



OWLET: An automated, open-source method for infant gaze tracking using smartphone and webcam recordings

Denise M. Werchan^{1,2} · Moriah E. Thomason^{1,2} · Natalie H. Brito³

Accepted: 20 August 2022
© The Psychonomic Society, Inc. 2022

Abstract

Groundbreaking insights into the origins of the human mind have been garnered through the study of eye movements in preverbal subjects who are unable to explain their thought processes. Developmental research has largely relied on in-lab testing with trained experimenters. This constraint provides a narrow window into infant cognition and impedes large-scale data collection in families from diverse socioeconomic, geographic, and cultural backgrounds. Here we introduce a new open-source methodology for automatically analyzing infant eye-tracking data collected on personal devices in the home. Using algorithms from computer vision, machine learning, and ecological psychology, we develop an online webcam-linked eye tracker (OWLET) that provides robust estimation of infants' point of gaze from smartphone and webcam recordings of infant assessments in the home. We validate OWLET in a large sample of 7-month-old infants ($N = 127$) tested remotely, using an established visual attention task. We show that this new method reliably estimates infants' point-of-gaze across a variety of contexts, including testing on both computers and mobile devices, and exhibits excellent external validity with parental-report measures of attention. Our platform fills a significant gap in current tools available for rapid online data collection and large-scale assessments of cognitive processes in infants. Remote assessment addresses the need for greater diversity and accessibility in human studies and may support the ecological validity of behavioral experiments. This constitutes a critical and timely advance in a core domain of developmental research and in psychological science more broadly.

Keywords Webcam eye tracking · Online studies · Infancy · Development

Introduction

Looking is one of the earliest behaviors to develop in young infants and provides a gateway into the human mind and brain before description of thoughts and experiences can originate firsthand. The study of infant looking behavior has had an unparalleled impact on our understanding of social, cognitive, and emotional processing at the beginning of postnatal life (Aslin, 2007). Over the past decade, advances in eye-tracking technology have afforded precise, automatic

quantification of infant looking behavior with high spatial and temporal resolution. Paired with clever experimental designs, this tool has allowed scientists to test previously intractable hypotheses about fundamental aspects of the human mind, including the origins of object perception (Johnson et al., 2003, 2004), attention (Amso et al., 2014; Werchan et al., 2019), face processing (Frank et al., 2009; Liu et al., 2011), and infants' remarkable capacity for learning (Kirkham et al., 2007; Werchan et al., 2015, 2016; Werchan & Amso, 2020). While this methodological advance has led to foundational discoveries into how infants experience and understand the world, it also typically requires specialized, technical expertise and expensive hardware housed in research laboratories. These constraints present two key challenges. First, testing in artificial laboratory settings provides a narrow window into the full repertoire of infant behavior and constrains ecological validity. Second, and most importantly, the challenges of recruiting families for in-person research limits large-scale data collection in

✉ Denise M. Werchan
denise.werchan@nyulangone.org

¹ Department of Population Health, New York University School of Medicine, 227 E 30th St, 7th Fl, New York, NY 10016, USA

² Department of Child & Adolescent Psychiatry, New York University School of Medicine, New York, NY 10016, USA

³ Department of Applied Psychology, New York University, New York, NY 10003, USA

diverse samples, impeding initiatives to increase reproducibility and equity in developmental science.

One promising and timely solution is to take research out of the lab. Recent platforms developed for online testing of infants using webcam videos, such as LookIt, have potential to facilitate the collection of data from larger and more diverse samples of infants (Scott & Schulz, 2017). In addition, these efforts may also have benefits for helping standardize and explicitly document best practices for infant research, aligning with an open science framework that emphasizes methods to increase the reproducibility and transparency of developmental science (Frank et al., 2017). Despite online platforms increasing the ease of testing larger samples of infants, however, videos from online testing sessions still require manual annotation by extensively trained coders. This slow, labor-intensive process hinders the feasibility of collecting large sample sizes through online data collection. Importantly, it is also subject to replicability-challenges introduced by potential systematic biases in subjective judgements of infant behavior across laboratories.

Automated coding of infant looking behavior is an area of active development that may help address this issue. Existing computational algorithms attempt to classify infant gaze direction from webcam videos using computer-vision facial landmark extraction and machine learning classifiers (Erel et al., 2022a; Chouinard et al., 2019). These algorithms show success in quantifying the macrostructure of infant gaze, such as differentiating looks towards or away from the screen. However, due in part to difficulties introduced by excessive motion in non-compliant infant subjects, current methods fail to quantify more detailed information about gaze patterns and eye movements. Using existing techniques, it is not possible to estimate coordinates of *where* infants are looking on a display. This limits the array of experimental methods appropriate for use in online testing and constrains the ability to collect data on more complex aspects of infant learning, cognition, and neurological function. Moreover, existing algorithms for infant gaze detection thus far have been evaluated using computer webcam videos. Validating remote methods suitable for use with smartphones and tablets is important to support socio-demographic diversity in online data collection, particularly given the digital divide in access to computers relative to mobile devices across racial, geographic, and socioeconomic strata (Perrin & Atske, 2021; Vogels, 2021).

The present goals are, first, to develop an open-source methodology that supports the extraction, processing, and analysis of infant gaze data from videos recorded on computer webcams (laptops or desktops) and mobile devices (smartphones or tablets) in the home. By integrating algorithms from computer vision, machine learning, and ecological psychology, we develop an online webcam-linked eye tracker (OWLET) that implements principles of

perception-action coupling (Lee, 1998) to solve two challenges in infant webcam-based eye tracking: (1) detecting and integrating changes in infant's position with changes in eye movements to estimate gaze direction, (2) mapping changes in infant's estimated gaze direction to screen coordinates. To support rapid and broad uptake of this methodology for progress in online data collection, we provide administration details, minimal conditions for the quality of infant videos, and open-source scripts to dynamically estimate point-of-gaze at a temporal resolution of 30 Hz.

Second, we examine the accuracy, reliability, and validity of this new open-source methodology for analyzing data from remote infant assessments. We evaluate the spatial accuracy of this methodology by measuring offsets between infants' cued point-of-gaze and OWLET's estimated point-of-gaze using calibration/validation videos from a large and relatively racially and socioeconomically diverse sample of 5- to 8-month-old infants tested on personal devices in the home. In addition, we assess construct validity by using this tool to automatically code infant gaze behavior and looking times during an established visual attention task. The visual attention measures obtained using OWLET are assessed for replication of prior lab-based experimental findings, as well as convergence with subjective, parent-report measures of infant attention. We also assess reliability of OWLET in comparison to human coded evaluations of infant looking times. We evaluate feasibility by assessing eye-tracking data quality in videos recorded using computers relative to mobile devices, and across infant racial and ethnic categories.

Approach

Overview

Figure 1 provides a broad overview of the framework underlying OWLET. It consists of three major components: 1) extraction of the infant's face/eye/pupil from each video frame using computer vision and machine learning algorithms, 2) estimation of the infant's gaze direction, which is grounded in principles from ecological psychology, 3) estimation of the infant's point-of-gaze on the screen, which uses a simple polynomial transfer function to map the infant's gaze direction to precise screen coordinates. All components of OWLET were developed using Python version 3.9.4. To accommodate the diverse conditions encountered during remote infant assessments, OWLET was developed with relatively minimal requirements for the quality of webcam/smartphone recordings: (1) the infant's face should be in line and close to the camera (~8-24 inches away); (2) the frame rate of the video should be at least 30 fps (the default for Zoom or QuickTime recordings; equivalent to

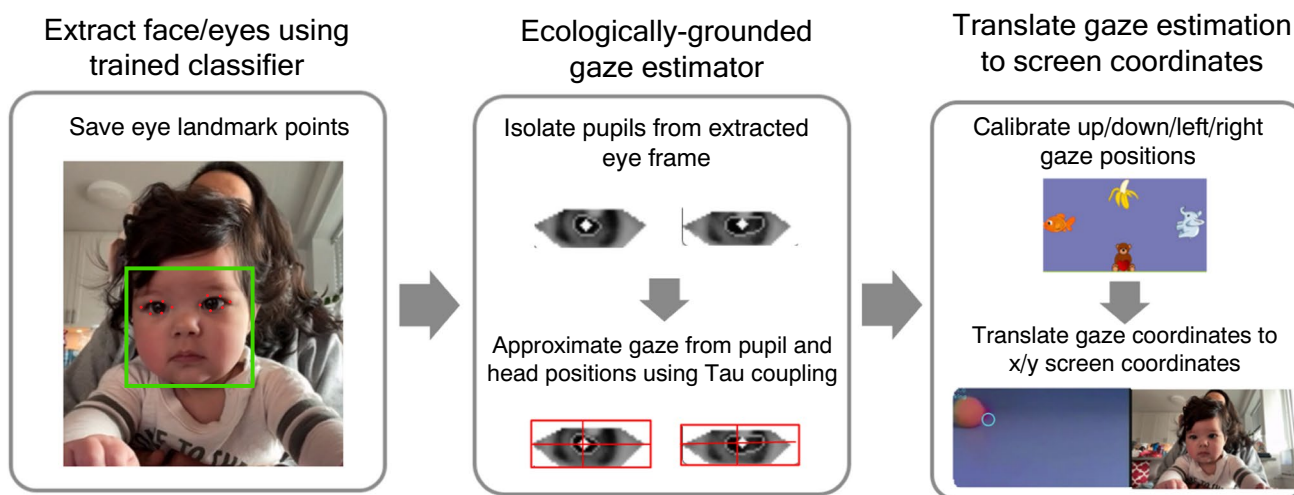


Fig. 1 Overview of the broad framework underlying OWLET. The core aspects of OWLET involve the application of algorithms from computer vision and machine learning to extract eye/pupil information and theoretical principles from ecological psychology to approximate gaze

