Gaze in Action: Head-mounted Eye Tracking of Children's Dynamic Visual Attention During Naturalistic Behavior

Lauren K. Slone[1], **Drew H. Abney**[1], **Jeremy I. Borjon**[1], **Chi-hsin Chen**[2], **John M. Franchak**[3], **Daniel Pearcy**[1], **Catalina Suarez-Rivera**[1], **Tian Linger Xu**[1], **Yayun Zhang**[1], **Linda B. Smith**[1], **Chen Yu**[1]

[1]Department of Psychological and Brain Sciences, Indiana University

[2]Department of Otolaryngology-Head and Neck Surgery, The Ohio State University

[3]Department of Psychology, University of California, Riverside

Abstract

Young children's visual environments are dynamic, changing moment-by-moment as children physically and visually explore spaces and objects and interact with people around them. Head-mounted eye tracking offers a unique opportunity to capture children's dynamic egocentric views and how they allocate visual attention within those views. This protocol provides guiding principles and practical recommendations for researchers using head-mounted eye trackers in both laboratory and more naturalistic settings. Head-mounted eye tracking complements other experimental methods by enhancing opportunities for data collection in more ecologically valid contexts through increased portability and freedom of head and body movements compared to screen-based eye tracking. This protocol can also be integrated with other technologies, such as motion tracking and heart-rate monitoring, to provide a high-density multimodal dataset for examining natural behavior, learning, and development than previously possible. This paper illustrates the types of data generated from head-mounted eye tracking in a study designed to investigate visual attention in one natural context for toddlers: free-flowing toy play with a parent. Successful use of this protocol will allow researchers to collect data that can be used to answer questions not only about visual attention, but also about a broad range of other perceptual, cognitive, and social skills and their development.

Keywords

Behavior; Issue 141; Head-mounted eye tracking; egocentric vision; development; infant; toddler; visual attention

Correspondence to: Lauren K. Slone at laurenkslone@gmail.com.

Video Link
The video component of this article can be found at https://www.jove.com/video/58496/

Disclosures
The authors declare that they have no competing or conflicting interests.

Introduction

The last several decades have seen growing interest in studying the development of infant and toddler visual attention. This interest has stemmed in large part from the use of looking time measurements as a primary means to assess other cognitive functions in infancy and has evolved into the study of infant visual attention in its own right. Contemporary investigations of infant and toddler visual attention primarily measure eye movements during screen-based eye-tracking tasks. Infants sit in a chair or parent's lap in front of a screen while their eye movements are monitored during the presentation of static images or events. Such tasks, however, fail to capture the dynamic nature of natural visual attention and the means by which children's natural visual environments are generated - active exploration.

Infants and toddlers are active creatures, moving their hands, heads, eyes, and bodies to explore the objects, people, and spaces around them. Each new development in body morphology, motor skill, and behavior - crawling, walking, picking up objects, engaging with social partners - is accompanied by concomitant changes in the early visual environment. Because what infants do determines what they see, and what they see serves for what they do in visually guided action, studying the natural development of visual attention is best carried out in the context of natural behavior[1].

Head-mounted eye trackers (ETs) have been invented and used for adults for decades[2,3]. Only recently have technological advances made head-mounted eye-tracking technology suitable for infants and toddlers. Participants are outfitted with two lightweight cameras on the head, a scene camera facing outward that captures the first person perspective of the participant and an eye camera facing inward that captures the eye image. A calibration procedure provides training data to an algorithm that maps as accurately as possible the changing positions of the pupil and corneal reflection (CR) in the eye image to the corresponding pixels in the scene image that were being visually attended. The goal of this method is to capture both the natural visual environments of infants and infants' active visual exploration of those environments as infants move freely. Such data can help to answer questions not only about visual attention, but also about a broad range of perceptual, cognitive, and social developments[4,5,6,7,8]. The use of these techniques has transformed understandings of joint attention[7,8,9], sustained attention[10], changing visual experiences with age and motor development[4,6,11], and the role of visual experiences in word learning[12]. The present paper provides guiding principles and practical recommendations for carrying out head-mounted eye-tracking experiments with infants and toddlers and illustrates the types of data that can be generated from head-mounted eye tracking in one natural context for toddlers: free-flowing toy play with a parent.

Protocol

This tutorial is based on a procedure for collecting head-mounted eye-tracking data with toddlers approved by the Institutional Review Board at Indiana University. Informed parental consent was obtained prior to toddlers' participation in the experiment.

1. Preparation for the Study

1. **Eye-Tracking Equipment**. Select one of the several head-mounted eye-tracking systems that are commercially available, either one marketed as specifically for children or modify the system to work with a custom-made infant cap, for instance as shown in Figures 1 and 2. Ensure that the eye-tracking system has the necessary features for testing infants and/or toddlers by following these steps:

    1. Select a scene camera that is adjustable in terms of positioning and has a wide enough angle to capture a field of view appropriate for addressing the research questions. To capture most of toddler's activity in a free-play setting like that described here, select a camera that captures an at least 100 degree diagonal field of view.

    2. Select an eye camera that is adjustable in terms of positioning and has an infrared LED either built into the camera or adjacent to the camera and positioned in such a way that the eye's cornea will reflect this light. Note that some eye-tracking models have fixed positioning, but models that afford flexible adjustments are recommended.

    3. Choose an eye-tracking system that is as unobtrusive and lightweight as possible to provide the greatest chance that infants/toddlers will tolerate wearing the equipment.

        1. Embed the system into a cap by attaching the scene and eye cameras to a Velcro strap that is affixed to the opposite side of Velcro sewn onto the cap, and positioning the cameras out of the center of the toddler's view.

           NOTE: Systems designed to be similar to glasses are not optimal. The morphology of the toddler's face is different from that of an adult and parts that rest on the toddler's nose or ears can be distracting and uncomfortable for the participant.

        2. If the ET is wired to a computer, bundle the cables and keep them behind the participant's back to prevent distraction or tripping. Alternatively, use a self-contained system that stores data on an intermediate device, such as a mobile phone, that can be placed on the child, which allows for greater mobility.

