	0 ms	0.57	0.22					
	200 ms	0.61	0.22	200 vs. 0	2805.71	3.88	< 0.001	0.15*
	400 ms	0.64	0.22	400 vs. 200	2805.56	2.85	0.004	0.11*
	600 ms	0.65	0.23	600 vs. 400	2805.95	2.21	0.027	0.08
	800 ms	0.67	0.22	800 vs. 600	2805.56	1.64	0.101	0.06
	1000 ms	0.67	0.22	1000 vs. 800	2805.56	0.54	0.586	0.02
8 months	0 ms	0.62	0.25					
	200 ms	0.65	0.25	200 vs. 0	1325.00	2.52	0.012	0.14
	400 ms	0.67	0.25	400 vs. 200	1325.00	1.74	0.081	0.10
	600 ms	0.69	0.25	600 vs. 400	1325.00	1.38	0.168	0.08
	800 ms	0.70	0.25	800 vs. 600	1325.00	0.97	0.330	0.05
	1000 ms	0.71	0.23	1000 vs. 800	1325.00	0.58	0.559	0.03
14 months	0 ms	0.62	0.22					
	200 ms	0.65	0.21	200 vs. 0	800.00	1.40	0.160	0.10
	400 ms	0.67	0.21	400 vs. 200	800.00	1.33	0.182	0.09
	600 ms	0.69	0.21	600 vs. 400	800.00	1.09	0.277	0.08
	800 ms	0.70	0.21	800 vs. 600	800.00	0.57	0.569	0.04
	1000 ms	0.71	0.21	1000 vs. 800	800.00	0.08	0.933	0.01

SD = standard deviation. * $p < \alpha$ adj (0.01)

and a main effect of AOI duration, $F(5, 1860) = 4.42$, $p = 0.001$, $\eta_p^2 = 0.01$. The best-fitting model did not include an AOI size × AOI duration interaction or an AOI size × AOI duration × age interaction. We did, however, detect an AOI size × age interaction, $F(10, 1860) = 11.96$, $p < 0.001$, $\eta_p^2 = 0.06$, and an AOI duration × age interaction, $F(10, 1860) = 2.84$, $p = 0.002$, , $\eta_p^2 = 0.02$, each explored below.

AOI size effect at each age To explore the statistically significant AOI size × age interaction effect, we conducted a follow-up one-way ANOVA at each age, which revealed a main effect of AOI size at each age, $ps < .001$ (Table 6).

We evaluated the AOI size effect with post-hoc pairwise comparisons between consecutive levels of spatial enlargement of AOI (repeated contrast coding) within each age (Table 7). As shown in Figs. 10 and 11, at 2 weeks, spatial enlargement of 1° larger than the target disk and enlargement from 4° to 5° larger than the disk both increased the proportion of AOI hits. At 3 weeks, spatial enlargement of the AOI from 1° to 2° larger than the disk increased the proportion of AOI hits. At 6 months, spatial enlargement up to 2° larger than the target disk increased the proportions of AOI hits. Furthermore, as the macaque infants aged, their fixations became increasingly concentrated around the target disk. Notably, among 2- and 3-week-olds, the medians were close to zero, suggesting that either the macaque infants were not looking, or the eye tracker was unable to capture gaze signals from some of these very young macaques.

AOI duration effect at each age We explored the statistically significant AOI duration × age interaction, with three

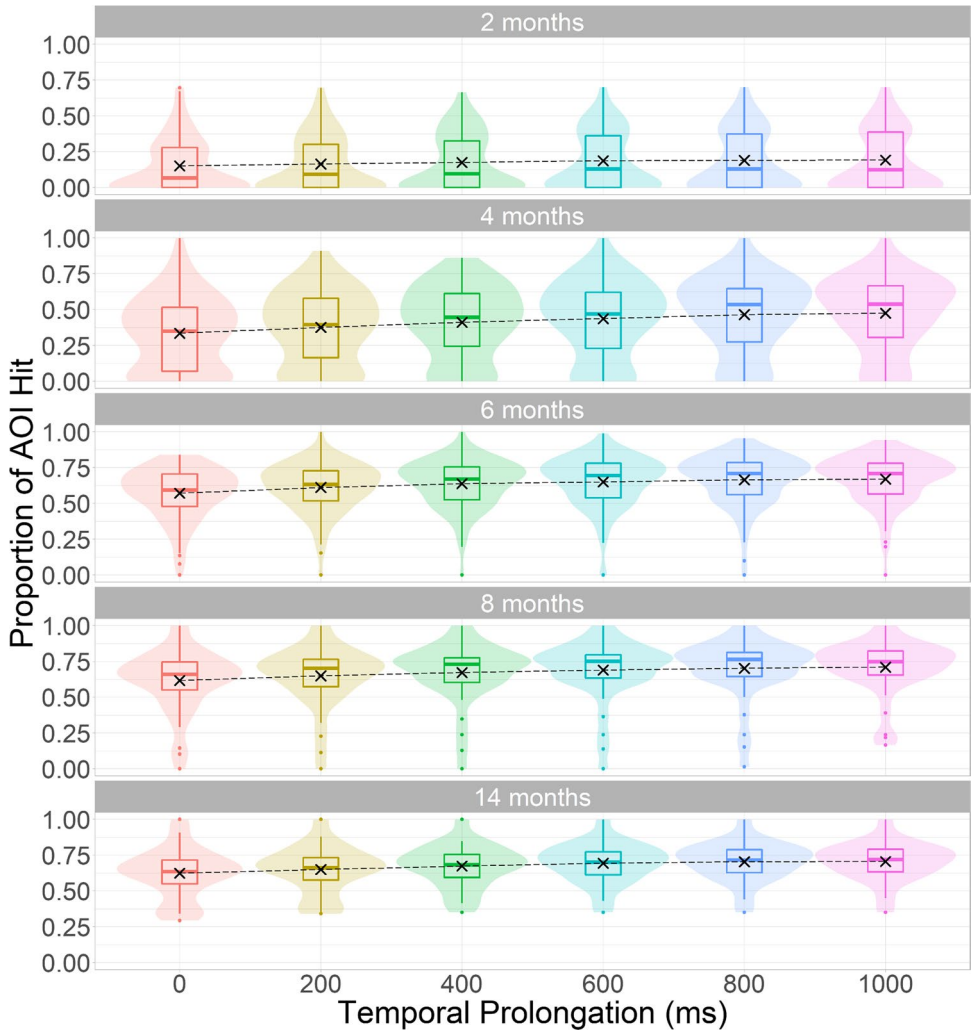

Fig. 8 Effect of AOI duration (temporal prolongation of AOI) on proportion of AOI hits, averaging across AOI size (spatial enlargement of AOI) at each age, in human infants, from 2 months of age (*top*) to 14 months of age (*bottom*). Boxplots: *Horizontal lines* within the boxplots indicate the medians. The *hinges* of the boxplots show the first (*bottom*) and third (*top*) quartiles. The *whiskers* extend up to 1.5 × interquartile range (IQR; distance between top and bottom hinges), above and below the hinges. The *violin plots* show the distribution of the AOI hits. The black "*X*" indicates the means

follow-up one-way ANOVAs, one at each age (critical α level was corrected with Bonferroni correction, adjusted $\alpha = 0.05/3 = 0.017$). We detected a main effect of AOI duration only at 6 months, $p < 0.001$ (Table 6). Therefore, temporal prolongation only appeared to increase the proportion of AOI hits in the oldest age group for macaque infants. Post hoc pairwise comparisons between consecutive levels of temporal prolongation of AOI duration (repeated contrast coding) revealed that, at 6 months, AOI temporal prolongation from 0 to 200 ms after the disk disappeared increased the proportion of AOI hits (Table 8; Figs. 11 and 12). There were no other statistically significant effects, $ps > 0.05$.