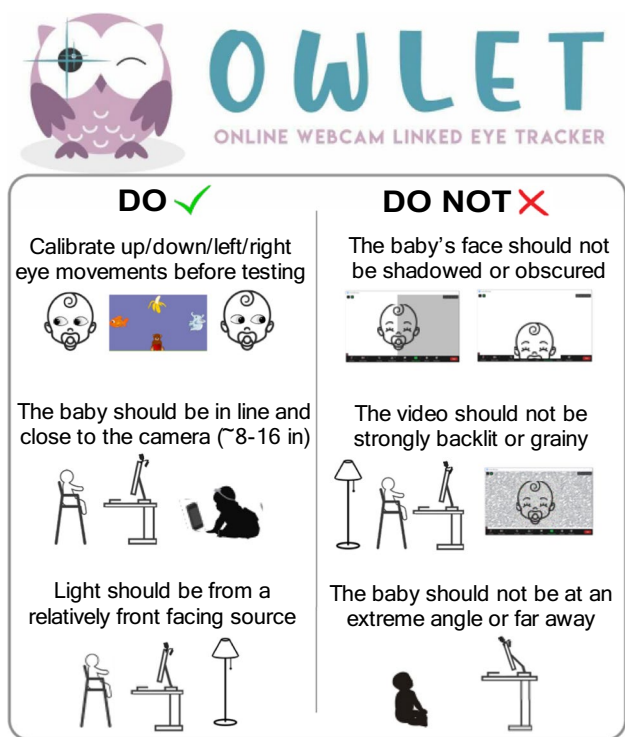


Fig. 2 Minimal video conditions for high-quality infant eye tracking using OWLET

sampling at a rate of 30 Hz); and (3) the lighting should be relatively even and not back-lit (Fig. 2). In addition, our eye tracker is designed to perform optimally through the use of a short calibration procedure prior to testing, where the infant looks at the edge of each side of the screen. This calibration procedure is used to validate the infant’s point-of-gaze

and to determine the spatial accuracy of the estimated gaze position.

Face/eye/pupil extraction

The infant’s video feed is processed frame-by-frame using the OpenCV library in Python (Bradski, 2000), which can be adapted to occur online or offline. First, the Dlib Machine Learning Toolkit is used to extract the infant’s face from the video frame and the coordinates of the associated facial landmarks (King, 2009). If more than one face is detected, the lower face is selected. The Dlib facial landmarks are then used to isolate the eye region in the detected face. Next, a series of image processing steps are applied to the eye frame, including contrast enhancement to increase the perceptual distance between the iris and sclera, followed by a bilateral filter and then a Gaussian blur to smooth over noise in the image while preserving edge information. The eye frame is then thresholded to isolate the iris. This threshold is determined dynamically by calculating the average pixel color of the eye frame. From the thresholded eye frame, a contour detector is applied to segment the iris from the rest of the eye. A convex hull is used with the contour detector to smooth over potential indents in the image (e.g., from reflections of light on the iris). The pupil is then isolated by calculating the centroid of the segmented iris from the moments of the image.

Ecologically grounded gaze direction estimator

After isolating the pupil, the location of the pupil is calculated relative to the width of the eye (to estimate horizontal gaze) and the height of each eye (to estimate vertical

gaze). In a second step, we account for the infant's head pose in estimating gaze direction. Existing computer vision algorithms for determining head pose direction from video feeds rely on knowledge of the intrinsic parameters of each camera (e.g., focal length, optical center, radial distortion). This knowledge is typically not openly available and must be manually calculated through a detailed calibration procedure that is not scalable for at-home, remote testing. Moreover, even if intrinsic camera parameters are known, existing head pose estimation algorithms have predominantly been developed using adult data sets and may not generalize well to infants.

To address prior limits on explicitly modeling infant 3D head position, we apply principles from tau-coupling theory to estimate this information (Lee, 1998). This theory has been widely used in ecological psychology to explain the control of movement through the perception of affordances, and has been supported by empirical findings in adults, infants, and other species (Agvei et al., 2016; Lee, 1998; Regan & Hamstra, 1993; Wann, 1996; Yilmaz & Warren, 1995). Tau-coupling theory assumes that the distance of movement (defined as a motion gap) and the distance of external perceptual information (defined as an action gap) is amodal and invariant (Lee, 1998). That is, the tau of a motion-gap and the tau of an action-gap are intrinsically coupled and remain in constant proportion over a specified time frame.

We use tau-coupling theory to estimate changes in the infant's head pose using changes in the observable perceptual information. We conceptualize the change in the infant's left/right head direction as a movement gap change in the perceived change in ratio of the left to the right eye area as an action gap. Similarly, we conceptualize the change in the infant's up/down head direction as a movement gap and the change in the perceived height of the eyes as an action gap. Given that changes in the perceived eye area ratio and the perceived height of the eyes should be coupled with the actual change in the infant's horizontal and vertical head angle, these ratios can serve as a proxy for approximating changes in the infant's head position. We scale the infant's vertical and horizontal pupil/eye positions by these respective ratios, thus combining infant eye movement and head position information to estimate gaze direction.

Point-of-gaze estimation

In a final step, we apply a polynomial transfer function to map the infant's estimated point-of-gaze to coordinates on the screen following a four-point calibration, again applying principles from tau-coupling theory. During calibration, the infant is cued to look at animated objects at the far right, left, top, and bottom of the screen. According to tau-coupling theory, the change in the infant's perceived gaze from the

far-left to the far-right of the screen (action gap) should be proportional to the width of the screen (movement gap). Similarly, the change in the infant's perceived gaze from the bottom to the top of the screen should be directly proportional to the height of the screen. We use this information both as a boundary for determining when the infant is not looking at the screen and as a scaling factor to translate the infant's current point-of-gaze to screen coordinates using a simple polynomial transfer function. A six-frame moving average filter is also applied to the raw gaze signal to smooth over noise in fixations arising from frame-by-frame variations in video quality or lighting conditions. The moving average filter is reset when a gaze shift is detected. This approach aligns with prior lab-based eye-tracking algorithms (e.g., in Tobii systems; Olsen, 2012) and more recent webcam-based eye trackers that use moving average windows of up to ten samples (Aljaafreh et al., 2020; Kumar et al., 2008; Lewandowska, 2019).

Output measures

The csv output from OWLET saves estimates of infant's point-of-gaze on a frame-by-frame basis, providing the researcher with flexibility in operationalizing the repertoire of output measures. Currently, OWLET's post-processing pipelines are configured to automatically calculate the following gaze measures: total looking time to the screen (i.e., the sum of all looks to the screen during a video), the duration of the longest consecutive look (i.e., the maximum duration of looking to anywhere on the screen prior to looking away for 1 s or longer), and the total number of left/right gaze shifts across at least 1/6 of the screen width. We selected these output measures to align with common dependent measures used in prior infant visual attention tasks (e.g., Cuevas & Bell, 2014; Kraybill et al., 2019; Rose et al., 2001, 2012). Importantly, however, the annotated frame-by-frame csv output also provides infant gaze coordinates at a temporal resolution of 30 Hz. This allows researchers utilizing this open-source tool to tailor the output measures for their specific use cases.

To calculate looking time, OWLET is currently configured to set look onsets when the infant's point-of-gaze falls within the screen boundaries for 1 s or longer, and offsets are set when it falls outside of the screen boundaries for 1 second or longer. Gaze shifts are tagged based on changes in the infant's horizontal point-of-gaze that exceed a specific threshold. Currently, a relative threshold of at least 1/6 of the screen width over a period of 33 ms is used (equivalent to one frame, for a standard 30-fps video). A relative threshold is applied, rather than an absolute, velocity-based threshold, to account for infants tested on variable screen sizes; however, this threshold is modifiable and can be adjusted for different use cases. When the change in the infant's point of

gaze exceeds the set threshold, our algorithm first assesses whether the change is an aberration due to signal noise. To determine this, the Euclidean distance of the current point-of-gaze (n) to both the prior point-of-gaze ($n-1$) and the subsequent point-of-gaze ($n+1$) is calculated, similar to prior webcam-based eye tracking algorithms (e.g., Kumar et al., 2008). If the current point-of-gaze is closer to the subsequent point-of-gaze than the prior point-of gaze, a gaze shift is tagged in the output and the moving average filter is reset; otherwise, the aberrant gaze point is discarded, and the prior gaze point is used to interpolate the missing gaze point. In a final step, gaze shifts that are identified within short succession (less than 50 ms apart) are considered to reflect the same gaze shift and are thus merged (Olsen, 2012; Kumar et al., 2008).