    4. Select a calibration software package that allows for offline calibration.

2. **Recording Environment**.

    1. Consider the extent to which the child will move throughout the space during data collection. If a single position is preferable, mention this to the child's caregiver so they can help the child stay in the desired location. Remove all potential distractors from the space except for those the child should interact with, which should be within reach.

        2. Employ a third-person camera to assist in the later coding of children's behavior as well as to identify moments when the ET may become displaced. If the child will move throughout the space, consider additional cameras as well.

2. **Collect the Eye-Tracking Data.**

    1. **Personnel and Activity**. Have two experimenters present, one to interact with and occupy the child, and one to place and position the ET.

        1. Fully engage the child in an activity that occupies the child's hands so that the child does not reach up to move or grab the ET while it is being placed on their head. Consider toys that encourage manual actions and small books that the child can hold while the experimenter or the parent reads to the child.

    2. **Place the ET on the Child**. Because toddlers' tolerance of wearing the head-mounted ET varies, follow these recommendations to promote success in placing and maintaining the ET on the child:

        1. In the time leading up to the study, ask caregivers to have their child wear a cap or beanie, similar to what is used with the ET, at home to get them accustomed to having something on their head.

        2. At the study, have different types of caps available to which the ET can be attached. Customize caps by purchasing different sizes and styles of caps, such as a ball cap that can be worn backward or a beanie with animal ears, and adding Velcro to which the eye-tracking system, fitted with the opposite side of the Velcro, can be attached. Also consider having hats to be worn by the caregiver and experimenters, to encourage the child's interest and willingness to also wear a cap.

            1. Before putting the cap on the child, have an experimenter desensitize the toddler to touches to the head by lightly touching the hair several times when the attention and interest of the toddler is directed to a toy.

        3. To place the ET on the child, be behind or to the side of the child (see Figure 2A). Place the ET on the child when their hands are occupied, such as when the child is holding a toy in each hand.

            1. If the child looks towards the experimenter placing the ET, say hello and let the child know what is being done while proceeding to quickly place the ET on the child's head. Avoid moving too slowly while placing the ET, which can cause child distress and may lead to poor positioning as the child has greater opportunity to move their head or reach for the ET.

            2. To reduce time spent adjusting the camera after placement, before placing the ET on the participant, set the cameras to be

in their anticipated position when upon the child's head (see Sections 2.3.1 and 2.3.2).

3. **Position the ET's Scene and Eye Cameras**. Once the ET is on the child's head, make adjustments to the position of the scene and eye cameras while monitoring these cameras' video feeds:

    1. Position the scene camera low on the forehead to best approximate the child's field of view (see Figure 1B); center the scene camera view on what the child will be looking at during the study.

        1. Keep in mind that hands and held objects will always be very close to the child and low in the scene camera view, while further objects will be in the background and higher in the scene camera view. Position the scene camera to best capture the type of view most relevant to the research question.

        2. Test the position of the scene camera by attracting the child's attention to specific locations in their field of view by using a small toy or laser pointer. Ensure these locations are at the anticipated viewing distance of the regions that will be of interest during the study (see Figure 3).

        3. Avoid tilt by checking that horizontal surfaces appear flat in the scene camera view. Mark the upright orientation of the scene camera to mitigate the possibility of the camera getting inadvertently inverted during repositioning, but note that extra steps during post-processing can revert the images to the correct orientation if necessary.

    2. To obtain high quality gaze data, position the eye camera to detect both the pupil and corneal reflection (CR) (see Figure 2).

        1. Position the eye camera so it is centered on the child's pupil, with no occlusion by cheeks or eyelashes throughout the eye's full range of motion (see Figure 2C–F for examples of good and bad eye images). To aid with this, position the eye camera below the eye, near the cheek, pointing upward, keeping the camera out of the center of the child's view. Alternatively, position the eye camera below and to the outer side of the eye, pointing inward.

        2. Ensure that the camera is close enough to the eye that its movement produces a relatively large displacement of the pupil in the eye camera image.

        3. Avoid tilt by making sure the corners of the eye in the eye image can form a horizontal line (see Figure 2C).

        4. Ensure that the contrast of the pupil versus the iris is relatively high so that the pupil can be accurately distinguished from iris

(see Figure 2C). To aid with this, adjust either the position of the LED light (if next to the eye camera) or the distance of the eye camera from the eye (if the LED is not independently adjustable). For increased pupil detection, position the LED light at an angle and not straight into the eye. Be sure that any adjustments to the LED light still produce a clear CR (see Figure 2C).

4. **Obtain Points During the Study for Offline Calibration**.

    1. Once the scene and eye images are as high quality as they can be, collect calibration data by drawing the child’s attention to different locations in their field of view.

        1. Obtain calibration points on various surfaces with anything that clearly directs the child’s attention to a small, clear point in their field of view (see Figure 3). For instance, use a laser pointer against a solid background, or a surface with small independently activated LED lights.

        2. Limit the presence of other interesting targets in the child’s view to ensure that the child looks at the calibration targets.

    2. Alternate between drawing attention to different locations that require large angular displacements of the eye.

        1. Cover the field of view equally and do not move too quickly between points, which will aid in finding clear saccades from the child during offline calibration to help to infer when they looked to the next location.

        2. If the child does not immediately look to the new highlighted location, get their attention to the location by wiggling the laser, turning off/on the LEDs, or touching the location with a finger.

        3. If feasible, obtain more calibration points than needed in case some turn out to be unusable later.

    3. Be sure that the child’s body position during calibration matches the position that will be used during the study.

        1. For example, do not collect calibration points when the child is sitting if it is expected that the child will later be standing.

        2. Ensure that the distance between the child and the calibration targets is similar to the distance between the child and regions that will be of interest during the study.

        3. Do not place calibration points very close to the child’s body if, during the experiment, the child will primarily be looking at objects that are further away. If one is interested in both near

and far objects, consider obtaining two different sets of calibration points that can later be used to create unique calibrations for each viewing distance (see Section 3.1 for more information).