Discussion

Remote eye tracking is increasingly used in developmental research involving human and primate infants given its non-invasive procedures and ability to quickly produce a large amount of data (Aslin & McMurray, 2004; Hopper et al., 2021). However, many questions remain about the best methods to maximize the quality of these data. Researchers must make a variety of methodological choices when designing eye-tracking studies, which can be particularly difficult with these populations—especially when comparing infants of differing ages and species—given that there are no empirically established guidelines (Holmqvist

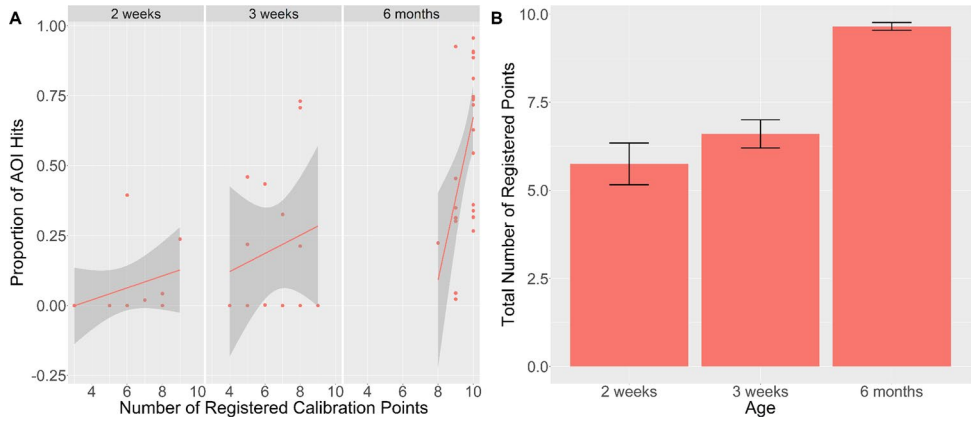

Fig. 9 **A** Scatter plot displaying the correlation between the number of registered calibration points and the proportion of AOI hits detected in the areas of interest, averaging across age, AOI sizes, and AOI durations, for macaque infants. The *lines* indicate the regression line and the *shaded area* surrounding indicates standard error of the regression line. **B** Means and standard errors of the total number of registered calibration points in macaque infants at each age. All macaque infants used 5-point calibration

et al., 2022). To begin to address these gaps, we explored how calibration methods (procedure and quality) and AOI characteristics (sizes and durations) influence the fixation-AOI mappings in human infants (2- to 14-month-old) and macaque infants (2-week-old to 6-month-old) tested longitudinally using a Tobii TX300 eye tracker. We found that a greater number of registered calibration points was associated with a greater proportion of AOI hits, suggesting there may be advantages of using a built-in Tobii 9-point calibration over a 5-point calibration. Moreover, we discovered that enlarging and prolonging AOIs increased the proportion of AOI hits, suggesting larger and longer AOIs may be advantageous. Moreover, we found that these increases varied by age and species, suggesting that infant researchers need to consider their specific populations' characteristics to select the most appropriate study designs. We make recommendations for data inclusion/exclusion decisions to maximize participant retention without jeopardizing the quality of fixation-AOI mappings.

Tobii's built-in calibration: 5-point versus 9-point procedure

Calibration is necessary to account for individual characteristics of infants' eyes for better eye-tracking accuracy and precision (Gredebäck et al., 2009). We detected no differences in the proportions of AOI hits in human infants when using a 5-point compared to a 9-point calibration method, regardless of age. However, we discovered that, in both human and macaque infants, averaging across all age groups, the proportion of AOI hits captured increased as the total number of successfully registered calibration points increased, regardless of the calibration method used and the infants' ages. Admittedly, while these findings may be because better calibration improves subsequent fixation-AOI mappings, we cannot rule out the possibility that both better calibration quality and better fixation-AOI mappings are driven by infants' characteristics, such as their attentional and emotional states during testing. Regardless of which

Table 6 Post hoc ANOVAs of AOI size and AOI duration effects on proportion of AOI hits at each age in macaque infants

Age	Effect	df_B	df_W	F	p	η_p^2
2 weeks	AOI size	5	420	22.62	< 0.001*	0.212
	AOI duration	5	420	0.07	0.997	0.001
3 weeks	AOI size	5	525	37.87	< 0.001*	0.265
	AOI duration	5	525	0.18	0.970	0.002
6 months	AOI size	5	915	48.74	< 0.001*	0.210
	AOI duration	5	915	17.50	< 0.001*	0.087

Experimental manipulations to areas of interest (AOI) for macaque infants: AOI size (spatial enlargement degree: 0°, 1°, 2°, 3°, 4°, and 5°) and AOI duration (temporal prolongation time: 0 ms, 200 ms, 400 ms, 600 ms, 800 ms, 1000 ms). Critical α level was corrected with Bonferroni correction, adjusted $\alpha = 0.05/3 = 0.017$. df_B = between-group degrees of freedom. df_W = within-group degrees of freedom. η_p^2 = partial eta squared. * $p < \alpha_{adj}$ (0.017)

Table 7 Descriptive statistics and post hoc pairwise comparisons of AOI size effects on the proportion of AOI hits at each age in macaque infants averaging across all AOI durations

Age	AOI size	Mean	SD	Comparison	*df*	*t*	*p*	*d*
2 weeks	0°	0.01	0.05					
	1°	0.05	0.12	1° vs. 0°	415	4.02	< 0.001	0.39*
	2°	0.05	0.13	2° vs. 1°	415	0.40	0.689	0.04
	3°	0.06	0.13	3° vs. 2°	415	0.31	0.755	0.03
	4°	0.06	0.13	4° vs. 3°	415	0.02	0.982	0.00
	5°	0.11	0.19	5° vs. 4°	415	5.71	< 0.001	0.56*
3 weeks	0°	0.08	0.16					
	1°	0.10	0.16	1° vs. 0°	520	0.82	0.410	0.07
	2°	0.22	0.29	2° vs. 1°	520	5.28	< 0.001	0.46*
	3°	0.24	0.31	3° vs. 2°	520	0.79	0.426	0.07
	4°	0.28	0.38	4° vs. 3°	520	2.21	0.027	0.19
	5°	0.32	0.39	5° vs. 4°	520	1.52	0.128	0.13
6 months	0°	0.47	0.28					
	1°	0.55	0.29	1° vs. 0°	910	6.65	< 0.001	0.44*
	2°	0.58	0.31	2° vs. 1°	910	2.81	0.005	0.17*
	3°	0.60	0.31	3° vs. 2°	910	1.52	0.127	0.10
	4°	0.61	0.31	4° vs. 3°	910	0.96	0.334	0.06
	5°	0.62	0.30	5° vs. 4°	910	0.57	0.568	0.04

SD = standard deviation. * $p < \boldsymbol{\alpha}$ adj (0.01)

mechanism underlies the association between calibration and subsequent fixation-AOI mappings, the total number of registered calibration points could be used to set minimum standards of data acquisition and to assess the usability of data collected from each test session to determine if certain sessions should be excluded.

Furthermore, human infants registered more successful calibration points when using the 9-point method compared to the 5-point, suggesting that attempting a greater number of points may maximize the number of registered calibration points. We, therefore, recommend that, when testing infants with the built-in Tobii calibration procedures at these young ages, researchers consider using the 9-point calibration method, which is less demanding of young infants in terms of the distances and angles between each point. Another advantage of the calibration approach in the Tobii TX300 system is that, even if not all points are registered for each eye, researchers have the option to repeat just the specific points that have not yet been captured. While this process appears straightforward, our experience is that sometimes the calibration will fail altogether with the addition of newly attempted, but failed points (resulting in the screen depicted in Fig. S1B). That is, repeating calibrations to obtain more points is not without risk. Therefore, trying to achieve a "perfect" calibration in a young infant is not always realistic, especially if the infant appears to be growing fussy or disinterested.

Our analyses of calibration methods were limited in some regards. One limitation is that, given that these analyses were not planned prior to data collection and lacked systematic manipulation, we only were able to conduct them in human infants at the ages of 2, 4, and 6 months. Therefore, it is unclear whether older human infants and other species would show similar advantages of a 9-point calibration approach. Our findings, while preliminary, nonetheless offer insights into potential advantages of using a 9-point over a 5-point calibration approach, at least when testing very young human infants (aged 2 to 6 months).