Open-source availability

The source code for OWLET is freely available to download at <https://github.com/denisemw/OWLET>. Instructions for downloading OWLET and all relevant dependencies are found there as well. The source code of OWLET is accessible via GitHub, licensed under the GNU General Public License v.3, to ensure that users can freely use, share, and modify OWLET.

Experimental validation

Participants

The spatial accuracy of OWLET, external validity, and reliability relative to manual annotation of infant looking behavior was examined using videos collected as part of a large remote, longitudinal study of infant development. Data were collected from 127 infants when they were approximately 7 months of age. Five infants did not complete the task, two were excluded for significant distractions during the task, and ten were excluded due to poor video quality. In addition,

five videos did not meet the requirements of OWLET ($n = 2$, where only the infant’s eyes were visible; $n = 3$ where half of the infant’s face was in shadow). Thus, the final sample used to evaluate OWLET consisted of 105 infants (M age = 6.78 months, range = 5.57–8.33, $SD = 0.68$ months; $n = 41$ females). Sociodemographic characteristics for the final sample are presented in Fig. 3.

Procedures

Infants were seated with their primary caregiver during the visual attention task. Testing occurred either using computers ($n = 76$; 72%) or mobile devices ($n = 29$; 28%). Caregivers were asked to prop up the testing device and hold their infant during the study to ensure that the distance of the infant from the screen remained relatively constant throughout the study. Prior to testing, experimenters first instructed mothers on how to change their Zoom settings to hide the participant and experimenter videos. The experimenter also expanded the infant’s video feed to the maximum size and recorded the video (at 30+ fps). In addition, parents were asked to measure the distance they were sitting from the screen using an 18-inch tape measure that was included in a testing kit mailed to families prior to participation. The experimenter also collected information on the device used for testing to determine the screen size. This information was used to calculate the approximate visual angle of the screen for subjects.

Gaze calibration/validation Prior to initiating the visual attention task, the experimenter performed a calibration procedure where four objects were presented on the top, bottom, left, and right of the screen. The experimenter animated each of the objects one at a time to cue infants to attend to each calibration location, ensuring that the infant fixated at each cued location prior to animating the next object. This procedure is similar to lab-based calibration methods, which often require the experimenter to manually advance the calibration point after infants attend to it, given the inability to explicitly

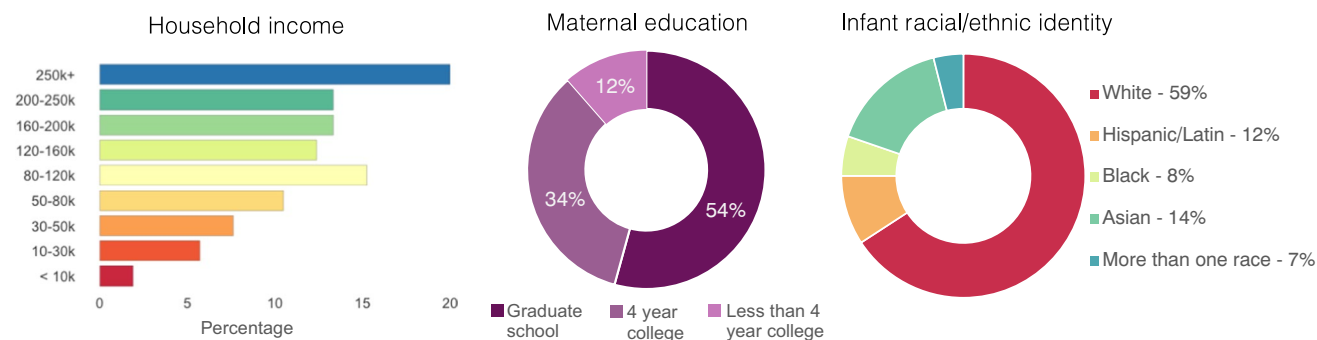


Fig. 3 Socio-demographic characteristics of the full experimental sample of infants

instruct infants to fixate on each location. The calibration procedure was repeated until infants visibly looked at all four objects. The recorded calibration video was subsequently used to validate the spatial accuracy of OWLET's gaze estimation (Fig. 4).

Visual attention assessment After calibration, infants watched a short, 80-s Sesame Street video (*Cecile - Up Down, In Out, Over and Under*), which has been used in prior work to assess individual differences in attention in similar-aged infants (Kraybill et al., 2019). Before beginning the Sesame Street video, the experimenter verified that the mother could clearly hear sound in a short test clip. They also instructed the mother to not interfere or redirect her baby's attention during the video.

Dependent measures

Looking time and gaze shifts Infants' total looking time to the screen, duration of the longest consecutive look to the screen, and left/right gaze shifts across the screen were measured using OWLET (see approach for details on how these automated output measures were configured). These variables were selected, as they are validated indices of infant attention in prior developmental literature (e.g., Colombo et al., 1991; Cuevas & Bell, 2014; Kraybill et al., 2019; Rose et al., 2001, 2012). In our analyses, we examined total looking time and the maximum look duration as both continuous measures and as dichotomous measures (short vs. long lookers) based on median splits of looking durations. Dichotomous measures were also included as variables of interest to smooth individual variability in infant looking times, following prior studies (Colombo et al., 1991; Cuevas & Bell, 2014; Rose et al., 2001, 2012). To validate the measures from OWLET, one experienced, expert coder naïve to the design of OWLET annotated infant's looking durations and left/right gaze shifts across the screen on a frame-by-frame basis in a randomly selected subsample of 50 infants using Datavyu (Datavyu Team, 2014). Inter-rater reliability was calculated for 20% of the manually coded videos, which indicated excellent reliability ($r = .97$, 95% CI [.95, .98]).

Infant orienting/regulatory capacity Parents reported on subjective measures of infant attention using the revised Infant Behavior Questionnaire – very short form (IBQ-R; Gartstein & Rothbart, 2003). In the current analysis, we focused on the orienting/regulatory capacity dimension of the IBQ, which is considered to reflect one of the earliest manifestations of self-regulation and executive attention in young infants (Rothbart et al., 2000; Sheese et al., 2008). Indeed, prior studies have shown that infant looking behavior is associated with parent-report measures of orienting/regulatory capacity, with shorter looking times generally predicting higher levels of orienting/regulatory capacity (Colombo & Cheatham, 2006; Gartstein et al., 2013; Hendry et al., 2018). Thus, we use orienting/regulatory capacity as a subjective measure of infant attentional control. We use this measure as a benchmark for external validity of the attention measures obtained using OWLET by testing the prediction that shorter looking times should be associated with higher parent-report measures of orienting/regulatory capacity.

Covariates

Prior work has indicated associations between family socioeconomic status (SES) and infant attention (e.g., Brandes-Aitken et al., 2019; Werchan et al., 2019). Thus, to control for potential SES effects when assessing the external validity of our attention measures, we included family income and maternal educational attainment as proxies for family SES. Both maternal educational attainment and family income were measured categorically (see Fig. 1).

Results

Spatial accuracy

The estimated diagonal visual angle of the testing screen varied across subjects from 22.16 to 76.91 ($M = 49.11$, $SD = 11.83$). The average x/y spatial offsets between OWLET's estimated point-of-gaze and the cued gaze location during the calibration video was calculated to estimate absolute spatial accuracy (in degrees visual angle). Results

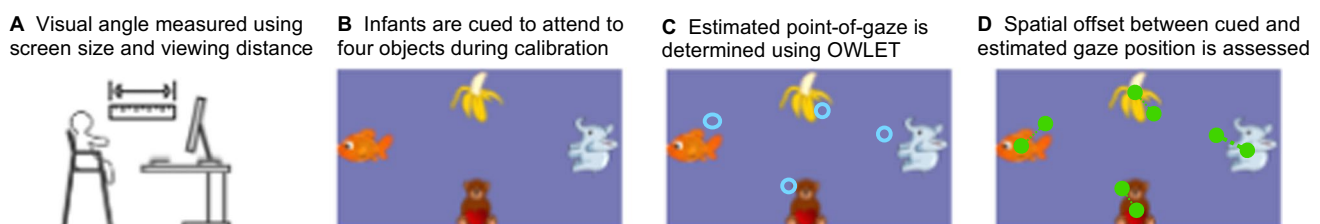


Fig. 4 Visual illustration of the procedures used to calculate spatial accuracy (the average offset between the cued and estimated gaze positions in degrees visual angle)

indicated that the mean absolute x/y calibration deviations were $3.36^{\circ}/2.67^{\circ}$ (SD of $1.89^{\circ}/1.55^{\circ}$) across subjects. Distributions of x/y calibration deviations are shown in Fig. 5. In addition, the absolute x/y calibration deviations were smaller for infants tested using smartphones ($M = 2.47^{\circ}/1.95^{\circ}$, $SD = 1.11^{\circ}/1.35^{\circ}$) relative to those tested using computer webcams ($M = 3.62^{\circ}/2.88^{\circ}$, $SD = 1.99^{\circ}/1.55^{\circ}$), $ps < .02$. Correlations indicated no associations between absolute x/y calibration deviations and infant age, $rs < .08$, $ps > .48$, total looking time, $rs < .20$, $ps > .08$, maximum look duration, $rs < .11$, $ps > .31$, or gaze shift rate, $rs < .05$, $ps > .66$.