NOTE: Binocular eye tracking is a developing technology[13,14] that promises advances in tracking gaze in depth.

4. To accommodate for drift or movement of the ET during the study, collect calibration points at both the beginning and end of the study at minimum. If feasible, collect additional calibration points at regular intervals during the session.

5. **Monitor the ET and Third-Person Video Feeds During the Study**.

   1. If the ET gets bumped or misaligned due to other movements/actions, take note of when in the study this happened because it may be necessary to recalibrate and code the portions of the study before and after the bump/misalignment separately (see Section 3.1.1).

   2. If possible, interrupt the study after each bump/misalignment to reposition the scene and eye cameras (see Section 2.3), then obtain new points for calibration (see Section 2.4).

3. After the Study, Calibrate the ET Data Using Calibration Software.

Note: A variety of calibration software packages are commercially available.

1. **Consider Creating Multiple Calibrations**. Customize calibration points to different video segments to maximize the accuracy of the gaze track by not feeding the algorithm incorrectly mismatched data.

   1. If the ET changed position at any time during the study, create separate calibrations for the portions before and after the change in ET position.

   2. If interested in attention to objects at very different viewing distances, create separate calibrations for the portions of the video where the child is looking to objects at each viewing distance. Bear in mind that differences in viewing distance may be created by shifts in the child’s visual attention between very close and vary far objects, but also by changes in the child’s body position relative to an object, such as shifting from sitting to standing.

2. **Perform Each Calibration**. Establish the mapping between scene and eye by creating a series of calibration points - points in the scene image to which the child’s gaze was clearly directed during that frame. Note that the calibration software can extrapolate and interpolate the point of gaze (POG) in all frames from a set of calibration points evenly dispersed across the scene image.

   1. Assist the calibration software in detecting the pupil and CR in each frame of the eye camera video to ensure that the identified POG is

reliable. In cases where the software cannot detect the CR reliably and consistently, use the pupil only (note, however, that data quality will suffer as a result).

1. Obtain a good eye image in the eye camera frames by adjusting the thresholds of the calibration software’s various detection parameters, which may include: the brightness of the eye image, the size of the pupil the software expects, and a bounding box that sets the boundaries of where the software will look for the pupil. Draw the bounding box as small as possible while ensuring that the pupil remains inside the box throughout the eye’s complete range of motion. Be aware that a larger bounding box that encompasses space that the pupil never occupies increases the likelihood of false pupil detection and may cause small movements of the pupil to be detected less accurately.
2. Be aware that even after adjusting the software’s various detection thresholds, the software may sometimes still incorrectly locate the pupil or CR; for instance, if eyelashes cover the pupil.

2. Find good calibration points based on the scene and eye camera frames. Note that the best calibration points provided to the software are those in which the pupil and CR are accurately detected, the eye is stably fixated on a clearly identifiable point in space in the scene image, and the points are evenly dispersed across the entire range of the scene image.
   1. Ensure that pupil detection is accurate for each frame in which a calibration point is plotted, so that both valid x-y scene coordinates and valid x-y pupil coordinates are fed into the algorithm.
   2. During the first pass at calibration, identify calibration points at moments when the child is clearly looking to a distinct point in the scene image. Keep in mind that these can be points intentionally created by the experimenter during data collection, for instance with a laser pointer (see Figure 3A–B), or they can be points from the study in which the POG is easily identifiable (see Figure 3C), as long as the pupil is accurately detected for those frames.
   3. To find moments of gaze to more extreme x-y scene image coordinates, scan through the eye camera frames to find moments with accurate pupil detection when the child’s eye is at its most extreme x-y position.

3. Do multiple "passes" for each calibration to iteratively hone in on the most accurate calibration possible. Note that after completing a first "pass" at calibration, many software programs will allow the deletion of points previously used without losing the current track (*e.g.* crosshair). Select a new set of calibration points to train the algorithm from scratch but with the additional aid of the POG track generated by the previous calibration pass, allowing one to gradually increase calibration accuracy by progressively "cleaning up" any noise or inaccuracies introduced by earlier passes.

3. **Assess the quality of calibration by observing how well the POG corresponds to known gaze locations, such as the dots produced by a laser pointer during calibration, and reflects the direction and magnitude of the child's saccades**. Avoid using points to assess calibration quality that were also used as points during the calibration process.

   1. Remember that because children's heads and eyes are typically aligned, children's visual attention is most often directed toward the center of the scene image, and an accurate track will reflect this. To assess the centeredness of the track, plot the frame-by-frame x-y POG coordinates in the scene image generated by the calibration (see Figure 4). Confirm that the points are most dense in the center of the scene image and distributed symmetrically, except in cases where the scene camera was not centered on the center of the child's field of view when originally positioned.

   2. Note that some calibration software will generate linear and/or homography fit scores that reflect calibration accuracy. Keep in mind that these scores are useful to some extent since, if they are poor, the track will likely also be poor. However, do not use fit scores as the primary measure of calibration accuracy as they reflect the degree to which the chosen calibration points agree with themselves, which provides no information about the fit of those points to the ground truth location of the POG.

   3. Remember that there are moments in the study that the target of gaze is easily identifiable and therefore can be used as ground truth. Calculate accuracy in degrees of visual angle by measuring the error between known gaze targets and the POG crosshair (error in pixels from the video image can be approximately converted to degrees based on lens characteristics of the scene camera)[4].

4. Code Regions of Interest (ROIs).

NOTE: ROI coding is the evaluation of POG data to determine what region a child is visually attending to during a particular moment in time. ROI may be coded with high accuracy and high resolution from the frame-by-frame POG data. The output of this coding is a stream of data points -one point per video frame - that indicate the region of POG over time (see Figure 5A).

1. **Prior to beginning ROI coding, compile a list of all ROIs that should be coded based on the research questions**. Be aware that coding ROIs that are not needed to answer the research questions makes coding unnecessarily time-consuming.

2. **Principles of ROI Coding**.

