

Journal of Visualized Experiments

Gaze in action: Head-mounted eye tracking of children's dynamic visual attention during naturalistic behavior --Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE58496R1
Full Title:	Gaze in action: Head-mounted eye tracking of children's dynamic visual attention during naturalistic behavior
Keywords:	Head-mounted eye tracking; egocentric vision; Development; infant; toddler; visual attention
Corresponding Author:	Lauren K Slone Indiana University Bloomington, IN UNITED STATES
Corresponding Author's Institution:	Indiana University
Corresponding Author E-Mail:	laurenkslone@gmail.com
Order of Authors:	Lauren K Slone Drew H Abney Jeremy I Borjon Chi-hsin Chen John M Franchak Daniel Pearcy Catalina Suarez-Rivera Tian Linger Xu Yayun Zhang Linda B Smith Chen Yu
Additional Information:	
Question	Response
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US\$2,400)
Please indicate the city, state/province, and country where this article will be filmed . Please do not use abbreviations.	Department of Psychological and Brain Sciences 1101 East 10th Street, Room A112, Indiana University Bloomington, IN, 47405



DEPARTMENT OF
PSYCHOLOGICAL AND
BRAIN SCIENCES

INDIANA UNIVERSITY

Bloomington

August 10th, 2018

Dear Editor:

Journal of Visualized Experiments

Thank you for the opportunity to revise manuscript JoVE58496, "Gaze in action: Head-mounted eye tracking of children's dynamic visual attention during naturalistic behavior." We found the editor's and reviewers' comments very helpful in revising this manuscript, and have taken care to consider each comment thoroughly. In the response letter I describe the changes we made to the manuscript in response to each comment. All changes to the manuscript are tracked to allow identification of all the edits.

The revised manuscript is 17 pages in length. Five figures are included in this submission. The material contained in this manuscript has not been published and is not under consideration for publication elsewhere. All authors have made meaningful contributions to the manuscript, and have approved the author order and the manuscript being submitted. As the corresponding author, I will keep my co-authors informed of the progress of the manuscript.

This work provides a novel contribution to the field of psychology by offering a complete, practical guide for researchers interested in using head-mounted eye tracking to capture children's egocentric views and visual attention within those views. We believe it is perfectly suited to publication in the *Journal of Visualized Experiments*, as the video created from this manuscript will allow researchers to clearly visualize the visual data that can be obtained using head-mounted eye tracking with young children.

Thank you for consideration of our revised manuscript. I look forward to hearing from you in due course.

Sincerely,

A handwritten signature in black ink that reads "Lauren K. Slone".

Lauren K. Slone
Postdoctoral Researcher
Indiana University Bloomington
Department of Psychological and Brain Sciences
1101 E 10th Street
Bloomington, IN 47405-7007
laurenkslone@gmail.com

TITLE:

Head-mounted Eye Tracking of Children's Dynamic Visual Attention During Naturalistic Behavior

AUTHORS & AFFILIATIONS:

Lauren K Slone¹, Drew H. Abney¹, Jeremy I. Borjon¹, Chi-hsin Chen², John M. Franchak³, Daniel Percy¹, Catalina Suarez-Rivera¹, Tian Linger Xu¹, Yayun Zhang¹, Linda B Smith¹, Chen Yu¹

¹Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

²Department of Otolaryngology-Head and Neck Surgery, The Ohio State University, Columbus, OH, USA

³Department of Psychology, University of California, Riverside, California, USA

Corresponding Author:

Lauren K Slone (laurenkslone@gmail.com)

Tel: (619) 857-0255

Email Addresses of Co-authors:

Drew H. Abney	(dhabney@indiana.edu)
Jeremy I. Borjon	(jborjon@iu.edu)
Chi-hsin Chen	(chi-hsin.chen@osumc.edu)
John M. Franchak	(john.franchak@ucr.edu)
Daniel Percy	(drpercy@umail.iu.edu)
Catalina Suarez-Rivera	(csuarezr@iu.edu)
Tian Linger Xu	(txu@indiana.edu)
Yayun Zhang	(yayzhang@indiana.edu)
Linda B. Smith	(smith4@indiana.edu)
Chen Yu	(chenyu@indiana.edu)

KEYWORDS:

Head-mounted eye tracking, egocentric vision, development, infant, toddler, visual attention

SUMMARY:

Young children do not passively observe the world, but rather actively explore and engage with their environment. This protocol provides guiding principles and practical recommendations for using head-mounted eye trackers to record infants' and toddlers' dynamic visual environments and visual attention in the context of natural behavior.