We also applied linear mixed effects models, using the “lme4” package in R, to evaluate whether there were differences in mean absolute x/y calibration deviations (in degrees visual angle) by calibration point (“top”, “bottom”, “left”, and “right”). Separate models were used for horizontal and vertical accuracy. Results indicated that there was a significant difference in horizontal spatial accuracy, $F(3, 238.77) = 5.18$, $p < .001$, with the “top” calibration point showing significantly higher mean absolute x-deviations than both the “right” calibration point, $b = .035$, $p < .001$, and the “left” calibration point, $b = .035$, $p < .001$. There were no significant differences between calibration points when examining vertical spatial accuracy, $F(3, 232.24) = 2.37$, $p = .07$.

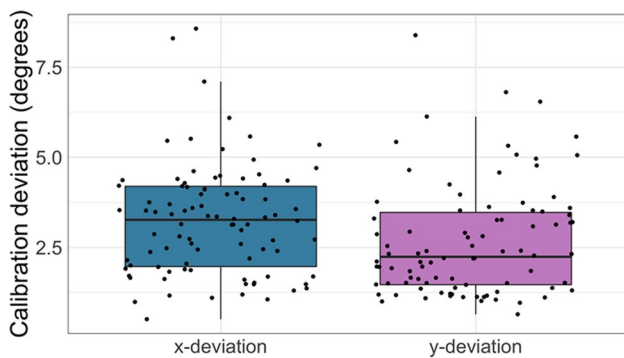


Fig. 5 Distributions of estimated spatial accuracy in degrees visual angle across subjects

Finally, given variation in the estimated visual angle of the testing screen, particularly for infants tested on smartphones relative to laptops, we also calculated the relative spatial accuracy (e.g., the absolute x/y calibration deviations divided by the x/y visual angles of the screen). Results indicated that the mean relative x/y calibration deviations were $0.08/0.11$ (SD of $0.04/0.06$). In addition, there were no differences in relative spatial accuracy for infants tested using smartphones ($M = .08/.10$, $SD = .03^{\circ}/.05$) compared to computer webcams ($M = .08/.11$, $SD = .04/.07$), $ps > .52$. Visual illustrations of the relative spatial accuracy across subjects at the 25th, 50th, and 75th percentiles are shown in Fig. 6.

Evaluation of gaze data by device type and sociodemographic characteristics

Prior to testing reliability in comparison to manual annotation and external validity, we evaluated whether there were systematic differences between data collected on mobile devices (smartphones or tablets) relative to computers (laptops or desktops). Two-sided independent samples t tests indicated that there were no differences between groups in the continuous measures of total looking time, maximum look duration, or gaze shift rate, all $ts < 1.20$, all $ps > .23$ (Fig. 7). Chi-square tests were also used to examine the dichotomous measures of total looking time and maximum look duration, which indicated no differences in the proportion of short relative to long lookers based on testing device for either total looking time, $\chi^2(1) = 1.88$, $p = .17$, or maximum look duration, $\chi^2(1) < .001$, $p = 1.00$.

Next, sociodemographic characteristics were explored for families who participated in the remote assessments using mobile devices or computers (Fig. 8). We observed significant differences in household income, $t(102) = 2.69$, $p = .01$, with families using mobile devices reporting mean incomes of approximately \$90,000 and families using computers reporting mean incomes of approximately \$150,000. In addition, a logistic regression examining maternal educational attainment (graduate degree or higher, four-year college degree, less than four-year college) as a predictor of testing device, with graduate degree or higher as a

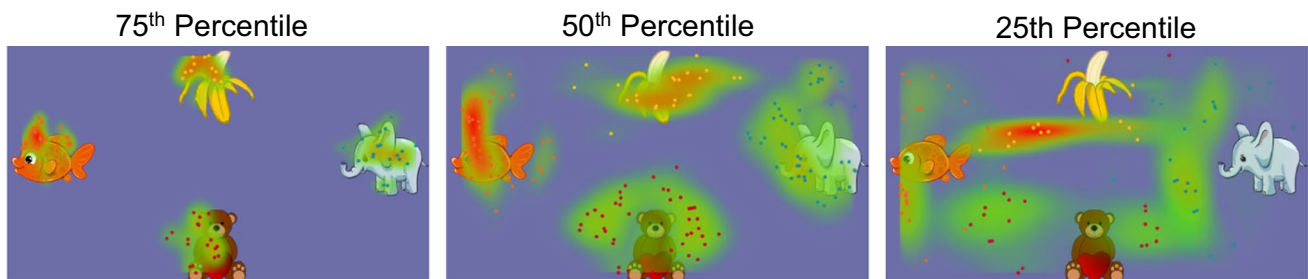


Fig. 6 Heatmap of the point-of-gaze estimated by OWLET relative to the centroid of each cued calibration point, split by spatial accuracy percentile groups

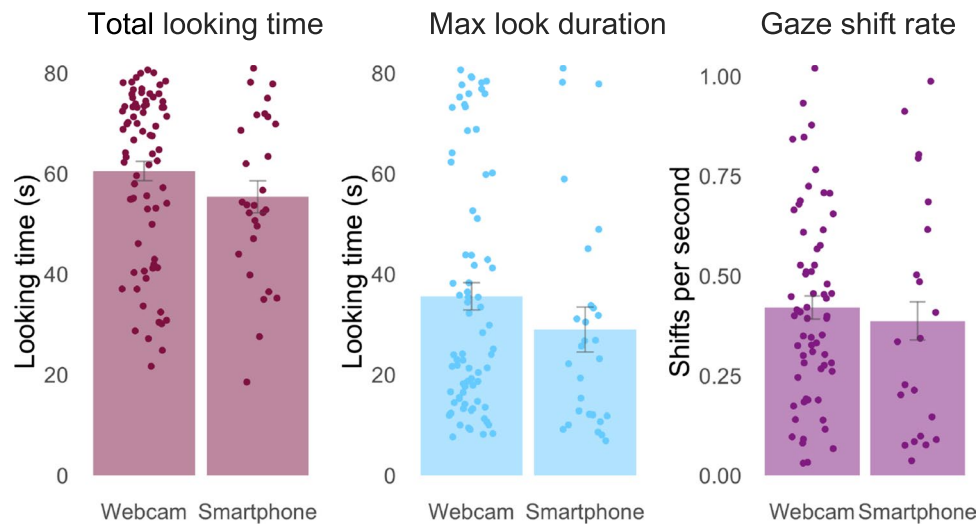
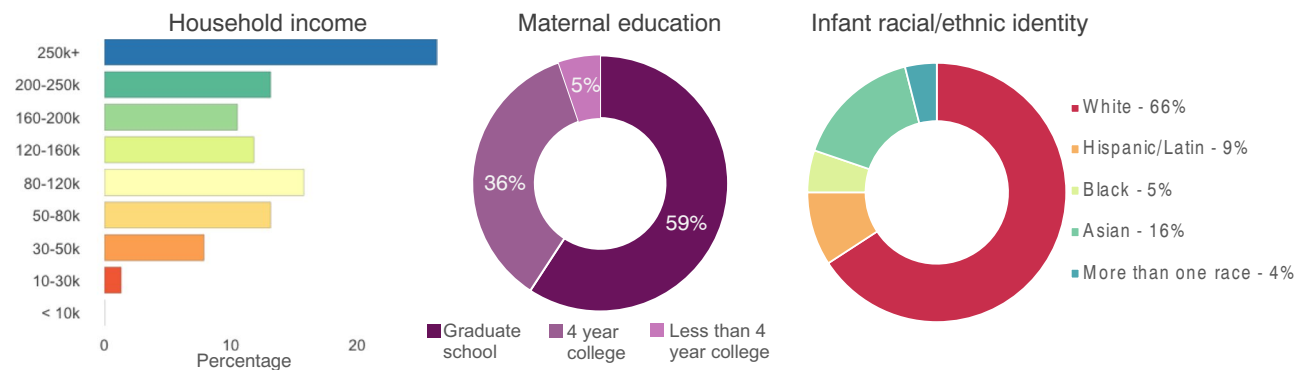


Fig. 7 Eye-tracking output measures separated by testing device (videos recorded using computer webcams in comparison to mobile devices)