    1. Remember that successful coding requires relinquishing the coder’s assumptions about where the child should be looking, and instead carefully examining each frame’s eye image, scene image, and computed POG. For example, even if an object is being held by the child and is very large in the scene image for a particular frame, do not infer that the child is looking at that object at that moment unless also indicated by the position of the eyes. Note that ROIs indicate what region the child is foveating, but do not capture the complete visual information the child is taking in.

    2. **Use the eye image, scene image, and POG track to determine which ROI is being visually attended to**.

        1. **Use the POG track as a guide, not as ground-truth**. Though ideally the POG track will clearly indicate the exact location gazed upon by the child for each frame, be aware that this will not always be the case due to the 2 dimensional (2D) nature of the scene image relative to the 3D nature of the real world viewed by the child and variation in calibration accuracy between participants.

            1. Remember that the computed POG track is an estimate based on a calibration algorithm and that reliability of the POG track for a particular frame therefore depends on how well the pupil and CR are detected; if either or both are not detected or are incorrect, the POG track will not be reliable.

               NOTE: Occasionally, the crosshair will be consistently off-target by a fixed distance. Newer software may allow one to computationally correct for this discrepancy. Otherwise, a trained researcher may do the correction manually.

        2. **Use movement of the pupil in the eye image as the primary cue that the ROI may have changed**.

            1. Scroll through frames one by one watching the eye image. When a visible movement of the eye occurs,

check whether the child is shifting their POG to a new ROI or to no defined ROI.

2. Note that not all eye movements indicate a change in ROI. If the ROI constitutes a large region of space (*e.g*., an up close object), bear in mind that small eye movement may reflect a look to a new location within the same ROI. Similarly, remember that eye movements can occur as the child tracks a single moving ROI, or as a child who is moving their head also moves their eyes to maintain gaze on the same ROI.

3. Note that with some ETs the eye image is a mirrored-image of the child's eye, in which case if the eye moves to the left that should correspond to a shift to the right in the scene.

3. **Because the POG track serves only as a guide, make use of available contextual information as well to guide coding decisions**.

   1. Integrate information from different sources or frames when coding ROI.Even though the ROI is coded separately for each frame, utilize frames before and after the current frame to gain contextual information that may aid in determining the correct ROI. For instance, if the POG track is absent or incorrect for a given frame due to poor pupil detection, but the eye did not move based on the preceding and subsequent frames in which the pupil was accurately detected, then ignore the POG track for that frame and code the ROI based on the surrounding frames.

   2. Make other decisions specific to the users' research questions.

      1. For example, make a protocol for how to code ROI when two ROIs are in close proximity to one another, in which case it can be difficult to determine which one is the "correct" ROI. In cases where the child appears to be fixating at the junction of the two ROIs, decide whether to code both ROIs simultaneously or whether to formulate a set of decision rules for how to select and assign only one of the ROI categories.

      2. As an additional example, when an object of interest is held such that a hand is occluding the object, decide whether to code the POG as an ROI for the hand or as an ROI for the held object.

3. **Code ROI for Reliability**. Implement a reliability coding procedure after the initial ROI coding protocol has been completed. There are many different types of reliability coding procedures available; choose the most relevant procedure based on the specific research questions.

Representative Results

The method discussed here was applied to a free-flowing toy play context between toddlers and their parents. The study was designed to investigate natural visual attention in a cluttered environment. Dyads were instructed to play freely with a set of 24 toys for six minutes. Toddlers' visual attention was measured by coding the onset and offset of looks to specific regions of interest (ROIs) -- each of the 24 toys and the parent's face -- and by analyzing the duration and proportion of looking time to each ROI. The results are visualized in Figure 5.

Figure 5A shows sample ROI streams for two 18-month-old children. Each colored block in the streams represents continuous frames in which the child looked at a particular ROI. The eye-gaze data obtained demonstrate a number of interesting properties of natural visual attention.

First, the children show individual differences in their selectivity for different subsets of toys. Figure 5B shows the proportion of the 6-minute interaction that each child spent looking at each of 10 selected toy ROIs. Though the total proportion of time Child 1 and Child 2 spent looking at toys (including all 24 toy ROIs) was somewhat similar, 0.76 and 0.87, respectively, proportions of time spent on individual toys varied greatly, both within and between subjects.

How these proportions of looking time were achieved also differed across children. Figure 5C shows each child's mean duration of looks to each of 10 selected toy ROIs. The mean duration of looks to all 24 toy ROIs for Child 2 ($M = 2.38$ s, $SD = 2.20$ s) was almost twice as long as that of Child 1 ($M = 1.20$ s, $SD = 0.78$ s). Comparing the looking patterns to the red ladybug rattle (purple bars) in Figure 5B,C illustrates why computing multiple looking measures, such as proportions and durations of looking, is important for a complete understanding of the data; the same proportion of looking to this toy was achieved for these children through different numbers of looks of different durations.

Another property demonstrated by these data is that both children rarely looked to their parent's face: the proportions of face looking for Child 1 and Child 2 were .015 and .003, respectively. Furthermore, the duration of these children's looks to their parent's face were short, on average 0.79 s ($SD = 0.39$ s) and 0.40 s ($SD = 0.04$ s) for Child 1 and Child 2, respectively.

Discussion

This protocol provides guiding principles and practical recommendations for implementing head-mounted eye tracking with infants and young children. This protocol was based on the study of natural toddler behaviors in the context of parent-toddler free play with toys in a laboratory setting. In-house eye-tracking equipment and software were used for calibration and data coding. Nevertheless, this protocol is intended to be generally applicable to researchers using a variety of head-mounted eye-tracking systems to study a variety of topics in infant and child development. Though optimal use of this protocol will involve study-specific tailoring, the adoption of these general practices have led to successful use of this protocol in a variety of contexts (see Figure 1), including the simultaneous head-mounted eye tracking of parents and toddlers[7,8,9,10], and head-mounted eye tracking of clinical populations including children with cochlear implants[15] and children diagnosed with autism spectrum disorders[16,17].