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.

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 behavior¹.

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 developments⁴⁻⁸. The use of these techniques has transformed understandings of joint attention⁷⁻⁹, sustained attention¹⁰, changing visual experiences with age and motor development^{4,6,11}, and the role of visual experiences in

word learning¹². 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

1.1. 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:

1.1.1. 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.

1.1.2. 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.

1.1.3. 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.

1.1.3.1. 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.

1.1.3.2. 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.

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

1.2. Recording Environment.

1.2.1. 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.

1.2.2. 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.

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

2.1.1. 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.

2.2. **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:

2.2.1. 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.

2.2.2. 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.

2.2.2.1. 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.

2.2.3. 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.

2.2.3.1. 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.

2.2.3.2. 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).

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:

2.3.1. 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.

2.3.1.1. 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.

2.3.1.2. 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**).

2.3.1.3. 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.

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

2.3.2.1. 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.

2.3.2.2. 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.

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

2.3.2.4. 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**).

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

2.4.1. 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.

2.4.1.1. 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.

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

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

2.4.2.1. 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.

2.4.2.2. 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.

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

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

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

2.4.3.2. 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.

2.4.3.3. 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.

2.4.4. 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.

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

2.5.1. 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).

2.5.2. 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.

3.1. 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.

3.1.1. 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.

3.1.2. 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.

306
307 3.2. **Perform Each Calibration.** Establish the mapping between scene and eye by creating a series
308 of calibration points – points in the scene image to which the child's gaze was clearly directed
309 during that frame. Note that the calibration software can extrapolate and interpolate the point
310 of gaze (POG) in all frames from a set of calibration points evenly dispersed across the scene
311 image.

312
313 3.2.1. Assist the calibration software in detecting the pupil and CR in each frame of the eye
314 camera video to ensure that the identified POG is reliable. In cases where the software cannot
315 detect the CR reliably and consistently, use the pupil only (note, however, that data quality will
316 suffer as a result).

317
318 3.2.1.1. Obtain a good eye image in the eye camera frames by adjusting the thresholds of the
319 calibration software's various detection parameters, which may include: the brightness of the
320 eye image, the size of the pupil the software expects, and a bounding box that sets the
321 boundaries of where the software will look for the pupil. Draw the bounding box as small as
322 possible while ensuring that the pupil remains inside the box throughout the eye's complete
323 range of motion. Be aware that a larger bounding box that encompasses space that the pupil
324 never occupies increases the likelihood of false pupil detection and may cause small movements
325 of the pupil to be detected less accurately.

326
327 3.2.1.2. Be aware that even after adjusting the software's various detection thresholds, the
328 software may sometimes still incorrectly locate the pupil or CR; for instance, if eyelashes cover
329 the pupil.

330
331 3.2.2. Find good calibration points based on the scene and eye camera frames. Note that the best
332 calibration points provided to the software are those in which the pupil and CR are accurately
333 detected, the eye is stably fixated on a clearly identifiable point in space in the scene image, and
334 the points are evenly dispersed across the entire range of the scene image.

335
336 3.2.2.1. Ensure that pupil detection is accurate for each frame in which a calibration point is
337 plotted, so that both valid x-y scene coordinates and valid x-y pupil coordinates are fed into the
338 algorithm.

339
340 3.2.2.2. During the first pass at calibration, identify calibration points at moments when the child
341 is clearly looking to a distinct point in the scene image. Keep in mind that these can be points
342 intentionally created by the experimenter during data collection, for instance with a laser pointer
343 (see **Figure 3A-B**), or they can be points from the study in which the POG is easily identifiable (see
344 **Figure 3C**), as long as the pupil is accurately detected for those frames.

345
346 3.2.2.3. To find moments of gaze to more extreme x-y scene image coordinates, scan through
347 the eye camera frames to find moments with accurate pupil detection when the child's eye is at
348 its most extreme x-y position.

349

3.2.3. 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.

3.3. **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.

3.3.1. 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.

3.3.2. 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.

3.3.3. 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**).

4.1. **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.

4.2. Principles of ROI Coding.

4.2.1. 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.

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

4.2.2.1. **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.

4.2.2.1.1. 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.

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

4.2.2.2.1. 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.