Sociodemographic characteristics of families using laptops or desktops



Sociodemographic characteristics of families using mobile devices

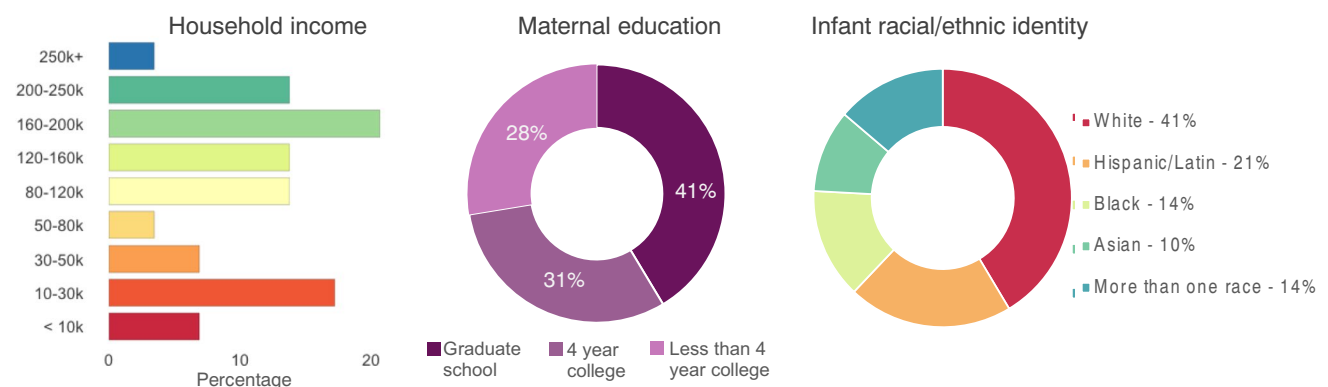


Fig. 8 Sociodemographic characteristics of families opting to use computers (top panel) or mobile devices (bottom panel) for study participation

reference, indicated that mothers who used mobile devices were over 6 times more likely to have less than a four-year college degree. $\beta = .55$, OR = 6.06, 95% CI [1.59, 26.32].

We then characterized the racial/ethnic identities of families who used mobile devices relative to computers. A logistic regression with White as a reference indicated that families

who used mobile devices were over four times as likely to identify as Black, $\beta = .38$, OR = 4.16, 95% CI [0.88, 20.07], over three times as likely to identify as Hispanic/Latin, $\beta = .42$, OR = 3.57, 95% CI [.99, 12.80], and over five times as likely to identify as more than one race/other, $\beta = .43$, OR = 5.56, 95% CI [1.09, 31.52].

We then evaluated whether there were systematic differences in eye-tracking quality based on infant race or ethnicity. Multiple linear regressions comparing OWLET output measures by race, controlling for infant age, indicated no differences in total looking time, maximum look duration, or shift rate for White infants relative to Asian, Black, or Hispanic/Latin infants, all $ps > .13$. Taken together, these analyses indicate equivalent eye tracking quality between videos recorded using mobile devices relative to computers, and across infants of different racial and ethnic backgrounds. Importantly, it also reveals that there is greater socioeconomic variability and diversity in racial/ethnic identities when families can use mobile devices to participate in research.

Reliability relative to manual-annotation

We next evaluated the reliability of looking durations and left/right gaze shifts estimated by OWLET relative to manual-annotation by human coders. We compared both overall looking time as well as the duration of the longest consecutive look, in addition to the total number of left/right gaze shifts, controlling for testing device type. Correlations between were excellent for all variables (Fig. 9): overall looking time, $r(49) = .97$, $p < .001$; maximum look duration, $r(49) = .99$, $p < .001$; gaze shift rate, $r(49) = .95$, $p < .001$.

We additionally examined the sensitivity of OWLET in identifying infants' left/right gaze shifts across the screen within ± 250 ms of those identified by human coders.

Sensitivity was calculated by the total number of left/right shifts correctly identified by OWLET within 250 ms relative to the total number of left/right shifts identified by human coders. To maximize power for these sensitivity analyses, we focused on infants with 15 or more left/right gaze shifts across the screen ($n = 30$, out of the 50 infant videos that were manually annotated). Results indicated excellent sensitivity ($M = .87$, $SD = .05$), with OWLET correctly identifying 87% of all human-identified left/right gaze shifts across the screen within 250 ms.

External validity

Relation with subjective attention measures We used linear regressions to examine external validity in comparison to maternal-report measures of infants' orienting/regulatory capacity, a temperamental measure of attentional control (Fig. 10). Controlling for infant age, maternal education, and family income, we observed significant associations between orienting/regulatory capacity and the continuous measure of overall looking time, $\beta = -.24$, $p = .03$, such that infants with higher orienting/regulatory capacity showed shorter looking times. There was only a trending association between orienting/regulatory capacity and the continuous measure of maximum look duration, $\beta = -.19$, $p = .08$. When examining dichotomous measures of looking durations, we observed significant effects of orienting/regulatory capacity on both total looking time, $\beta = -.26$, $p = .02$, and on maximum look duration, $\beta = -.25$, $p = .02$. There was no association between orienting/regulatory capacity and gaze shift rate, $\beta = -.09$, $p = .39$.

Relation with infant age Finally, age-related differences in the attention measures generated by OWLET were examined as an additional exploratory index of external validity, given prior findings indicating age-related declines in

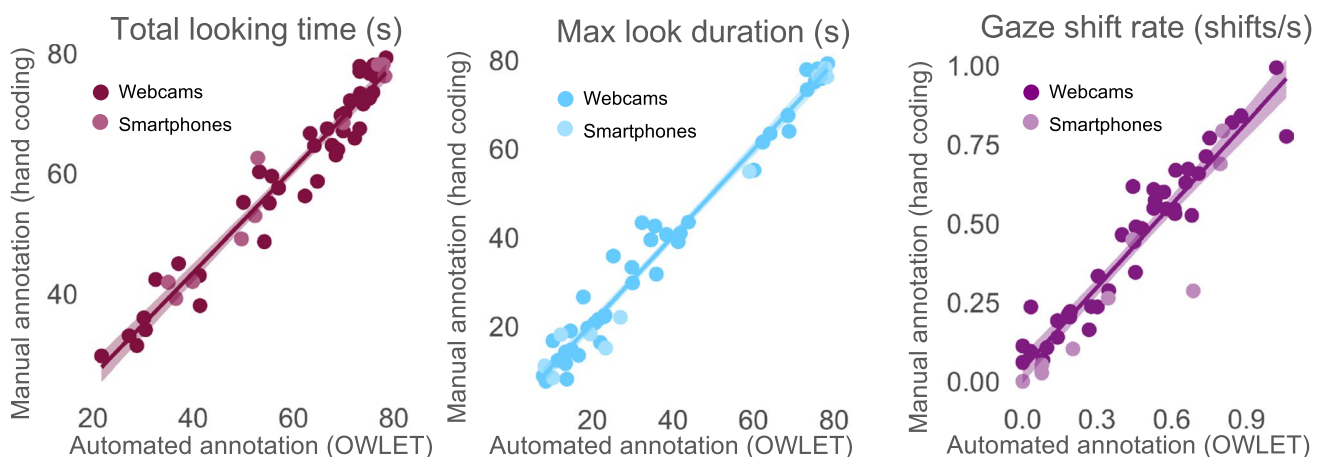


Fig. 9 Reliability of automated eye-tracking output measures using OWLET in comparison to manual annotation of infant looking behavior by expert coders

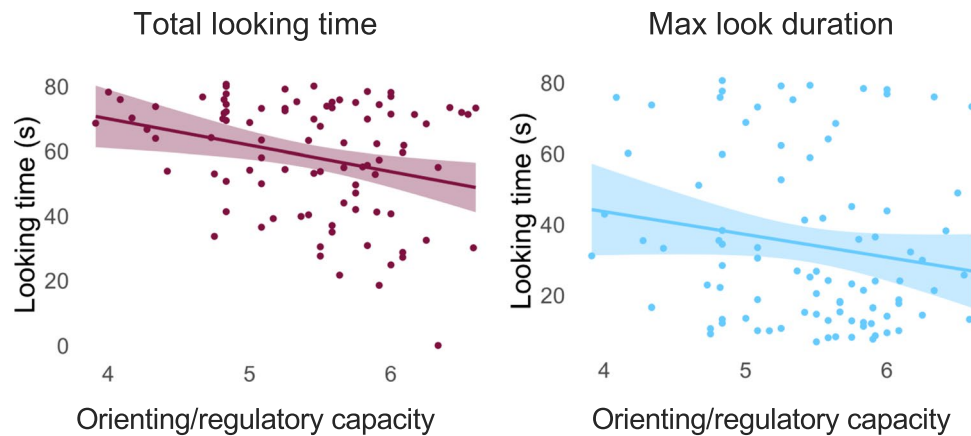


Fig. 10 Correlations between parental-report measures of infant attention using the orienting/regulatory capacity dimension of the IBQ-R and the automated task-based measures of attention produced by OWLET

look durations (Colombo & Mitchell, 1990; Colombo, 2001; Colombo et al., 2004). Note that the broader study from which the data for these analyses were drawn was designed to evaluate infants at 6 months of age. However, there was a moderate amount of variability around this target age ($M = 6.78$ months, $SD = 0.68$ months, range = [5.57, 8.33]). As such, to conduct exploratory analyses of age-related effects as an additional metric of external validity, we created two post hoc groups of younger and older infants by splitting the sample into subgroups of infants 1 SD below the mean age ($n = 20$ younger infants, M age = 5.87 months, $SD = 0.15$ months) and infants 1 SD above the mean age ($n = 18$ older infants, M age = 7.90 months, $SD = 0.28$ months). Linear regressions were then used to explore age as a predictor of total looking time, maximum look duration, and shift rate, controlling for orienting/regulatory capacity, family income, and maternal education (Fig. 11). We used both continuous measures and dichotomous measures (short vs. long lookers) of total looking time and maximum look duration.