This protocol provides numerous advantages for investigating the development of a variety of natural competencies and behaviors. The freedom of head and body movement that head-mounted ETs allow gives researchers the opportunity to capture both participants' self-generated visual environments and their active exploration of those environments. The portability of head-mounted ETs enhances researchers' ability to collect data in more ecologically valid contexts. Due to these advantages, this method provides an alternative to screen-based looking time and eye-tracking methods for studying development across domains such as visual attention, social attention, and perceptual-motor integration, and complements and occasionally challenges the inferences researchers can draw using more traditional experimental methods. For instance, the protocol described here increases the opportunity for participants to exhibit individual differences in looking behavior, because participants have control not only over where and for how long they focus their visual attention in a scene, as in screen-based eye tracking, but also over the composition of those scenes through their eye, head, and body movements and physical manipulation of elements in the environment. The two participants' data presented here demonstrate individual differences in how long toddlers look and what objects toddlers sample when they are able to actively create and explore their visual environment. Additionally, the data presented here, as well as other research employing this protocol, suggest that in naturalistic toy play with their parents, toddlers look to their parent's face much less than suggested by previous research[4,5,7,8,9,10].

Despite these benefits, head-mounted eye tracking with infants and toddlers poses a number of methodological challenges. The most critical challenge is obtaining a good calibration. Because the scene image is only a 2D representation of the 3D world that was actually viewed, a perfect mapping between eye position and gazed scene location is impossible. By following the guidelines provided in this protocol, the mapping can become reliably close to the "ground truth", however special attention should be paid to several issues. First, the freedom of head and body movement allowed by head-mounted eye tracking also means that young participants will often bump the eye-tracking system. This is a problem because any change in the physical position of the eye relative to the eye or scene cameras will change the mapping between the pupil/CR and the corresponding pixels attended in the scene image. Conducting separate calibrations for these portions of the study is therefore critical, as failure to do so will result in an algorithm that only tracks the child's gaze accurately for one portion of the study, if only points during one portion are used to calibrate. Second, accurate detection of the child's pupil and CR are critical. If a calibration point in the scene image is plotted while the pupil is incorrectly detected or not detected at all, then the algorithm either learns to associate this calibration x-y coordinate in the scene image with an incorrect pupil x-y coordinate, or the algorithm is being fed blank data in the case where the pupil is not detected at all. Thus, if good detection is not achieved for a segment of the study, calibration quality for these frames will be poor and should not be trusted for coding POG. Third, because children's heads and eyes are typically aligned, visual attention is most often directed toward the center of the scene image. Nevertheless, extreme x-y calibration points in the scene image are also necessary for establishing an accurate gaze track across the entire scene image. Thus, although calibration points should typically be chosen at moments when the eye is stable on an object, this may not be possible for calibration points in the far corners of the scene image. Finally, keep in mind that even when a good eye image is obtained and the system calibrates, this does not ensure that the data is of sufficient quality for the intended analyses. Differences in individual factors such as eye physiology, as well as environmental factors such as lighting and differences in eye-tracking hardware and software can all influence data quality and have the potential to create offsets or inaccuracies in the data.[18,19] provide more information and possible solutions for such issues (see also Franchak 2017[20]).

Working with infants and toddlers also involves the challenge of ensuring tolerance of the head-mounted ET throughout the session. Employing the recommendations included in this protocol, designed for use with infants from approximately 9–24 months of age, a laboratory can obtain high-quality head-mounted eye-tracking data from approximately 70% of participants[20]. The other 30% of participants may either not begin the study due to intolerance of the eye tracker or fuss out of the study before sufficient data (*e.g.*, >3–5 minutes of play) with a good eye track can be obtained. For the successful 70% of infant and toddler participants, these sessions typically last for upwards of 10 minutes, however much longer sessions may be infeasible with current technologies, depending on the age of the participant and the nature of the task in which the participant is engaged. When designing the research task and environment, researchers should keep in mind the developmental status of the participants, as motor ability, cognitive ability, and social development including sense of security around strangers, can all influence participants' attention span and ability to perform the intended task. Employing this protocol with infants much younger than 9 months will also involve additional practical challenges such as propping up infants that cannot yet sit on their own, as well as consideration of eye morphology and physiology, such as binocular disparity, which differ from that of older children and adults[19,21]. Moreover, this protocol is most successful when carried out by experienced trained experimenters, which can constrain the range of environments in which data may be collected. The more practice experimenters have, the more likely they will be able to conduct the experiment smoothly and collect eye tracking data of high quality.

Head-mounted eye tracking can also pose the additional challenge of relatively more time-consuming data coding. This is because, for the purpose of finding ROIs, head-mounted eye-tracking data is better coded frame by frame than by "fixations" of visual attention. That is, fixations are typically identified when the rate of change in the frame-by-frame x-y POG coordinates is low, taken as an indication that the eyes are stable on a point. However, because the scene view from a head-mounted eye tracker moves with the participant's head and body movements, the eye's position can only be accurately mapped to a physical location being foveated by considering how the eyes are moving relative to head and body movements. For instance, if a participant moves their head and eyes together, rather than their eyes only, the x-y POG coordinates within the scene can remain unchanged even while a participant scans a room or tracks a moving object. Thus, "fixations" of visual attention cannot be easily and accurately determined from only the POG data. For further information on issues associated with identifying fixations in head-mounted eye tracking data, please consult other work[15,22]. Manually coding data frame-by-frame for ROI can require extra time compared to coding fixations. As a reference, it took highly trained coders between 5 and 10 minutes to manually code for ROI each minute of the data presented here, which was collected at 30 frames per second. The time required for coding is highly variable and depends on the quality of the eye tracking data; the size, number, and visual discriminability of ROI targets; the experience of the coder; and the annotation tool used.

Despite these challenges, this protocol can be flexibly adapted to a range of controlled and naturalistic environments. This protocol can also be integrated with other technologies, such as motion tracking and heart-rate monitoring, to provide a high-density multimodal dataset for examining natural behavior, learning, and development than previously possible. Continued advances in head-mounted eye-tracking technology will undoubtedly alleviate many current challenges and provide even greater frontiers for the types of research questions that can be addressed using this method.