4.2.2.2.2. 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.

4.2.2.2.3. 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.

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

4.2.3.1. 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.

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

4.2.3.2.1. 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.

4.2.3.2.2. 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.

4.3. **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, .76 and .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.

FIGURE AND TABLE LEGENDS:

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.

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 toddlers⁷⁻¹⁰, and head-mounted eye tracking of clinical populations including children with cochlear implants¹⁵ 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^{4,5,7-10}.

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 2017²⁰).

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 participants²⁰. 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.

ACKNOWLEDGMENTS:

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.

DISCLOSURES:

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

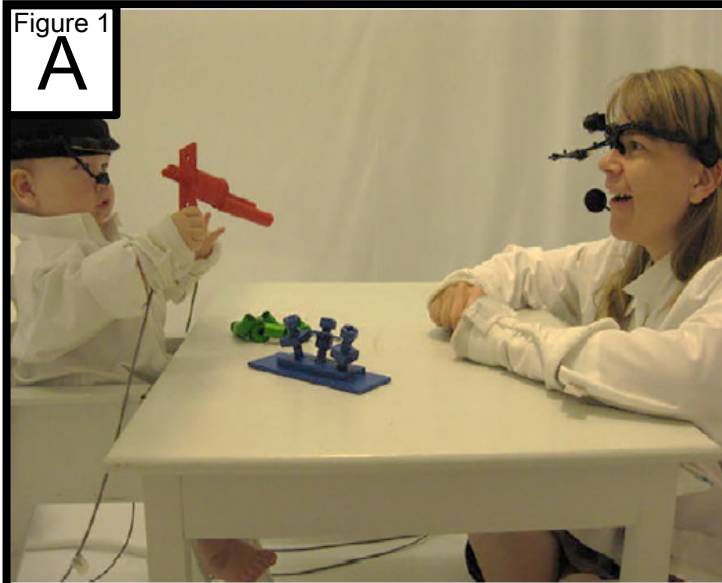
REFERENCES:

1. Tatler, B. W., Hayhoe, M. M., Land, M. F. & Ballard, D. H. 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).
4. Franchak, J. M., Kretch, K. S., & Adolph, K. E. See and be seen: Infant–caregiver social looking during locomotor free play. *Developmental Science* **21** (4), e12626 (2018).
5. Franchak, J. M., Kretch, K. S., Soska, K. C. & Adolph, K. E. Head-mounted eye tracking: a new method to describe infant looking. *Child Development* **82** (6), 1738–50 (2011).
6. Kretch, K. S. & Adolph, K. E. The organization of exploratory behaviors in infant locomotor planning. *Developmental Science* **20** (4), e12421 (2017).
7. Yu, C. & Smith, L. B. Hand–Eye Coordination Predicts Joint Attention. *Child Development* **88** (6), 2060–2078 (2017).
8. Yu, C. & Smith, L. B. 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).
9. Yu, C. & Smith, L. B. Multiple Sensory-Motor Pathways Lead to Coordinated Visual Attention. *Cognitive Science*. **41**, 5–31 (2016).
10. Yu, C. & Smith, L. B. The Social Origins of Sustained Attention in One-Year-Old Human Infants. *Current Biology* **26** (9), 1–6 (2016).
11. Kretch, K. S., Franchak, J. M., & Adolph, K. E. Crawling and walking infants see the world differently. *Child Development*, **85** (4), 1503–1518 (2014).
12. Yu, C., Suanda, S. H., & Smith, L. B. Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science* (in press).
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).
14. Elmadjian, C., Shukla, P., Tula, A. D., & Morimoto, C. H. 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, June).
15. Castellanos, I., Pisoni, D. B., Yu, C., Chen, C., & Houston, D. M. (in press). Embodied cognition in prelingually deaf children with cochlear implants: Preliminary findings. In H. Knoors, & M. Marschark (Eds.), *Educating Deaf Learners: New Perspectives*. New York: Oxford University Press.
16. Kennedy, D. P., Lisandrelli, G., Shaffer, R., Pedapati, E., Erickson, C. A., & 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* (May 2018).
17. Yurkovic, J. R., Lisandrelli, G., Shaffer, R., Pedapati, E., Erickson, C. A., Yu, C., & Kennedy, D. P. Using Dual Head-Mounted Eye Tracking to Index Social Responsiveness in Naturalistic Parent-Child Interaction. *Talk at the International Congress for Infant Studies Biennial Congress* (July