Additionally, other aspects of calibration still need to be explored. For example, while the built-in calibration procedures (such as those in the Tobii TX300 system we used here) are easy to use, some customized software toolboxes offer more flexibility and control over the built-in procedures, which may facilitate better and easier calibration in human and primate infants (Niehorster et al., 2020). For example, calibration routines that use large stimuli to attract attention, and which subsequently shrink to a small target for actual calibration, may enable capturing infants' attention while also retaining high precision (Schlegelmilch & Wertz, 2019). With new approaches that enable greater flexibility in calibration procedures, future studies are encouraged to explore how different variations of calibration targets–types, locations, sounds, and movements–may affect fixation-AOI mapping in human and

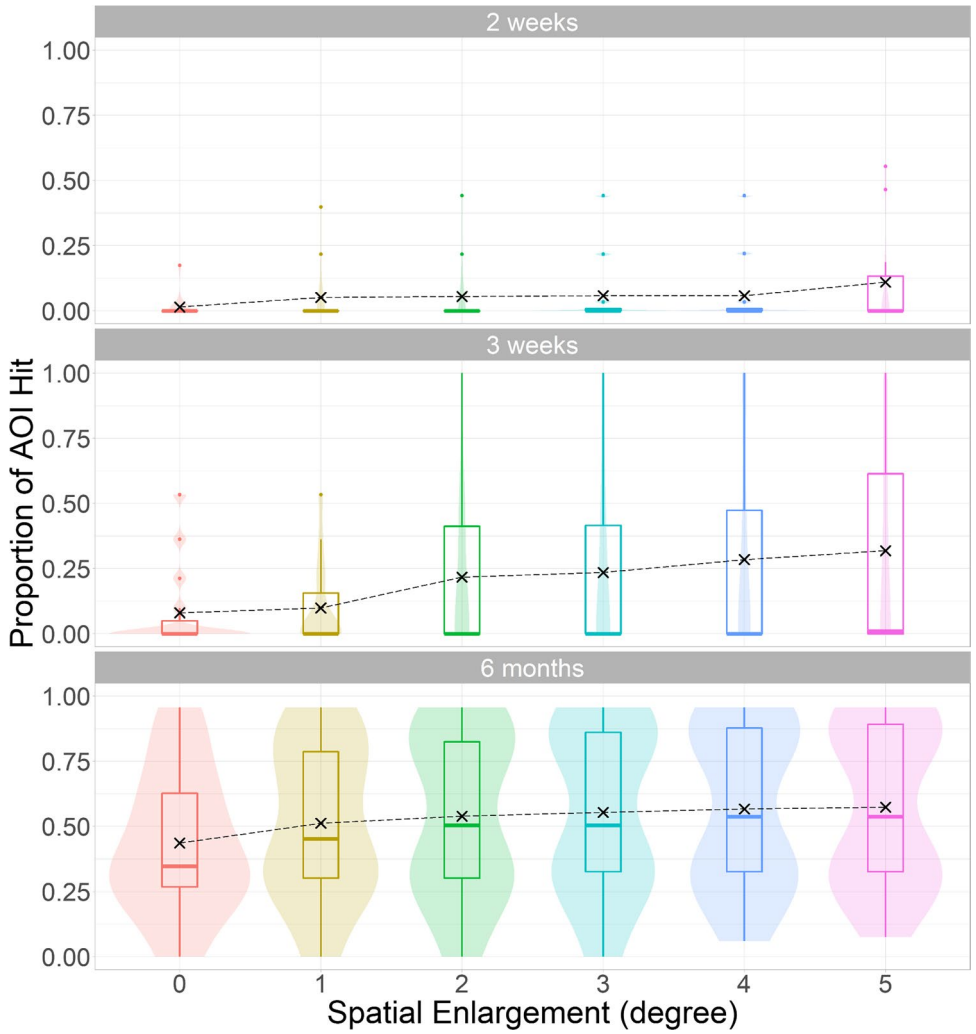

Fig. 10 Effect of AOI size (spatial enlargement of AOIs) in macaque infants averaging across AOI duration (temporal prolongation of AOI) at each age at 2 weeks (*top*), 3 weeks (*middle*), and 6 months (*bottom*) of age. Boxplots: *Horizontal lines* within the boxplots indicate the medians. The *hinges* of the boxplots show the first (*bottom*) and third (*top*) quartiles. The *whiskers* extend up to 1.5 × interquartile range (IQR; distance between top and bottom hinges), above and below the hinges. The *violin plots* show the distribution of the AOI hits. The black "*X*" indicates the means

primate infants, to further optimize calibration quality and thereby subsequent data quality.

Effect of AOI size and developmental changes

For infants of both species, enlarging the size of the AOIs resulted in a better ability to capture fixations around the target disk, but this effect differed across age and species. For human infants, we found an increase in the proportions of fixations captured by the AOIs with enlargement up to 5° between the age of 2 and 6 months, up to 4° at 8 months, but only up to 2° at 14 months. For macaque infants, increases in the mapping of fixations onto AOIs were found with an AOI enlargement up to 5° at the age of 2 weeks, and up to 2° between 3 weeks and 6 months. As previous studies reported spatial deviations from 1° to 5° in eye-tracking data in human infants from 3 to 30 months of age (Dalrymple et al., 2018; De Kloe et al., 2022; Morgante et al., 2012), our findings are consistent with such reports and extend them to a younger age of 2 months, as well as to infants of another primate species. Since fixations irrelevant to the target disk should spatially be distributed randomly on the screen and are unlikely to be located within a certain area around the disk, our findings of the increase in the proportions of AOI hits as the result of AOI size enlargements are likely driven by the spatial deviations of valid fixations rather than random noise. In fact, infants may not necessarily be focusing on the center of the stimulus, as adults can be instructed to do, and instead may focus on the high-contrast outer edge of the disk (Bronson, 1994; Johnson, 2019). Therefore, we

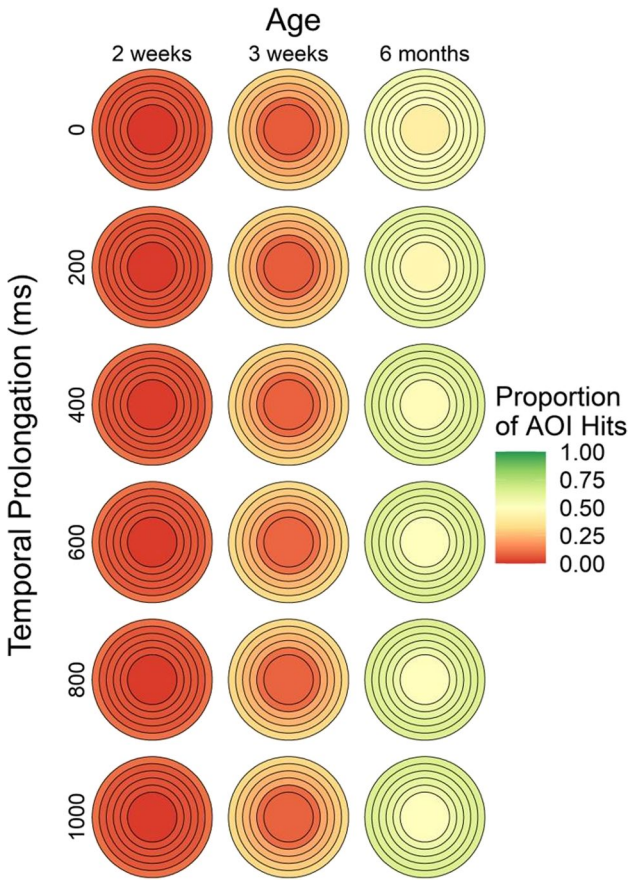

Fig. 11 Effect of AOI size (spatial enlargement of AOI), reflected by the *concentric circles* (from innermost circle to outermost: 0°, 1°, 2°, 3°, 4°, and 5°), on the proportion of AOI hits out of the total number of fixations on the screen, for 0-ms (*top*) to 1000-ms temporal prolongation (*bottom*), at each age in macaque infants, at 2 weeks (*leftmost*), 3 weeks (*middle*), and 6 months of age (*rightmost*), averaging across the 6 target disk locations. *Color shading* represents the cumulative proportion of AOI hits in an AOI with a corresponding spatial enlargement and temporal prolongation (*dark red* = 0, *light yellow* = 0.5, *dark green* = 1.0). The outer circles contain the inner circles, so if the proportion of AOI hits increases as the AOIs grow larger, this change reflects the larger AOIs capturing a greater proportion of AOI hits

recommend creating AOIs that are larger than the outer edges of stimuli for infant eye-tracking research. In sum, while an AOI that is 5° larger than the outer edge of the stimulus is likely to capture more fixation samples than random noise in human infants at 2, 4, and 6 months and macaque infants at 2 weeks, an AOI with the same size is likely to capture more noise than valid fixations as infants get older (e.g., 14-month-old human infants and 6-month-old macaque infants). These results suggest that AOI sizes need to be adjusted based on participants' age and species.

Further, our findings also provide insights for stimulus creation: For 2- to 6-month-old human infants and 2-week-old macaque infants, simultaneously presented stimuli need to be sufficiently spaced apart from each other to afford larger AOIs and to reduce the likelihood of capturing fixations on the wrong AOI. In other words, the distance between two stimuli (occurring simultaneously or in rapid succession) should be spaced far enough apart to afford enlarged and non-overlapping AOIs for each stimulus for infants at these young ages. However, for older infants–8- and 14-month-old human infants, as well as 3-week-old and 6-month-old macaque infants–eye-tracking studies may use stimuli that are closer to each other and may use smaller AOIs, capturing a greater degree of precision.