Acknowledgements

This research was funded by the National Institutes of Health grants R01HD074601 (C.Y.), T32HD007475-22 (J.I.B., D.H.A.), and F32HD093280 (L.K.S.); National Science Foundation grant BCS1523982 (L.B.S., C.Y.); and by Indiana University through the Emerging Area Research Initiative Learning: Brains, Machines, and Children (L.B.S.). The authors thank the child and parent volunteers who participated in this research and who agreed to be used in the figures and filming of this protocol. We also appreciate the members of the Computational Cognition and Learning Laboratory, especially Sven Bambach, Anting Chen, Steven Elmlinger, Seth Foster, Grace Lisandrelli, and Charlene Tay, for their assistance in developing and honing this protocol.

References

1. Tatler BW, Hayhoe MM, Land MF, & Ballard DH Eye guidance in natural vision: Reinterpreting salience. Journal of Vision. 11 (5), 1–23 (2011).
2. Hayhoe M Vision using routines: A functional account of vision. Visual Cognition. 7 (1–3), 43–64 (2000).
3. Land M, Mennie N, & Rusted J The Roles of Vision and Eye Movements in the Control of Activities of Daily Living. Perception. 28 (11), 1311–1328 (1999). [PubMed: 10755142]
4. Franchak JM, Kretch KS, & Adolph KE See and be seen: Infant-caregiver social looking during locomotor free play. Developmental Science. 21 (4), e12626 (2018). [PubMed: 29071760]
5. Franchak JM, Kretch KS, Soska KC, & Adolph KE Head-mounted eye tracking: a new method to describe infant looking. Child Development. 82 (6), 1738–50 (2011). [PubMed: 22023310]
6. Kretch KS, & Adolph KE The organization of exploratory behaviors in infant locomotor planning. Developmental Science. 20 (4), e12421 (2017).
7. Yu C, & Smith LB Hand-Eye Coordination Predicts Joint Attention. Child Development. 88 (6), 2060–2078 (2017). [PubMed: 28186339]
8. Yu C, & Smith LB Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. PLoS One. 8 (11), e79659 (2013). [PubMed: 24236151]
9. Yu C, & Smith LB Multiple Sensory-Motor Pathways Lead to Coordinated Visual Attention. Cognitive Science. 41, 5–31 (2016). [PubMed: 27016038]

10. Yu C, & Smith LB The Social Origins of Sustained Attention in One-Year-Old Human Infants. Current Biology. 26 (9), 1–6 (2016). [PubMed: 26725201]

11. Kretch KS, Franchak JM, & Adolph KE Crawling and walking infants see the world differently. Child Development, 85 (4), 1503–1518 (2014). [PubMed: 24341362]

12. Yu C, Suanda SH, & Smith LB Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. Developmental Science. (2018).

13. Hennessey C, & Lawrence P Noncontact binocular eye-gaze tracking for point-of-gaze estimation in three dimensions. IEEE Transactions on Biomedical Engineering, 56 (3), 790–799 (2009). [PubMed: 19272927]

14. Elmadjian C, Shukla P, Tula AD, & Morimoto CH 3D gaze estimation in the scene volume with a head-mounted eye tracker In Proceedings of the Workshop on Communication by Gaze Interaction. New York: Association for Computing Machinery, 3 (2018).

15. Castellanos I, Pisoni DB, Yu C, Chen C, & Houston DM (in press). Embodied cognition in prelingually deaf children with cochlear implants: Preliminary findings In Knoors H, & Marschark M (Eds.), Educating Deaf Learners: New Perspectives. New York: Oxford University Press (2018).

16. Kennedy DP, Lisandrelli G, Shaffer R, Pedapati E, Erickson CA, & Yu C Face Looking, Eye Contact, and Joint Attention during Naturalistic Toy Play: A Dual Head-Mounted Eye Tracking Study in Young Children with ASD. Poster at the International Society for Autism Research Annual Meeting 5 (2018).

17. Yurkovic JR, Lisandrelli G, Shaffer R, Pedapati E, Erickson CA, Yu C, & Kennedy DP Using Dual Head-Mounted Eye Tracking to Index Social Responsiveness in Naturalistic Parent-Child Interaction. Talk at the International Congress for Infant Studies Biennial Congress7 (2018).

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

19. Saez de Urabain IR, Johnson MH, & Smith TJ GraFIX: a semiautomatic approach for parsing low- and high-quality eye-tracking data. Behavior Research Methods. 47 (1), 53–72 (2015). [PubMed: 24671827]

20. Franchak JM Using head-mounted eye tracking to study development In Hopkins B, Geangu E, & Linkenauger S (Eds.), The Cambridge Encyclopedia of Child Development.(2nd ed.).Cambridge, UK: Cambridge University Press, 113–116 (2017).

21. Yonas A, Arterberry ME, & Granrud CE Four-month-old infants' sensitivity to binocular and kinetic information for three-dimensional object shape. Child Development. 58 (4), 910–917 (1987). [PubMed: 3608662]

22. Smith TJ, & Saez de Urabain IR Eye tracking In Hopkins B, Geangu E, & Linkenauger S (Eds.), The Cambridge Encyclopedia of Child Development.Cambridge, UK: Cambridge University Press, 97–101 (2017).