- 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, I. R., Johnson, M. H., & Smith, T. J. GraFIX: a semiautomatic approach for parsing low- and high-quality eye-tracking data. *Behavior Research Methods* **47** (1), 53–72 (2015).
20. Franchak, J. M. Using head-mounted eye tracking to study development. In B. Hopkins, E. Geangu, & S. Linkenauger (Eds.), *The Cambridge Encyclopedia of Child Development* (2nd ed.). Cambridge, UK: Cambridge University Press, 113–116 (2017).
21. Yonas, A., Arterberry, M. E., & Granrud, C. E. Four-month-old infants' sensitivity to binocular and kinetic information for three-dimensional-object shape. *Child Development* **58** (4), 910–917 (1987).
22. Smith, T. J. & Saez de Urabain, I. R. Eye tracking. In B. Hopkins, E. Geangu, & S. Linkenauger (Eds.), *The Cambridge Encyclopedia of Child Development*. Cambridge, UK: Cambridge University Press, 97–101 (2017).

Figure 1

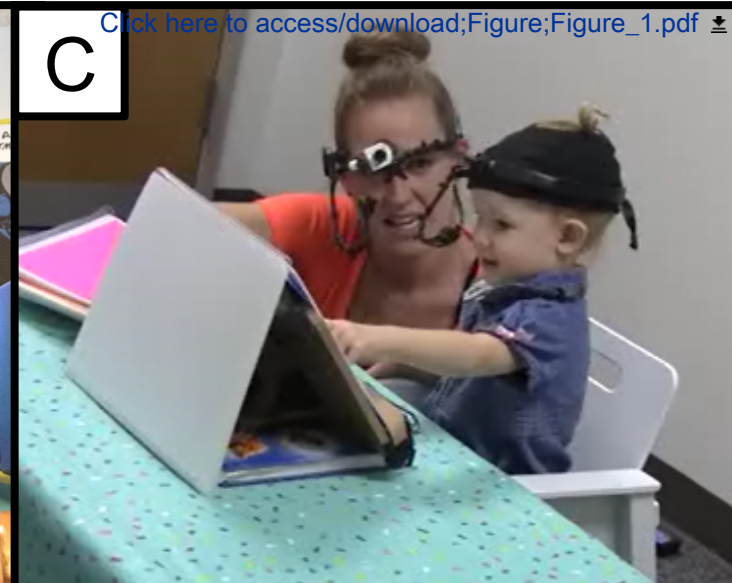
A



B



C



[Click here to access/download;Figure;Figure_1.pdf](#)