Our findings also indicate that, at older ages (14-month-old humans and 6-month-old macaques), both species showed a more condensed distribution of fixations around the target disk than they did at younger ages. Notably, these patterns are consistent with the overall age-related increases of fixation-AOI mappings we found in infants: in both human and macaque infants, the AOIs captured more fixations as infants grew older. Such age-related increases in capturing infants' fixations may, in part, be related to the rapid development in infants' visual and attentional systems across these ages for both species (Chandna, 1991; Dobson & Teller, 1978; Ordy et al., 1964; Richards, 2004; Teller, 1981; Xiang et al., 2021). Human and primate infants' visual acuity, tracking ability, and sustained attention undergo rapid development in their first year after birth (Maylott et al., 2020; Phillips et al., 2007; Teller, 1981; Von Hofsten & Rosander, 1997). In sum, older infants may be easier to capture eye gaze from than younger infants due to improvements in infants' visual and attentional abilities with age.

However, another likely factor contributing to the apparent age-related increase in fixation-AOI mappings is that the eye-tracking system can better detect the eyes and gaze locations of older compared to younger infants (Hessels & Hooge, 2019; Hopper et al., 2021; Wass et al., 2013). That is, there may be more error, noise, and data loss when using this eye-tracking system with very young infants due to limitations with the system itself (e.g., difficulty in identifying pupils of young infants; Wass et al., 2014). If so, these apparent age-related improvements may, at least in part, reflect enhanced measurement precision and accuracy in older infants (i.e., that the Tobii TX300 has a better ability to capture older infants' fixations for both species). While this hypothesis has yet to be empirically tested—which would require, for example, behavioral coding of infants' attention frame-by-frame from video and comparing to eye-tracking data—this interpretation is consistent with a report in human infants that with age, between 5 and 10 months, spatial accuracy increases and data loss decreases using a Tobii TX300 (De Kloe et al., 2022). Thus, older infants, compared to younger infants, may provide eye-tracking data that are more stable, smooth, and have less noise.

Table 8 Descriptive statistics and post hoc pairwise comparisons of AOI duration effects on the proportion of AOI hits in macaque infants averaging across all AOI sizes

Age	AOI duration	Mean	SD	Comparison	*df*	*t*	*p*	*d*
2 weeks	0 ms	0.06	0.13	-				
	200 ms	0.06	0.13	-				
	400 ms	0.06	0.14	-				
	600 ms	0.06	0.14	-				
	800 ms	0.06	0.14	-				
	1000 ms	0.06	0.14	-				
3 weeks	0 ms	0.20	0.31	-				
	200 ms	0.20	0.31	-				
	400 ms	0.20	0.31	-				
	600 ms	0.21	0.32	-				
	800 ms	0.21	0.31	-				
	1000 ms	0.21	0.31	-				
6 months	0 ms	0.51	0.37	-				
	200 ms	0.56	0.34	200 vs. 0	910	3.43	0.001	0.23*
	400 ms	0.59	0.31	400 vs. 200	910	2.41	0.016	0.16
	600 ms	0.60	0.28	600 vs. 400	910	0.77	0.440	0.04
	800 ms	0.59	0.26	800 vs. 600	910	-0.22	0.826	0.01
	1000 ms	0.59	0.24	1000 vs. 800	910	0.00	0.997	0.00

Descriptive statistics for AOI durations are reported for all ages. Follow-up comparisons were only conducted at 6 months for macaque infants given that the one-way ANOVA showed significant Temporal effect at 6 months, but not at 2 weeks or 3 weeks of age. SD = standard deviation, * $p < \alpha_{adj}$ (0.01)

These two potential interpretations of age-related improvements in fixation-AOI mapping in infant eye tracking—that there are both qualities of the infants, as well as limitations of the eye-tracking measurement system—are not mutually exclusive, and regardless of which may play a bigger role, both suggest that some methodological adjustments, such as using larger and longer duration AOIs, may be useful to increase data capture.

Effect of AOI duration and its developmental changes

Prolonging the time window of the AOIs also improved fixation-AOI mapping for human and macaque infants, but it did so differently across age and species. In humans, AOI duration prolongation increased the proportion of AOI hits when it was extended up to 800 ms at 4 months and up to 400 ms at 6 months, while in macaques, prolongation of up to 200 ms improved AOI hits only at 6 months, suggesting that, at particular ages, infants of both species tended to begin to fixate on the disk locations only after the stimulus disappeared. Such delays in attention shifting have been reported in very young human infants, which decrease (i.e., delays grow smaller as attention shifting grows faster) from 6 to 26 weeks (Butcher et al., 2000). Our findings suggest that this delay might also persist when the stimulus holding infants' attention disappears.

We also noted age-related changes in these AOI prolongation effects in both human and macaque infants: For human infants, the extended time window of the AOIs increased the proportion of AOI hits when it remained for up to 800 ms after the stimulus disappeared at the age of 4 months, but narrowed to only be beneficial when extended to 400 ms at 6 months, and appeared no longer to be beneficial with any extension at 8 and 14 months; for macaque infants, the AOI prolongation effect was effective up to 200 ms in the 6-month-olds but not the younger ages. Capture of noise/false positives by extending the AOI durations should have led to an increase in the proportion of AOI hits across all ages. Rather, the systematic, age-related changes in the effect of AOI prolongation are consistent with the interpretation that we captured a greater number of valid fixations. This pattern may also reflect a gradual improvement in infants' ability to more rapidly shift their attention over the first half year after birth (Boothe et al., 1982; Butcher et al., 2000; Johnson et al., 1991; McConnell & Bryson, 2005; Ross-Sheehy et al., 2015; Wass et al., 2013). This delay in attention shifting among young infants is noteworthy when we design eye-tracking tasks that require high temporal accuracy.

However, we did not detect any increases in the proportion of AOI hits with AOI prolongations in 2-month-old humans and 2- to 3-week-old macaques, the youngest groups in the current study. One possible reason for these null results may

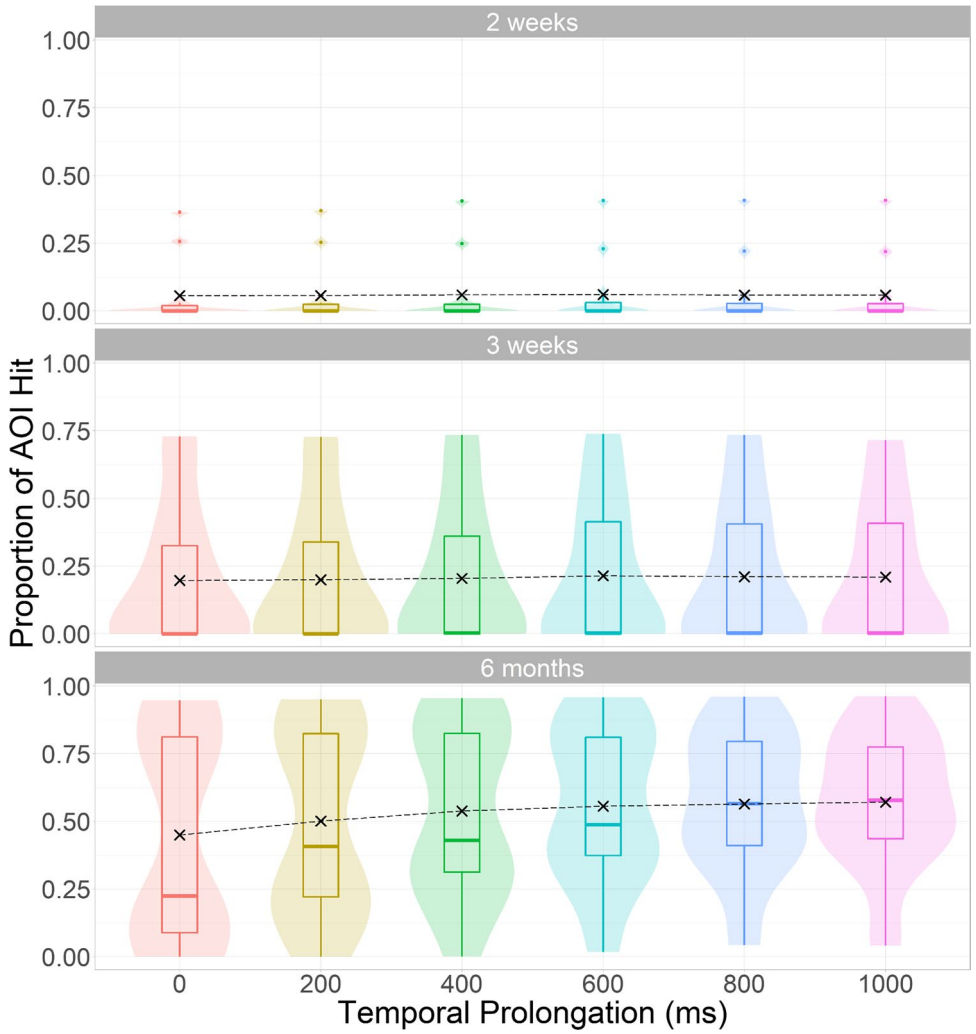

Fig. 12 Effect of AOI temporal prolongation in macaque infants averaging across spatial enlargement at each age at 2 weeks (*top*), 3 weeks (*middle*), and 6 months (*bottom*) of age. Boxplots: *Lines within the boxplots* indicate the medians. The *hinges* of the boxplots show the first (*bottom*) and third (*top*) quartiles. The *whiskers* extend up to 1.5 × interquartile range (IQR; distance between top and bottom hinges), above and below the hinges. The *violin plots* show the distribution of the AOI hits. The black "*X*" indicates the means