When examining continuous measures of looking durations, age was only a trending predictor of total looking time, $\beta = -.35$, $p = .08$, and did not predict maximum look duration, $\beta = -.24$, $p = .28$. When examining looking durations as dichotomous variables, we observed a significant effect of age on total looking, $\beta = -.44$, $p = .03$, and a marginally significant effect of age on maximum look duration, $\beta = -.41$, $p = .05$. There were no significant age-related differences in shift rate, $\beta = -.31$, $p = .12$.

Discussion

There has been significant innovation in infant remote testing procedures over the past few years, accelerated in part as a result of COVID-19 pandemic-related testing restrictions (Gustafsson et al., 2021; Sheskin et al., 2020). Shifts to remote testing have included regularization of methods for

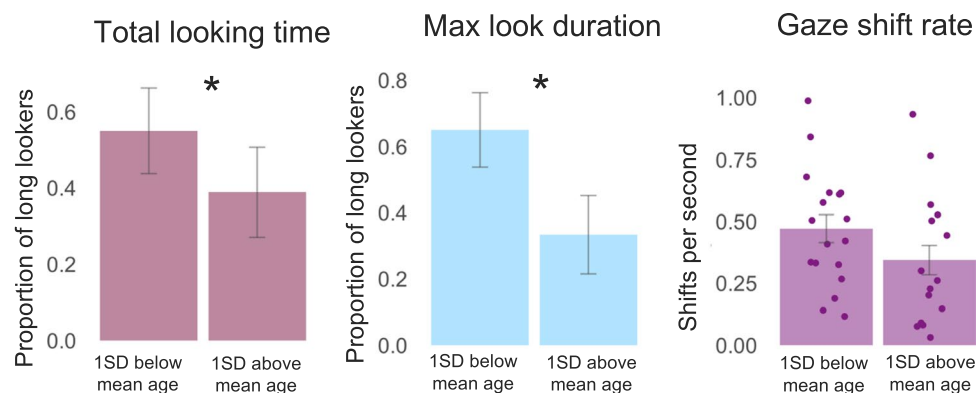


Fig. 11 Eye tracking automated output measures in younger relative to older infants

informed consent, survey administration, and adaptations of traditional infant paradigms for use in remote assessments (e.g., through Zoom; see Gustafsson et al., 2021). Online platforms specifically designed for infant research are also increasingly used (e.g., Scott & Schulz, 2017). Despite these innovations, online data acquisition largely necessitates time-consuming manual annotation by extensively trained coders. This constraint can be so costly as to be prohibitive, especially in large-scale studies, which are a major movement in developmental science today. For this reason, many population-based consortia studies of early human development to date, such as the Developing Human Connectome Project, have relied heavily on maternal questionnaire data. A lack of objective assessments to complement maternal-report measures risks confounding exposures (e.g., maternal depression) with outcomes (e.g., maternal report of infant self-regulation). Moreover, maternal-report is typically an unreliable indicator of whether a child's behavior is developmentally normative or not (Lord & Corsello, 1997; Wakschlag et al., 2005). Objective assessments are thus essential for capturing heterogeneity and key individual differences in developmental trajectories. This lofty goal, however, must be carefully balanced with competing demands on the feasibility and accessibility of conducting objective assessments at scale.

Development and experimental validation of OWLET

To address limitations in remote infant testing, we developed a novel methodology integrating computer vision, machine learning, and ecological psychology to estimate infants' gaze behavior from videos recorded using computer webcams or smartphones. This approach led to reliable estimates of infant gaze behavior across a variety of testing contexts. In addition, we found robust associations with parental-report measures of infants' attentional control, as well as age-related effects that match expectations about maturation of visual attention. Corroboration of priors provides further evidence of the reliability of this new methodology. We also found high correlation with manually coded estimates of total looking time and the duration of the longest single look. Overall, presented results indicate both high internal reliability of OWLET relative to costly and time-intensive manual coding, as well as robust external validity. With lifted constraints on manual coding, it is practical that this technique could be implemented to supplement large-scale, online developmental studies.

An important consideration in the development of OWLET was ensuring high efficacy regardless of whether testing occurred using a mobile device or computer, as well as when testing families from diverse racial and ethnic backgrounds. We verified that the output produced by OWLET

was robust regardless of whether testing occurred using a computer or a mobile device, and regardless of infant race or ethnicity. We also observed that providing the opportunity for families to participate using mobile devices was associated with substantially increased socioeconomic and racial/ethnic diversity in our sample. These findings substantiate our postulation that this tool may support greater inclusivity and accessibility for research participation in families from diverse socioeconomic, geographic, and cultural backgrounds. Importantly, the implementation of this tool has potential to greatly expand opportunities for remote infant testing in a variety of settings, including rural contexts or clinical settings.

Although we validated OWLET using a visual attention task, this tool can be used to test other complex aspects of infant cognition, such as face processing (Frank et al., 2009; Liu et al., 2011), object perception (Johnson et al., 2003; Johnson et al., 2004), memory (Richmond & Nelson, 2009; Sanders & Johnson, 2021), rule learning and early executive functions (Wass et al., 2011; Werchan et al., 2015; Werchan & Amso, 2020; Werchan & Amso, 2021), and risk for neurodevelopmental disorders, such as autism (Gluga et al., 2015; Jones & Klin, 2013). Importantly, OWLET records infant gaze coordinates at a temporal resolution of 30 Hz, which is output as a plain-text csv file that can be integrated with stimulus timing information. This allows researchers to flexibly tailor the measures of interest according to their specific needs. The translation of infant videos into frame-by-frame gaze coordinates also facilitates the secure and safe storage and transfer of data, given reduced file sizes and the removal of identifying information in videos. As such, OWLET is well-suited to support and accelerate large-scale online data collection in infants, without the time, cost, and privacy considerations entailed by manual coding of video data.

Comparison with existing platforms

A major difference between OWLET and other webcam eye trackers' approach to gaze classification is that OWLET is grounded in ecological psychology principles and relies only on the observable perceptual information to classify infants' gaze. In contrast, other infant gaze platforms such as iCatcher (Erel et al., 2022a) apply deep learning algorithms to classify gaze, using human-labeled images of infants looking to the left, right, or away as training input. While these algorithms can perform powerfully, they are also dependent on the quality and breadth of the training set. For instance, training on image sets generated from laboratory experiments may not generalize well to the range of non-ideal or unstandardized conditions encountered when testing in the home, such as poor lighting or variations in infant positioning relative to the camera. Moreover, an

additional contribution of OWLET is that this platform also provides information about infants' x/y gaze coordinates on a frame-by-frame basis. Currently, other infant gaze tracking platforms only support classification of looks to the left, right, and away from the screen. As such, OWLET provides greater flexibility in allowing researchers to define their own areas of interest. Finally, OWLET has also been tested and validated on smartphones, whereas existing platforms have only been validated using computer or tablet webcams. This is a critical consideration particularly in regard to increasing the accessibility, inclusivity, and equity of developmental research, given sociodemographic inequities in access to smartphones relative to laptops (Perrin & Atske, 2021; Vogels, 2021).

Several commercial webcam eye trackers have also been introduced in recent years (e.g., Finger et al., 2017; Lewandowska, 2019), which have been coopted for infant studies with some success (Bánki et al., 2022). However, the majority of commercial webcam eye-tracking software has not been directly validated for psychological research through peer review, nor have these platforms been explicitly designed to accommodate infant research. A major benefit of OWLET relative to commercial webcam eye-tracking software is that it is open-source and freely available to researchers. In addition, OWLET also supports post hoc processing of gaze data, whereas the majority of existing commercial platforms for webcam eye tracking are designed to be used concurrently. Finally, OWLET was explicitly designed and validated for use in infant subjects and shows robust performance across a variety of testing contexts.

Implications for supporting accessibility and ecological validity of developmental science

Our tool may also have important applications for efforts towards making developmental research more equitable and inclusive. Events occurring in the context of the COVID-19 pandemic have magnified existing structural and systemic inequities in our society, with developmental science being no exception (Nketia et al., 2021). Increased attention to these systemic issues creates opportunity for developmental scientists to address the multiple pathways through which traditional approaches have contributed to bias and inequity in our study of human development. While these issues go far beyond the scope of this paper, the introduction of new methods to help reduce structural barriers to participate in research is one avenue towards addressing these larger issues. For instance, tools like OWLET might increase the ease of recruiting and testing large samples of families from historically excluded communities; it also provides opportunities for families who live outside the proximity of major

research universities to participate in developmental studies in their own home.