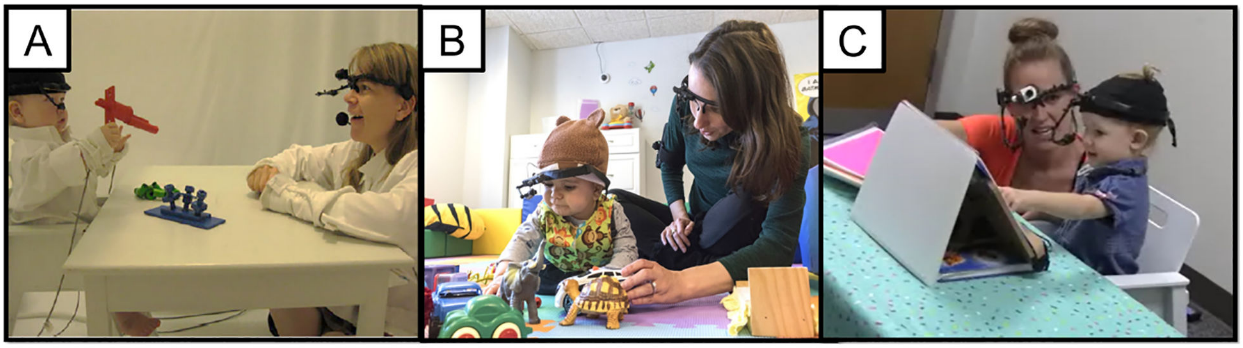

Figure 1. Head-mounted eye tracking employed in three different contexts:
(A) tabletop toy play, **(B)** toy play on the floor, and **(C)** reading a picture book.

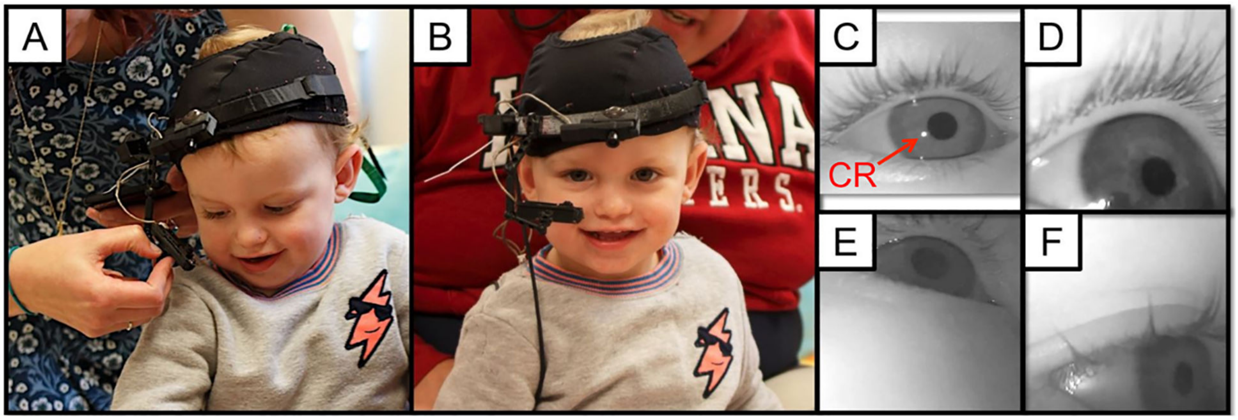

Figure 2. Setting up the head-mounted eye-tracking system.

(A) A researcher positioning an eye tracker on an infant. **(B)** A well-positioned eye tracker on an infant. **(C)** Good eye image with large centered pupil and clear corneal reflection (CR). **(D, E, F)** Examples of bad eye images.

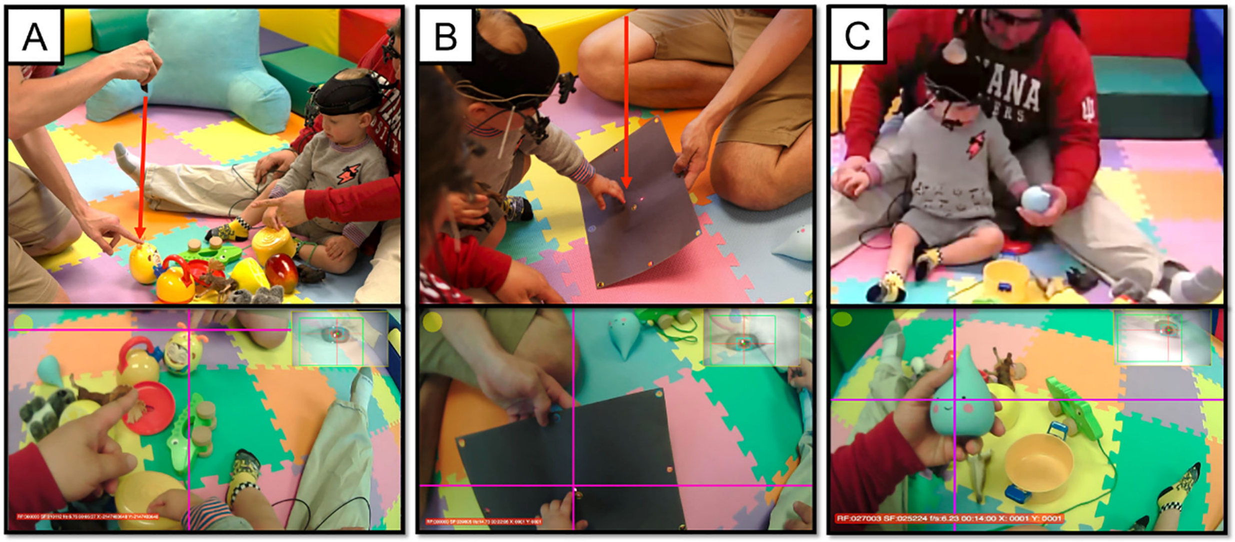

Figure 3. Three different ways of obtaining calibration points.

Two views of each moment are shown; top: third-person view, bottom: child's first-person view. Arrows in the third-person view illustrate the direction of a laser beam. Inset boxes in the upper right of the child's view show good eye images at each moment used for calibration and pink crosshairs indicate point of gaze based on the completed calibration. **(A)** Calibration point generated by an experimenter using a finger and laser pointer to direct attention to an object on the floor. **(B)** Calibration point generated by an experimenter using a laser pointer to direct attention to dots on a surface. **(C)** Calibration point during toy play with a parent in which the child's attention is directed to a held object.