be that the target disk was, in fact, displayed too briefly (i.e., only 2 s) which was insufficient time for very young infants to orient to it, particularly given that the intervals between the disk's disappearance at one location and reappearance at another location was also brief (i.e., only 1 s). Human infants' speed to shift attention from one location to the other increases from 2 to 6 months of age (McConnell & Bryson, 2005). For example, one study reported that, even with a central stimulus offset, 2-month-olds needed an average of about 2 s after a peripheral stimulus onset to shift their gaze to it, compared to 6-month-olds who need an average of less than 1 s (McConnell & Bryson, 2005). Therefore, in the current study, the youngest infants may not have had enough time to disengage their attention and shift to another location in rapid succession for all five locations. Consistent with this interpretation, out of the total fixations on the screen, we found that our AOIs, varying in sizes, only mapped an average of 9–23% of fixations on AOIs (out of total fixation on screen) in 2-month-old human infants and only 1–11% in 2-week-old macaque infants. These results suggest that future studies with infants this young may better test AOI temporal prolongations by displaying the target stimuli themselves for longer periods of time to ensure they are fixated on before they disappear. One approach that may ensure stimuli are presented in a way that is fair to different age groups is by using a system-controlled or experimenter-controlled procedure in which an infant must accumulate a certain amount of looking to the screen or to a stimulus before the trial ends (Slonecker et al., 2018).

Unfortunately, in the current study we were unable to distinguish between temporal inaccuracy of the system itself

and delays in infants' latencies to fixate on target AOIs. Nonetheless, the implications are the same: some adjustments to the durations of AOIs may be beneficial for overcoming both potential sources of error.

Limitations and future directions

To our knowledge, the current study is the first to systematically examine how, in young infants, enlarging AOI sizes and extending AOI temporal windows impacts fixation-AOI mapping. We found that prolonging the AOI duration after the stimulus disappearance increased the proportion of AOI hits for both human and macaque infants. This approach may help capture "sticky" fixations, which are theorized to reflect a delay in attention shifting at these early ages (Butcher et al., 2000). However, future studies are needed to further investigate whether this AOI prolongation effect is associated with infants' attention shifting ability, and how we may better design age-appropriate eye-tracking measures in line with infants' attention disengagement skills.

In recent years, primates have been increasingly popular as a model for studying human development using eye-tracking technology, which highlights the need to carefully examine eye-tracking methodology in primate infants at various ages (Nakamura et al., 2021; Ryan et al., 2020). We provided preliminary findings on rhesus macaque infants on how calibration quality and manipulating the sizes and durations of AOIs might improve the Tobii TX300's ability to capture valid fixations. However, the current study was not designed to directly compare eye-tracking performance of infants of both species, thereby lacking a sample of macaque infants that were chronologically age-equivalent, and/or developmentally equivalent in their visual attention systems, to the human infants. Eye-tracking studies on primate infants are uncommon and largely limited to only a few species, much like primate cognition research more generally (Altschul et al., 2019; Nelson et al., in press). Primate infant studies can therefore benefit from pooling resources, sharing protocols, and having well-recognized guidelines, which require systematic examinations of the eye-tracking methods and decisions on primate infants.

Another common practice that requires further systematic examination is the use of a same-aged peer's calibration when a given infant cannot be calibrated successfully. Calibrating young infants can be difficult, as human and primate infants cannot be instructed to look at a stimulus and remain still during testing. Researchers commonly exclude infants who cannot be calibrated reliably from studies (e.g., Gredebäck et al., 2009; Maylott et al., 2020). This exclusion may result in a high amount of data loss and potentially non-random infant dropout, jeopardizing study generalizability (Klein-Radukic & Zmyj, 2015; Segal et al., 2021). Subject dropout in primate studies is particularly troubling, given the small sample sizes to begin with (Farrar et al., 2021; Schubiger et al., 2019). In addition, even though calibration procedures can be repeated until an acceptable calibration is obtained, a previous study in 9- to 10-month-olds found that repeating calibrations multiple times was associated with poorer eye-tracking data accuracy (Hessels et al., 2015). Therefore, researchers sometimes adopt another age-matched infant's calibration profile when a personalized calibration cannot be completed, to maximize the ability to include as many infants as possible (Maylott et al., 2021; Ryan et al., 2020). Although it is ideal to use the infants' own calibration profile, here, we found no evidence in our human infant data that fixation-AOI mapping was poorer when we used another age-matched infant's calibration profile for those infants who failed in calibration compared to infants who used their own calibration profile. However, we had only a small sample of human infants (11 sessions at 2 months out of 79 sessions total; 2 sessions at 4 months out of 88 sessions total) who used others' calibration profiles, so replications with larger samples and extensions to other species are needed. Further, we did not experimentally manipulate whether an infant used their own or another infant's calibration; this method should be studied more systematically (rather than just opportunistically) in future work to better understand the advantages and limitations of this approach. While having fewer infants excluded is ideal, and some approaches may increase usability, it will be useful to better understand how including these infants may impact eye-tracking data quality and fixation-AOI mapping. Another direction that could be explored in future work is to compare operator-controlled (i.e., experimenter-controlled) to system-controlled (i.e., automated) calibration in infants, to determine if one is advantageous over the other in specific populations (Hessels et al., 2015).

While we successfully calibrated all infant macaques in the current study, this success may not reflect the ease with which macaque infants can be calibrated relative to human infants. Instead, quite the contrary: this success was possible mainly because infant macaques were available for repeated attempts at calibration throughout the day across multiple days, unlike human infants whose calibration had to be achieved during a more limited one-time visit to the laboratory at each age. For some infant macaques, repeated attempts were needed across multiple test sessions to obtain a usable calibration. As with human infants, future studies with macaque infants are needed to systematically explore and report the number of calibration attempts and the consequences on data quality when using another infant's calibration profile. Meanwhile we encourage infant researchers to be transparent in reporting these practices.

Here, we focused on a popular eye tracker model (i.e., Tobii TX300) and a widely used, noise-robust fixation classifying algorithm (i.e., I-VT filter). However, many factors may influence the quality of the raw gaze samples, including variation in eye tracker models, the age groups and species studied, and the eye-tracking setup (e.g., room luminance).

Different types of fixation filters and the associated decisions about which parameters to use for these filters (e.g., maximum gap length, smoothing and filtering windows, velocity cutoffs) may also influence the ability to extract reliable and valid fixation candidates from the raw gaze signal (Hooge et al., 2022). Therefore, it is critical for future studies to systematically compare across various eye trackers and fixation filtering algorithms and parameters to examine the extent to which the current findings can be generalized, and to find the best possible procedures to maximize fixation-AOI mapping.

In conclusion, our findings suggest adjustments to infant eye-tracking data collection and processing methods may help researchers collect more data from human and primate infants. When used in conjunction with other recommended practices—such as applying new algorithms for detecting fixations from raw gaze signals (Wass et al., 2013), optimizing the testing environments and infant states for eye tracking (Hessels & Hooge, 2019), and using infant-friendly calibration procedures (Gredebäck et al., 2009)—the approaches recommended here may improve fixation-AOI mapping. Determining how data can be used optimally, even if produced by less-than-ideal populations, will strengthen eye-tracking paradigms, as well as uncover points of commonality and difference between humans and animals at different ages, facilitating comparative and developmental science. Ultimately, establishing these evidence-based approaches will produce more robust data, replicable findings, and reliable interpretations, shedding light on the ontogenetic and phylogenetic emergence of perceptual, cognitive, social, and emotional development.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13428-022-02056-3.