Indeed, in our own data, we observed that our entire sample was skewed towards higher SES families, with 20% of the sample reporting incomes of greater than \$250k and only 8% reporting incomes of less than \$30K per year. These statistics differ greatly from the distribution of income in the broader U.S. population, in which only 6% of families report incomes of greater than \$250k and nearly 20% of families report incomes of less than \$30K (Donovan et al., 2021). Importantly, however, we observed substantial differences in SES distributions of families tested on laptops relative to smartphones (Fig. 8). In particular, while families tested using laptops were greatly skewed towards higher SES, we found that the demographics of families tested using smartphones were much closer to the demographics of the broader US population, with less than 5% reporting incomes greater than \$250k and approximately 25% reporting incomes of less than \$30K. These findings support our proposition that testing families on mobile devices using OWLET may afford greater inclusivity in developmental research.

Methods to bring research out of the laboratory and into the home are also relevant for ecological validity. Oftentimes, tightly controlled testing contexts do not reflect the complex, visually rich environments that infants' experience in their unique ecological niches (Werchan & Amso, 2017). This mismatch may contribute to incorrect inferences surrounding child development. To accurately capture heterogeneity and individual differences in developmental trajectories, it is important to ensure that studies reflect the unique environments experienced by the developing child. However, competing demands of balancing ecological validity with scientific precision is an inherent challenge to this goal. The application of new methods that support testing in more ecologically valid contexts, while affording high measurement precision, is essential to progress in this area.

Limitations and constraints

Currently, OWLET is developed for videos in which the majority of the infant's face is visible. It is not configured for videos in which the infant's face is obscured (e.g., videos where only the infant's eyes are visible, or videos where half of the infant's face is in shadow). An additional limitation is that when more than one face is detected, OWLET is currently configured to select the lower face, given that the majority of infants are positioned lower than their caregivers; however, this could lead to incorrect detection of the infant's face at times. In future iterations, this issue could be addressed using a face extractor trained to exclusively extract infant faces, such as the infant face detector implemented in iCatcher+, a method posted as a preprint that has not yet been peer reviewed (Erel et al., 2022b). Another

consideration is that our algorithm for dynamically calculating an appropriate threshold for pupil extraction may perform poorly for infants with very light or dark eyes that have significant reflections from external light, which are also common issues with specialized lab-based eye trackers (Hessels et al., 2015). In our platform, the threshold for pupil extraction can be manually modified to improve performance. We found that a pixel value of ~30–50 typically worked well for thresholding the iris from the rest of the image in very dark eyes with light reflections, and a pixel value of ~80–110 typically worked well for thresholding the iris in very light eyes. However, future iterations may reduce experimenter overhead by developing novel algorithms to dynamically configure appropriate threshold values for pupil extraction.

Finally, while OWLET is a significant advance from current platforms that only classify looks to the left, right, or away from the screen, OWLET still performs below the accuracy level of most lab-based eye trackers. Indeed, we observed that the average calibration offset was equivalent to approximately 10% of the screen width/height across devices. Given the fairly diffuse spatial accuracy, this tool is not recommended for analysis of more fine-grained gaze information (e.g., eye movements while reading text, looking to a person's eyes vs. mouth). Rather, OWLET is primarily recommended for analyzing infants' point-of-gaze in larger regions of interest (e.g., quadrants of the screen). An additional limitation of OWLET is the application of a six-frame moving average filter (equivalent to 200 ms of recorded data) to smooth the raw fixation signal. While this filter improves data quality by smoothing over noise arising from frame-by-frame fluctuations in lighting or video quality in remote recordings, it also limits the spatiotemporal precision of this platform. These methodological limitations should be carefully considered during the design and interpretation of infant eye-tracking studies (Oakes, 2012; Wass et al., 2014). Further testing and validation of infant webcam eye tracking is critical for increasing insight into the possibilities and limitations of remote infant testing.

Conclusions

In sum, here we introduce a novel open-source methodology for estimating gaze during remote online experiments with infants. Our novel gaze analysis platform fills a significant gap in the current tools available for scalable online data collection, particularly for testing young infants and toddlers. This tool enables rapid data collection and coding of cognitive processes in more ecologically valid environments. Importantly, remote research affords easier access to greater sociodemographic and geographic diversity when testing participants, in addition to lowering the time and

cost investments for families to participate in studies. We believe the approach presented here broadens the possibilities for rapid, scalable online data acquisition. This provides a significant step towards helping ensure that developmental science accurately reflects the diverse, intersectional environments occupied by infants and children and will help garner a more precise understanding of the drivers and origins of the human mind.

Author contributions DMW conceptualized and developed the eye-tracking methodology. NHB designed the infant remote assessment. DMW analyzed the data and wrote the manuscript, with critical input and revisions by MET and NHB.

Funding This work was funded by NIH R01MH125870 (NHB) and by NIH R01MH126468 (MET).

References

- Agyei, S. B., Van der Weel, F. R., & Van der Meer, A. L. (2016). Development of visual motion perception for prospective control: brain and behavioral studies in infants. *Frontiers in Psychology*, 7, 100. <https://doi.org/10.3389/fpsyg.2016.00100>
- Aljaafreh, A., Alaqdash, M., Al-Oudat, N., Abukhait, J., & Saleh, M. (2020). A low-cost webcam-based eye tracker and saccade measurement system. *International Journal of Circuits, Systems and Signal Processing*, 14. <https://doi.org/10.46300/9106.2020.14.16>
- Amso, D., Haas, S., & Markant, J. (2014). An eye tracking investigation of developmental change in bottom-up attention orienting to faces in cluttered natural scenes. *PLoS ONE*, 9(1). <https://doi.org/10.1371/journal.pone.0085701>
- Aslin, R. N. (2007). What's in a look? *Developmental Science*, 10(1), 48–53.
- Bánki, A., de Eccher, M., Falschlehner, L., Hoehl, S., & Markova, G. (2022). Comparing online webcam- and laboratory-based eye-tracking for the assessment of infants' audio-visual synchrony perception. *Frontiers in Psychology*, 12, 733933. <https://doi.org/10.3389/fpsyg.2021.733933>
- Bradski, G. (2000). The openCV library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, 25(11), 120–123.
- Brandes-Aitken, A., Braren, S., Swingler, M., Voegtline, K., & Blair, C. (2019). Sustained attention in infancy: A foundation for the development of multiple aspects of self-regulation for children in poverty. *Journal of Experimental Child Psychology*, 184, 192–209.
- Chouinard, B., Scott, K., & Cusack, R. (2019). Using automatic face analysis to score infant behaviour from video collected online. *Infant Behavior & Development*, 54, 1–12.
- Colombo, J. (2001). The Development of Visual Attention in Infancy. *Annual Review of Psychology*, 52(1), 337–367. <https://doi.org/10.1146/annurev.psych.52.1.337>
- Colombo, J., & Mitchell, D. W. (1990). Individual differences in early visual attention: Fixation time and information processing. In J. Colombo & J. W. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, prediction* (pp. 193–227). Lawrence Erlbaum Associates Inc.
- Colombo, J., & Cheatham, C. L. (2006). The emergence and basis of endogenous attention in infancy and early childhood. *Advances in Child Development and Behavior*, 34, 283–322.
- Colombo, J., Mitchell, D. W., Coldren, J. T., & Freeseaman, L. J. (1991). Individual differences in infant visual attention: are short lookers