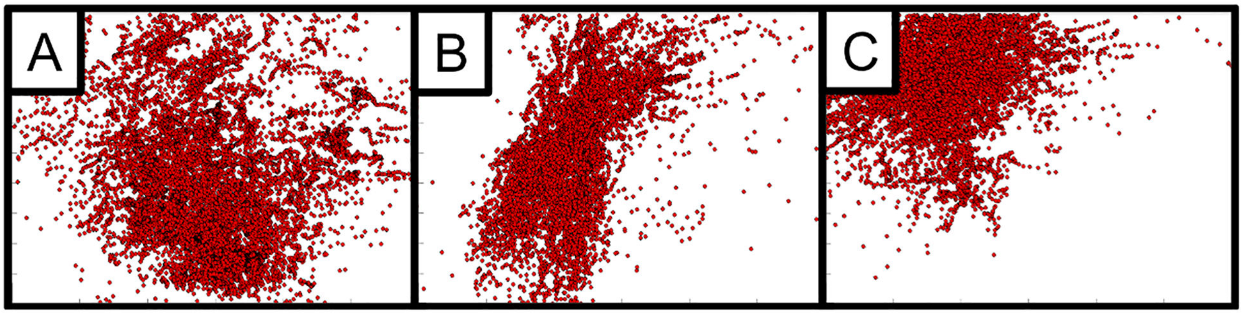

Figure 4. Example plots used to assess calibration quality.

Individual dots represent per-frame x-y point of gaze (POG) coordinates in the scene camera image, as determined by the calibration algorithm. **(A)** Good calibration quality for a child toy-play experiment, indicated by roughly circular density of POG that is centered and low (child POG is typically directed slightly downward when looking at toys the child is holding), and roughly evenly distributed POG in the remaining scene camera image. **(B)** Poor calibration quality, indicated by elongated and tilted density of POG that is off-centered, and poorly distributed POG in the remaining scene camera image. **(C)** Poor calibration quality and/or poor initial positioning of the scene camera, indicated by off-centered POG.

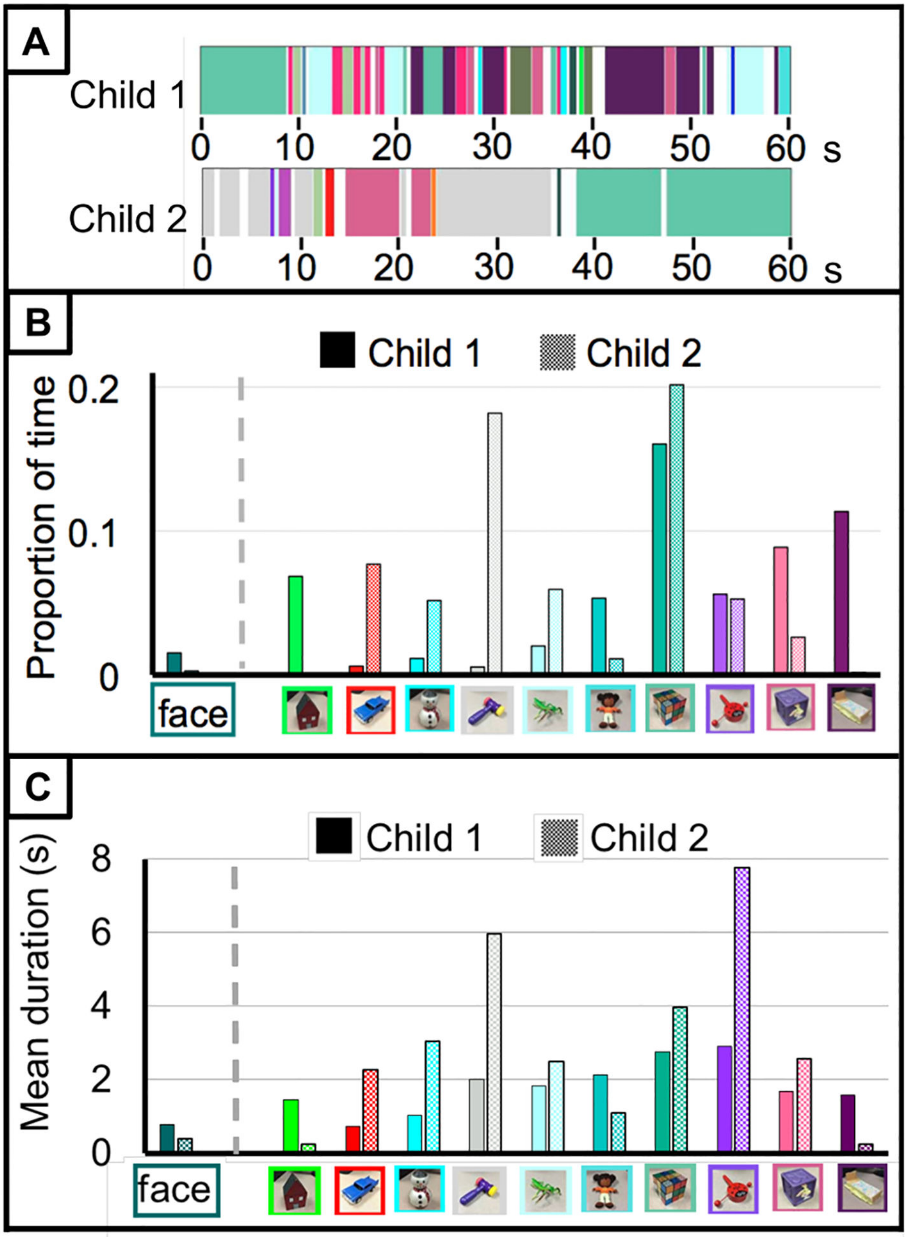

Figure 5. Two children's eye-gaze data and statistics.

(A) Sample ROI streams for Child 1 and Child 2 during 60 s of the interaction. Each colored block in the streams represents continuous frames in which the child looked at an ROI for either a specific toy or the parent's face. White space represents frames in which the child did not look at any of the ROIs. **(B)** Proportion of time looking at the parent's face and 10 toy ROIs, for both children. Proportion was computed by summing the durations of all looks to each ROI, and dividing the summed durations by the total session time of 6 minutes. **(C)** Mean duration of looks to the parent's face and ten toy ROIs, for both children. Mean duration was computed by averaging the durations of individual looks to each ROI during the 6-minute interaction.