Acknowledgements We are grateful to the families who participated, researchers in the Social Cognition Laboratory at the University of Miami for assistance with human infant data collection, especially Roberto Lazo and Sarah Maylott, and researchers and animal care staff in the Laboratory of Comparative Ethology at the National Institutes of Health for assistance with macaque infant data collection, especially Sheila Sutti, Emily Slonecker, Kristin Byers, and Michelle Miller. Thanks to Scott P. Johnson for sharing stimulus materials, and to Tiffany Leung, Thea Saunders, Lindsey Faucher, Baylee Brochu, and Catalina Sanchez Montenegro for feedback on an earlier version of the paper.

Code availability R markdown for replicating data analyses is available as a supplementary file and at https://osf.io/p9mwk/?view_only=a0800300342b44f883c95145d45b411c

Authors' contributions AP, EAS, and GZ designed the study. GZ and EAS collected the human data. AP and EAS collected the macaque data. GZ analyzed the data and created the graphs. GZ and EAS wrote the manuscript. GZ, EAS, and AP edited the manuscript.

Funding Funded by a National Science Foundation CAREER Award 1653737, the Division of Intramural Research (Project 1ZIAHD001106), *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, USA, and an Association for Psychological Science James McKeen Cattell Fund Fellowship Sabbatical Award.

Data availability The video stimulus and de-identified data for both human and macaque infant eye tracking are included in the supplementary materials and are also available at https://osf.io/p9mwk/?view_only=a0800300342b44f883c95145d45b411c.

Declarations

Conflicts of interest The authors have no relevant financial or non-financial interests to disclose.

Ethics approval The Institutional Review Board for Human Subject Research at the University of Miami approved this study for human participants. The *Eunice Kennedy Shriver* National Institute of Child Health and Human Development Animal Care and Use Committee approved the procedures for macaques. We conducted the study in accordance with the Guide for the Care and Use of Laboratory Animals and complied with the Animal Welfare Act.

Consent to participate Informed consent was obtained from the caregivers of all human infants included in the study.

Consent for publication The authors affirm that both human adults consented to publication of the images in Fig. 3.

References

Altschul, D. M., Beran, M. J., Bohn, M., Caspar, K. R., Fichtel, C., Försterling, M., et al. (2019). Collaborative open science as a way to reproducibility and new insights in primate cognition research. *Japanese Psychological Review, 62*(3), 205–220. https://doi.org/10.24602/sjpr.62.3_205

Aslin, R. N. (2007). What's in a look? *Developmental Science, 10*(1), 48–53. https://doi.org/10.1111/j.1467-7687.2007.00563.x

Aslin, R. N., & McMurray, B. (2004). Automated corneal-reflection eye tracking in infancy: Methodological developments and applications to cognition. *Infancy, 6*(2), 155–163. https://doi.org/10.1207/s15327078in0602_1

Akbas, E., & Eckstein, M. P. (2017). Object detection through search with a foveated visual system. *PLOS Computational Biology, 13*(10), e1005743. https://doi.org/10.1371/journal.pcbi.1005743

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

Boothe, R. G., Kiorpes, L., Regal, D. M., & Lee, C. P. (1982). Development of visual responsiveness in *Macaca nemestrina monkeys*. *Developmental Psychology, 18*(5), 665–670. https://doi.org/10.1037/0012-1649.18.5.665

Brémond-Gignac, D., Copin, H., Lapillonne, A., & Milazzo, S. (2011). Visual development in infants: Physiological and pathological mechanisms. *Current Opinion in Ophthalmology, 22*, S1–S8. https://doi.org/10.1097/01.icu.0000397180.37316.5d

Bronson, G. W. (1994). Infants' transitions toward adult-like scanning. *Child Development, 65*(5), 1243–1261. https://doi.org/10.1111/j.1467-8624.1994.tb00815.x

Butcher, P. R., Kalverboer, A. F., & Geuze, R. H. (2000). Infants' shifts of gaze from a central to a peripheral stimulus: A longitudinal study of development between 6 and 26 weeks. *Infant Behavior and Development, 23*(1), 3–21. https://doi.org/10.1016/S0163-6383(00)00031-X

Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology, 155*, 49–62. https://doi.org/10.1016/j.ijpsycho.2020.05.010

Chandna, A. (1991). Natural history of the development of visual acuity in infants. *Eye, 5*(1), 20–26. https://doi.org/10.1038/eye.1991.4

Dalrymple, K. A., Manner, M. D., Harmelink, K. A., Teska, E. P., & Elison, J. T. (2018). An examination of recording accuracy and precision from eye tracking data from toddlerhood to adulthood. *Frontiers in Psychology, 9*, 803. https://doi.org/10.3389/fpsyg.2018.00803

Damon, F., Méary, D., Quinn, P. C., Lee, K., Simpson, E. A., Paukner, A., et al. (2017). Preference for facial averageness: Evidence for a common mechanism in human and macaque infants. *Scientific Reports, 7*(1), 46303. https://doi.org/10.1038/srep46303

De Kloe, Y. J., Hooge, I. T., Kemner, C., Niehorster, D. C., Nyström, M., & Hessels, R. S. (2022). Replacing eye trackers in ongoing studies: A comparison of eye-tracking data quality between the Tobii Pro TX300 and the Tobii Pro Spectrum. *Infancy, 27*(1), 25–45. https://doi.org/10.1111/infa.12441

D'Entremont, B. (2000). A perceptual–attentional explanation of gaze following in 3- and 6-month-olds. *Developmental Science, 3*(3), 302–311. https://doi.org/10.1111/1467-7687.00124

Dobson, V., & Teller, D. Y. (1978). Visual acuity in human infants: A review and comparison of behavioral and electrophysiological studies. *Vision Research, 18*(11), 1469–1483. https://doi.org/10.1016/0042-6989(78)90001-9

Dupierrix, E., de Boisferon, A. H., Méary, D., Lee, K., Quinn, P. C., Di Giorgio, E., Simion, F., Tomonaga, M., & Pascalis, O. (2014). Preference for human eyes in human infants. *Journal of Experimental Child Psychology, 123*, 138–146. https://doi.org/10.1016/j.jecp.2013.12.010

Farrar, B. G., Voudouris, K., & Clayton, N. S. (2021). Replications, comparisons, sampling and the problem of representativeness in animal cognition research. *Animal Behavior and Cognition, 8*(2), 273. https://doi.org/10.26451/2Fabc.08.02.14.2021

Feng, G. (2011). Eye tracking: A brief guide for developmental researchers. *Journal of Cognition and Development, 12*(1), 1–11. https://doi.org/10.1080/15248372.2011.547447

Gluckman, M., & Johnson, S. (2013). Attentional capture by social stimuli in young infants. *Frontiers in Psychology, 4*. https://www.frontiersin.org/article/10.3389/fpsyg.2013.00527

Gredebäck, G., Johnson, S., & von Hofsten, C. (2009). Eye tracking in infancy research. *Developmental Neuropsychology, 35*(1), 1–19. https://doi.org/10.1080/87565640903325758

Groot, C., Ortega, F., & Beltran, F. S. (1994). Thumb rule of visual angle: A new confirmation. *Perceptual and Motor Skills, 78*(1), 232–234. https://doi.org/10.2466/pms.1994.78.1.232

Heiney, S. A., & Blazquez, P. M. (2011). Behavioral responses of trained squirrel and rhesus monkeys during oculomotor tasks. *Experimental Brain Research, 212*(3), 409–416. https://doi.org/10.1007/s00221-011-2746-4

Hessels, R. S., Andersson, R., Hooge, I. T., Nyström, M., & Kemner, C. (2015). Consequences of eye color, positioning, and head movement for eye-tracking data quality in infant research. *Infancy, 20*(6), 601–633. https://doi.org/10.1111/infa.12093

Hessels, R. S., & Hooge, I. T. (2019). Eye tracking in developmental cognitive neuroscience – The good, the bad and the ugly. *Developmental Cognitive Neuroscience, 40*, 100710. https://doi.org/10.1016/j.dcn.2019.100710

Hessels, R. S., Kemner, C., van den Boomen, C., & Hooge, I. T. C. (2016). The area-of-interest problem in eyetracking research: A noise-robust solution for face and sparse stimuli. *Behavior Research Methods, 48*(4), 1694–1712. https://doi.org/10.3758/s13428-015-0676-y

Hirata, S., Fuwa, K., Sugama, K., Kusunoki, K., & Fujita, S. (2010). Facial perception of conspecifics: Chimpanzees (Pan troglodytes) preferentially attend to proper orientation and open eyes. *Animal Cognition, 13*(5), 679–688. https://doi.org/10.1007/s10071-010-0316-y

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Holmqvist, K., Örbom, S. L., Hooge, I. T., Niehorster, D. C., Alexander, R. G., Andersson, R., et al. (2022). Eye tracking: Empirical foundations for a minimal reporting guideline. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01762-8