- faster processors or feature processors? *Child Development*, 62(6), 1247–1257.
- Colombo, J., Shaddy, D. J., Richman, W. A., Maikranz, J. M., & Blaga, O. M. (2004). The Developmental Course of Habituation in Infancy and Preschool Outcome. *Infancy*, 5(1), 1–38. https://doi.org/10.1207/s15327078in0501_1
- Cuevas, K., & Bell, M. A. (2014). Infant attention and early childhood executive function. *Child Development*, 85(2), 397–404.
- Datavyu Team. (2014). Datavyu: A video coding tool. *Datavary Project, New York University*. URL <http://www.datavyu.org>.
- Donovan, S. A., Labonte, M., Dalaker, J., & Romero, P. D. (2021). The U.S. income distribution: Trends and issues. *Congressional Research Service*. Retrieved August 29, 2022, from <https://fas.org/sgp/crs/misc/R44705.pdf>
- Erel, Y., Potter, C. E., Jaffe-Dax, S., Lew-Williams, C., & Bermano, A. H. (2022a). iCatcher: A neural network approach for automated coding of young children's eye movements. *Infancy*, 27(4), 765–779. <https://doi.org/10.1111/infa.12468>
- Erel, Y., Shannon, K. A., Scott, K. M., Cao, P., Tan, X., Hart, P. K., ... Liu, S. (2022b). iCatcher+: Robust and automated annotation of infant gaze from videos collected in the lab and online. *iCatcher+*. <https://doi.org/10.31234/osf.io/qp97k>
- Finger, H., Goeke, C., Diekamp, D., Standvoß, K., & König, P. (2017). LabVanced: a unified JavaScript framework for online studies. In: International Conference on Computational Social Science (Cologne).
- Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, 110(2), 160–170.
- Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., Hamlin, J. K., Hannon, E. E., Kline, M., Levelt, C., Lew-Williams, C., Nazzi, T., Panneton, R., Rabagliati, H., Soderstrom, M., Sullivan, J., Waxman, S., & Yurovsky, D. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy: The Official Journal of the International Society on Infant Studies*, 22(4), 421–435.
- Gartstein, M. A., & Rothbart, M. K. (2003). Studying infant temperament via the Revised Infant Behavior Questionnaire. *Infant Behavior & Development*, 26(1), 64–86.
- Gartstein, M. A., Bridgett, D. J., Young, B. N., Panksepp, J., & Power, T. (2013). Origins of Effortful Control: Infant and Parent Contributions. *Infancy*, 18(2), 149–183. <https://doi.org/10.1111/j.1532-7078.2012.00119.x>
- Gliga, T., Bedford, R., Charman, T., Johnson, M. H., & BASIS Team. (2015). Enhanced Visual Search in Infancy Predicts Emerging Autism Symptoms. *Current Biology*, 25(13), 1727–1730.
- Gustafsson, H. C., Young, A. S., Stamos, G., Wilken, S., Brito, N. H., Thomason, M. E., Graham, A., Nigg, J. T., & Sullivan, E. L. (2021). Innovative methods for remote assessment of neurobehavioral development. *Developmental Cognitive Neuroscience*, 52, 101015.
- Hendry, A., Jones, E. J. H., Bedford, R., Gliga, T., Charman, T., & Johnson, M. H. (2018). Developmental change in look durations predicts later effortful control in toddlers at familial risk for ASD. *Journal of Neurodevelopmental Disorders*, 10(1), 3. <https://doi.org/10.1186/s11689-017-9219-4>
- 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.
- Johnson, S. P., Amso, D., & Slemmer, J. A. (2003). Development of object concepts in infancy: Evidence for early learning in an eye-tracking paradigm. *Proceedings of the National Academy of Sciences of the United States of America*, 100(18), 10568–10573.
- Johnson, S. P., Slemmer, J. A., & Amso, D. (2004). Where infants look determines how they see: Eye movements and object perception performance in 3-month-Olds. *Infancy*, 6(2), 185–201.
- 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(7480), 427–431. <https://doi.org/10.1038/nature12715>
- King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research: JMLR*, 10, 1755–1758.
- Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, location, location: development of spatiotemporal sequence learning in infancy. *Child Development*, 78(5), 1559–1571.
- Kraybill, J. H., Kim-Spoon, J., & Bell, M. A. (2019). Infant Attention and Age 3 Executive Function. *The Yale Journal of Biology and Medicine*, 92(1), 3–11.
- Kumar, M., Klingner, J., Puranik, R., Winograd, T., & Paepcke, A. (2008). Improving the accuracy of gaze input for interaction. *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, 65–68.
- Lee, D. N. (1998). Guiding Movement by Coupling Taus. *Ecological Psychology: A Publication of the International Society for Ecological Psychology*, 10(3–4), 221–250.
- Lewandowska, B. (2019). *RealEye Eye-tracking system Technology Whitepaper*. Retrieved August 29, 2022, from <https://support.realeye.io/realeye-accuracy/>
- Liu, S., Quinn, P. C., Wheeler, A., Xiao, N., Ge, L., & Lee, K. (2011). Similarity and difference in the processing of same- and other-race faces as revealed by eye tracking in 4- to 9-month-olds. *Journal of Experimental Child Psychology*, 108(1), 180–189.
- Lord, C., & Corsello, C. (1997). Diagnostic instruments in autism spectrum disorders. *Handbook of Autism and Pervasive Developmental Disorders*, 2, 460–483.
- Nketia, J., Amso, D., & Brito, N. H. (2021). Towards a more inclusive and equitable developmental cognitive neuroscience. *Developmental Cognitive Neuroscience*, 52, 101014.
- 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. (2012). The Tobii I-VT fixation filter. *Tobii Technology*, 21, 4–19. Retrieved August 29, 2022, from <https://www.tobii.com/siteassets/tobii-pro/learn-and-support/analyze/how-do-we-classify-eye-movements/tobii-pro-i-vt-fixation-filter.pdf>
- Perrin, A., & Atske, S. (2021). 7% of Americans don't use the Internet. Who are they. *Pew Research Center*. Retrieved August 29, 2022, from <https://www.pewresearch.org/fact-tank/2021/04/02/7-of-americans-dont-use-the-internet-who-are-they/>
- Regan, D., & Hamstra, S. J. (1993). Dissociation of discrimination thresholds for time to contact and for rate of angular expansion. *Vision Research*, 33(4), 447–462. [https://doi.org/10.1016/0042-6989\(93\)90252-R](https://doi.org/10.1016/0042-6989(93)90252-R)
- Richmond, J., & Nelson, C. A. (2009). Relational memory during infancy: evidence from eye tracking. *Developmental Science*, 12(4), 549–556.
- Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2001). Attention and recognition memory in the 1st year of life: a longitudinal study of preterm and full-term infants. *Developmental Psychology*, 37(1), 135–151.
- Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2012). Implications of infant cognition for executive functions at age 11. *Psychological Science*, 23(11), 1345–1355.
- Rothbart, M. K., Derryberry, D., & Hershey, K. (2000). Stability of temperament in childhood: Laboratory infant assessment to parent report at seven years. In V. J. Molfese & D. L. Molfese (Eds.), *Temperament and personality development across the life span* (pp. 85–119). Hillsdale, NJ: Erlbaum.

- Sanders, A. J., & Johnson, S. P. (2021). Indexing Early Visual Memory Durability in Infancy. *Child Development, 92*(2), e221–e235.
- Scott, K., & Schulz, L. (2017). Lookit (part 1): A new online platform for developmental research. *Open Mind, 1*(1), 4–14.
- Sheese, B. E., Rothbart, M. K., Posner, M. I., White, L. K., & Fraundorf, S. H. (2008). Executive attention and self-regulation in infancy. *Infant Behavior & Development, 31*(3), 501–510.
- Sheskin, M., Scott, K., Mills, C. M., Bergelson, E., Bonawitz, E., Spelke, E. S., Fei-Fei, L., Keil, F. C., Gweon, H., Tenenbaum, J. B., Jara-Ettinger, J., Adolph, K. E., Rhodes, M., Frank, M. C., Mehr, S. A., & Schulz, L. (2020). Online Developmental Science to Foster Innovation, Access, and Impact. *Trends in Cognitive Sciences, 24*(9), 675–678.
- Vogels, E. A. (2021). Digital divide persists even as Americans with lower incomes make gains in tech adoption. *Pew Research Center, June, 22*. Retrieved August 29, 2022, from <https://www.pewresearch.org/fact-tank/2021/06/22/digital-divide-persists-even-as-americans-with-lower-incomes-make-gains-in-tech-adoption/>
- Wakschlag, L. S., Leventhal, B. L., Briggs-Gowan, M. J., Danis, B., Keenan, K., Hill, C., Egger, H. L., Cicchetti, D., & Carter, A. S. (2005). Defining the “Disruptive” in Preschool Behavior: What Diagnostic Observation Can Teach Us. *Clinical Child and Family Psychology Review, 8*(3), 183–201.
- Wann, J. P. (1996). Anticipating arrival: Is the tau margin a specious theory? *Journal of Experimental Psychology: Human Perception and Performance, 22*(4), 1031–1048. <https://doi.org/10.1037/0096-1523.22.4.1031>
- Wass, S., Porayska-Pomsta, K., & Johnson, M. H. (2011). Training attentional control in infancy. *Current Biology, 21*(18), 1543–1547.
- 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/inf.12055>
- Werchan, D. M., & Amso, D. (2017). A novel ecological account of prefrontal cortex functional development. *Psychological Review, 124*(6), 720–739.
- Werchan, D. M., & Amso, D. (2020). Top-down knowledge rapidly acquired through abstract rule learning biases subsequent visual attention in 9-month-old infants. *Developmental Cognitive Neuroscience, 42*, 100761.
- Werchan, D. M., & Amso, D. (2021). All contexts are not created equal: Social stimuli win the competition for organizing reinforcement learning in 9-month-old infants. *Developmental Science, https://doi.org/10.1111/desc.13088*
- Werchan, D. M., Collins, A. G. E., Frank, M. J., & Amso, D. (2015). 8-month-old infants spontaneously learn and generalize hierarchical rules. *Psychological Science, 26*(6), 805–815.
- Werchan, D. M., Collins, A. G. E., Frank, M. J., & Amso, D. (2016). Role of Prefrontal Cortex in Learning and Generalizing Hierarchical Rules in 8-Month-Old Infants. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 36*(40), 10314–10322.
- Werchan, D. M., Lynn, A., Kirkham, N. Z., & Amso, D. (2019). The emergence of object-based visual attention in infancy: A role for family socioeconomic status and competing visual features. *Infancy, 24*(5), 752–767.
- Yilmaz, E. H., & Warren, W. H. (1995). Visual control of braking: A test of the tau hypothesis. *Journal of Experimental Psychology: Human Perception and Performance, 21*(5), 996–1014. <https://doi.org/10.1037/0096-1523.21.5.996>

Open practices statement Source code for OWLET, and instructions for downloading all dependencies, is available at <https://github.com/denisemw/OWLET>. This study was not preregistered.

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

Springer Nature or its licensor 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.