Hooge, I. T., Niehorster, D. C., Nyström, M., Andersson, R., & Hessels, R. S. (2022). Fixation classification: How to merge and select fixation candidates. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01723-1

Hopper, L. M., Gulli, R. A., Howard, L. H., Kano, F., Krupenye, C., Ryan, A. M., & Paukner, A. (2021). The application of noninvasive, restraint-free eye-tracking methods for use with nonhuman primates. *Behavior Research Methods, 1–28*. https://doi.org/10.3758/s13428-020-01465-6

Jakobsen, K. V., Umstead, L., & Simpson, E. A. (2016). Efficient human face detection in infancy. *Developmental Psychobiology, 58*(1), 129–136. https://doi.org/10.1002/dev.21338

Johnson, M. H., Posner, M. I., & Rothbart, M. K. (1991). Components of visual orienting in early infancy: Contingency learning, anticipatory looking, and disengaging. *Journal of Cognitive Neuroscience, 3*(4), 335–344. https://doi.org/10.1162/jocn.1991.3.4.335

Johnson, S. P. (2019). Development of visual-spatial attention. In T. Hodgson (Ed.), *Processes of Visuospatial Attention and Working Memory. Current Topics in Behavioral Neurosciences* (Vol. 41). Springer. https://doi.org/10.1007/7854_2019_96

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature, 504*, 427–431. https://doi.org/10.1038/nature12715

Kano, F., & Tomonaga, M. (2009). How chimpanzees look at pictures: A comparative eye-tracking study. *Proceedings of the Royal Society B: Biological Sciences, 276*(1664), 1949–1955. https://doi.org/10.1098/rspb.2008.1811

Kano, F., Call, J., & Tomonaga, M. (2012). Face and eye scanning in gorillas (*Gorilla gorilla*), orangutans (*Pongo abelii*), and humans (*Homo sapiens*): Unique eye-viewing patterns in humans among hominids. *Journal of Comparative Psychology, 126*(4), 388–398. https://doi.org/10.1037/a0029615

Katus, L., Hayes, N. J., Mason, L., Blasi, A., McCann, S., Darboe, M. K., de Haan, M., Moore, S. E., Lloyd-Fox, S., & Elwell, C. E. (2019). Implementing neuroimaging and eye tracking methods to assess neurocognitive development of young infants in low- and middle-income countries. *Gates Open Research, 3*, 1113. https://doi.org/10.12688/gatesopenres.12951.2

Kiorpes, L. (2015). Visual development in primates: Neural mechanisms and critical periods. *Developmental Neurobiology, 75*(10), 1080–1090. https://doi.org/10.1002/dneu.22276

Klein-Radukic, S., & Zmyj, N. (2015). Dropout in looking time studies: The role of infants' temperament and cognitive developmental status. *Infant Behavior and Development, 41*, 142–153. https://doi.org/10.1016/j.infbeh.2015.10.001

Kulke, L., Atkinson, J., & Braddick, O. (2015). Automatic detection of attention shifts in infancy: Eye tracking in the fixation shift paradigm. *PLoS ONE, 10*(12), e0142505. https://doi.org/10.1371/journal.pone.0142505

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LmerTest: Package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13). https://doi.org/10.18637/jss.v082.i13

Liversedge, S. P., & Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences, 4*(1), 6–14. https://doi.org/10.1016/S1364-6613(99)01418-7

McConnell, B. A., & Bryson, S. E. (2005). Visual attention and temperament: Developmental data from the first 6 months of life. *Infant Behavior and Development, 28*(4), 537–544. https://doi.org/10.1016/j.infbeh.2005.09.002

Machado, C. J., & Nelson, E. E. (2011). Eye-tracking with nonhuman primates is now more accessible than ever before. *American Journal of Primatology, 73*(6), 562–569. https://doi.org/10.1002/ajp.20928

Machado, C. J., Whitaker, A. M., Smith, S. E., Patterson, P. H., & Bauman, M. D. (2015). Maternal immune activation in nonhuman primates alters social attention in juvenile offspring. *Biological Psychiatry, 77*(9), 823–832. https://doi.org/10.1016/j.biopsych.2014.07.035

Maylott, S. E., Sansone, J. R., Jakobsen, K. V., & Simpson, E. A. (2021). Superior detection of faces in male infants at 2 months. *Child Development, 92*(4), e621–e634. https://doi.org/10.1111/cdev.13543

Maylott, S. E., Paukner, A., Ahn, Y. A., & Simpson, E. A. (2020). Human and monkey infant attention to dynamic social and nonsocial stimuli. *Developmental Psychobiology, 62*(6), 841–857. https://doi.org/10.1002/dev.21979

Mendelson, M. J., Haith, M. M., & Goldman-Rakic, P. S. (1982). Face scanning and responsiveness to social cues in infant rhesus monkeys. *Developmental Psychology, 18*(2), 222–228. https://doi.org/10.1037/0012-1649.18.2.222

Milliren, C. E., Evans, C. R., Richmond, T. K., & Dunn, E. C. (2018). Does an uneven sample size distribution across settings matter in cross-classified multilevel modeling? Results of a simulation study. *Health & Place, 52*, 121–126. https://doi.org/10.1016/j.healthplace.2018.05.009

Mitchell, J. F., & Leopold, D. A. (2015). The marmoset monkey as a model for visual neuroscience. *Neuroscience Research, 93*, 20–46. https://doi.org/10.1016/j.neures.2015.01.008

Morgante, J. D., Zolfaghari, R., & Johnson, S. P. (2012). A critical test of temporal and spatial accuracy of the Tobii T60XL eye tracker. *Infancy, 17*(1), 9–32. https://doi.org/10.1111/j.1532-7078.2011.00089.x

Muschinski, J., Feczko, E., Brooks, J. M., Collantes, M., Heitz, T. R., & Parr, L. A. (2016). The development of visual preferences for direct versus averted gaze faces in infant macaques (*Macaca mulatta*). *Developmental Psychobiology, 58*(8), 926–936. https://doi.org/10.1002/dev.21421

Nakamura, T., Fujiwara, K., Saitou, M., & Tsukiyama, T. (2021). Nonhuman primates as a model for human development. *Stem Cell Reports, 16*(5), 1093–1103. https://doi.org/10.1016/j.stemcr.2021.03.021

Nelson, E. L., Alvarez, J., Jimenez, B., & Padron, K. (in press). State of the Field: Developmental Primate Cognition. In Schwartz, B.L. & Beran, M.J. (under contract). *Primate Cognitive Studies*. Cambridge University Press. Retrieved from https://hands.fiu.edu/PrimateCognitiveStudies_Nelson_Preprint.pdf

Niehorster, D. C., Andersson, R., & Nyström, M. (2020). Titta: A toolbox for creating PsychToolbox and PsychoPy experiments with Tobii eye trackers. *Behavior Research Methods, 52*(5), 1970–1979. https://doi.org/10.3758/s13428-020-01358-8

Nyström, M., Andersson, R., Holmqvist, K., & Van De Weijer, J. (2013). The influence of calibration method and eye physiology on eyetracking data quality. *Behavior Research Methods, 45*(1), 272–288. https://doi.org/10.3758/s13428-012-0247-4

Oakes, L. M. (2012). Advances in eye tracking in infancy research. *Infancy, 17*(1), 1–8. https://doi.org/10.1111/j.1532-7078.2011.00101.x

Olsen, A., & Matos, R. (2012). Identifying parameter values for an I-VT fixation filter suitable for handling data sampled with various sampling frequencies. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 317–320. https://doi.org/10.1145/2168556.2168625

Ordy, J. M., Latanick, A., Samorajski, T., & Massopust, L. C. (1964). Visual acuity in newborn primate infants. *Experimental Biology and Medicine, 115*, 677–680. https://doi.org/10.3181/00379727-115-29004

Orquin, J. L., Ashby, N. J., & Clarke, A. D. (2016). Areas of interest as a signal detection problem in behavioral eye-tracking research. *Journal of Behavioral Decision Making, 29*(2–3), 103–115. https://doi.org/10.1002/bdm.1867

O'Shea, R. P. (1991). Thumb's rule tested: Visual angle of thumb's width is about 2 deg. *Perception, 20*(3), 415–418. https://doi.org/10.1068/p200415

Parr, L. A., Brooks, J. M., Jonesteller, T., Moss, S., Jordano, J. O., & Heitz, T. R. (2016a). Effects of chronic oxytocin on attention to dynamic facial expressions in infant macaques. *Psychoneuroendocrinology, 74*, 149–157. https://doi.org/10.1016/j.psyneuen.2016.08.028

Parr, L. A., Murphy, L., Feczko, E., Brooks, J., Collantes, M., & Heitz, T. R. (2016b). Experience-dependent changes in the development of face preferences in infant rhesus monkeys. *Developmental Psychobiology, 58*(8), 1002–1018. https://doi.org/10.1002/dev.21434

Paukner, A., Bower, S., Simpson, E. A., & Suomi, S. J. (2013). Sensitivity to first-order relations of facial elements in infant rhesus macaques. *Infant and Child Development, 22*(3), 320–330. https://doi.org/10.1002/icd.1793

Paukner, A., Simpson, E. A., Ferrari, P. F., Mrozek, T., & Suomi, S. J. (2014). Neonatal imitation predicts how infants engage with faces. *Developmental Science, 17*(6), 833–840. https://doi.org/10.1111/desc.12207

Paukner, A., Slonecker, E. M., Murphy, A. M., Wooddell, L. J., & Dettmer, A. M. (2018). Sex and rank affect how infant rhesus macaques look at faces. *Developmental Psychobiology, 60*(2), 187–193. https://doi.org/10.1002/dev.21579

Phillips, J. O., Liberato, N. A., Grant, K., Burbacher, T. M., Jacobs, C. M., & Weiss, A. H. (2007). Development of optokinetic and smooth pursuit eye movements in infant monkeys. *Investigative Ophthalmology & Visual Science, 48*(13), 2837.

Pomaranski, K. I., Hayes, T. R., Kwon, M. K., Henderson, J. M., & Oakes, L. M. (2021). Developmental changes in natural scene viewing in infancy. *Developmental Psychology, 57*(7), 1025–1041. https://doi.org/10.1037/dev0001020

Pons, F., Bosch, L., & Lewkowicz, D. J. (2019). Twelve-month-old infants' attention to the eyes of a talking face is associated with communication and social skills. *Infant Behavior and Development, 54*, 80–84. https://doi.org/10.1016/j.infbeh.2018.12.003

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Richards, J. E. (2004). The development of sustained attention in infants. In M. I. Posner (Ed.), *Cognitive Neuroscience of Attention* (pp. 342–356). The Guilford Press.

Richards, J. E. (2010). The development of attention to simple and complex visual stimuli in infants: Behavioral and psychophysiological measures. *Developmental Review, 30*(2), 203–219. https://doi.org/10.1016/j.dr.2010.03.005

Ross-Sheehy, S., Schneegans, S., & Spencer, J. P. (2015). The infant orienting with attention task: Assessing the neural basis of spatial attention in infancy. *Infancy, 20*(5), 467–506. https://doi.org/10.1111/infa.12087

Ryan, A. M., Berman, R. F., & Bauman, M. D. (2019). Bridging the species gap in translational research for neurodevelopmental disorders. *Neurobiology of Learning and Memory, 165*, 106950. https://doi.org/10.1016/j.nlm.2018.10.006

Ryan, A. M., Murai, T., Lau, A. R., Hogrefe, C. E., McAllister, A. K., Carter, C. S., & Bauman, M. D. (2020). New approaches to quantify social development in rhesus macaques (*Macaca mulatta*): Integrating eye tracking with traditional assessments of social behavior. *Developmental Psychobiology, 62*(7), 950–962. https://doi.org/10.1002/dev.22003

Schlegelmilch, K., & Wertz, A. E. (2019). The effects of calibration target, screen location, and movement type on infant eye-tracking data quality. *Infancy, 24*(4), 636–662. https://doi.org/10.1111/infa.12294

Schubiger, M. N., Kissling, A., & Burkart, J. M. (2019). Does opportunistic testing bias cognitive performance in primates? *Learning from drop-outs. PLoS ONE, 14*(3), e0213727. https://doi.org/10.1371/journal.pone.0213727

Segal, S. C., Marquis, A. R., & Moulson, M. C. (2021). Are our samples representative? Understanding whether temperament influences infant dropout rates at 3 and 7 months. *Infant Behavior and Development, 65*, 101630. https://doi.org/10.1016/j.infbeh.2021.101630

Senju, A., & Csibra, G. (2008). Gaze Following in Human Infants Depends on Communicative Signals. *Current Biology, 18*(9), 668–671. https://doi.org/10.1016/j.cub.2008.03.059

Simpson, E. A., Jakobsen, K. V., Damon, F., Suomi, S. J., Ferrari, P. F., & Paukner, A. (2017). Face detection and the development of own-species bias in infant macaques. *Child Development, 88*(1), 103–113. https://doi.org/10.1111/cdev.12565

Simpson, E. A., Maylott, S. E., Leonard, K., Lazo, R. J., & Jakobsen, K. V. (2019). Face detection in infants and adults: Effects of orientation and color. *Journal of Experimental Child Psychology, 186*, 17–32. https://doi.org/10.1016/j.jecp.2019.05.001

Simpson, E. A., Miller, G. M., Ferrari, P. F., Suomi, S. J., & Paukner, A. (2016a). Neonatal imitation and early social experience predict gaze following abilities in infant monkeys. *Scientific Reports, 6*, 20233. https://doi.org/10.1038/srep20233

Simpson, E. A., Nicolini, Y., Shetler, M., Suomi, S. J., Ferrari, P. F., & Paukner, A. (2016b). Experience-independent sex differences in newborn macaques: Females are more social than males. *Scientific Reports, 6*(1), 19669. https://doi.org/10.1038/srep19669

Slonecker, E. M., Simpson, E. A., Suomi, S. J., & Paukner, A. (2018). Who's my little monkey? Effects of infant-directed speech on visual retention in infant rhesus macaques. *Developmental Science, 21*(2), e12519. https://doi.org/10.1111/desc.12519

SR Research, Inc. (2007). *EyeLink user manual. Version 1.3. 0 [Computer software manual]*. SR Research Inc..

Teller, D. Y. (1981). The development of visual acuity in human and monkey infants. *Trends in Neurosciences, 4*, 21–24. https://doi.org/10.1016/0166-2236(81)90009-6

Tobii Technology, A. B. (2016). *Tobii Studio user's manual v3. 4.5 [Computer software manual]*. Tobii Technology AB https://www.tobiipro.com/siteassets/tobii-pro/user-manuals/tobii-pro-studio-user-manual.pdf

Tobii Technology, A. B. (2017). *Accuracy and precision test report (TX300 fw 1.1.1 RC bright light illumination mode)*. Tobii Technology AB https://www.tobiipro.com/siteassets/tobii-pro/accuracy-and-precision-tests/tobii-tx300-eye-tracker-fw-1.1.1-accuracy-and-precision-test-report.pdf

Van Renswoude, D. R., Johnson, S. P., Raijmakers, M. E. J., & Visser, I. (2016). Do infants have the horizontal bias? *Infant Behavior and Development, 44*, 38–48. https://doi.org/10.1016/j.infbeh.2016.05.005

Vehlen, A., Standard, W., & Domes, G. (2022). How to choose the size of facial areas of interest in interactive eye tracking. *PLoS ONE, 17*(2), e0263594. https://doi.org/10.1371/journal.pone.0263594

Von Hofsten, C., & Rosander, K. (1997). Development of smooth pursuit tracking in young infants. *Vision Research, 37*(13), 1799–1810. https://doi.org/10.1016/S0042-6989(96)00332-X

Wang, A., Payne, C., Moss, S., Jones, W. R., & Bachevalier, J. (2020). Early developmental changes in visual social engagement in infant rhesus monkeys. *Developmental Cognitive Neuroscience, 43*, 100778. https://doi.org/10.1016/j.dcn.2020.100778

Wass, S. V., Forssman, L., & Leppänen, J. (2014). Robustness and Precision: How data quality may influence key dependent variables in infant eye-tracker analyses. *Infancy, 19*(5), 427–460. https://doi.org/10.1111/infa.12055

Wass, S. V., Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods, 45*(1), 229–250. https://doi.org/10.3758/s13428-012-0245-6

Wilkinson, K. M., & Mitchell, T. (2014). Eye tracking research to answer questions about augmentative and alternative communication assessment and intervention. *Augmentative and Alternative Communication, 30*(2), 106–119. https://doi.org/10.3109/07434618.2014.904435

Xiang, Y., Long, E., Liu, Z., Li, X., Lin, Z., Zhu, Y., et al. (2021). Study to establish visual acuity norms with Teller Acuity Cards II for infants from southern China. *Eye, 35*(10), 2787–2792. https://doi.org/10.1038/s41433-020-01314-y

Xue, J., Quan, C., Li, C., Yue, J., & Zhang, C. (2017). A crucial temporal accuracy test of combining EEG and Tobii eye tracker. *Medicine, 96*(13). https://doi.org/10.1097/MD.0000000000006444

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.