Figure 2

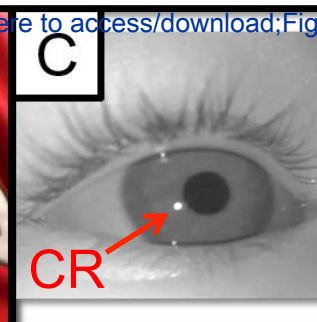
A



B



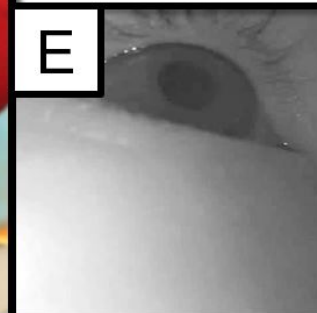
C



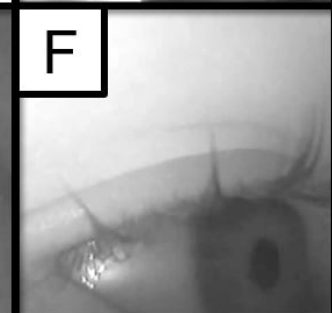
D



E



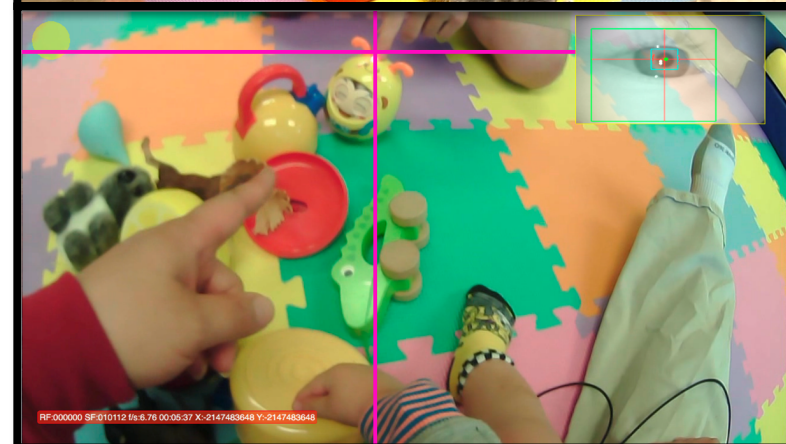
F



[Click here to access/download;Figure;Figure_2_withCR.pdf](#)

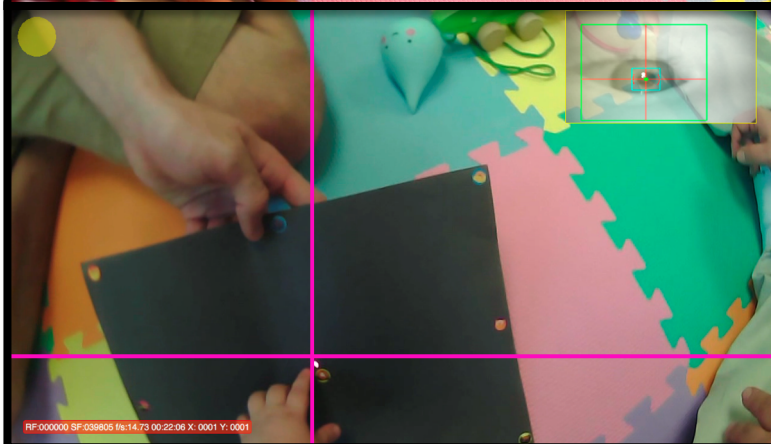
Figure 3

A



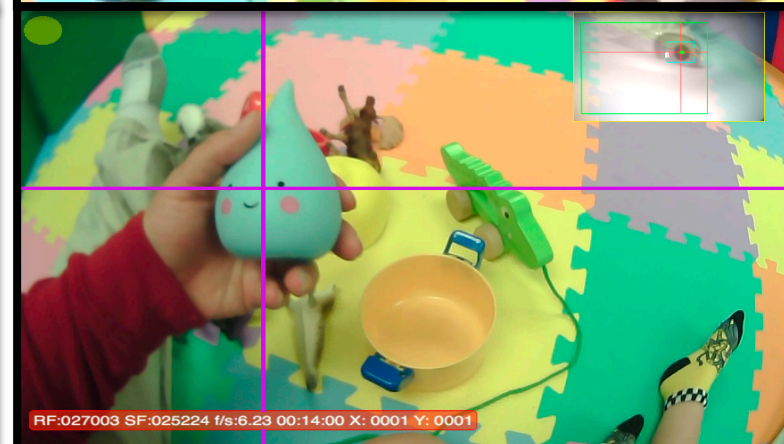
RF:000000 SF:010112 f/s:6.76 00:05:37 X:-2147483648 Y:-2147483648

B

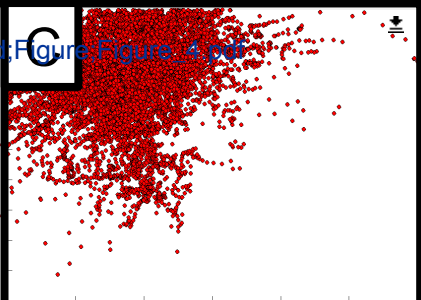
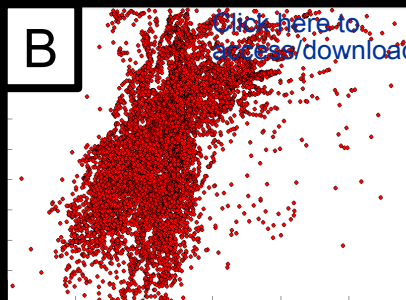
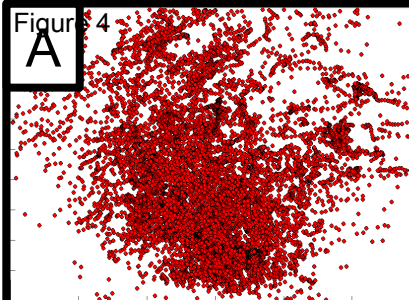


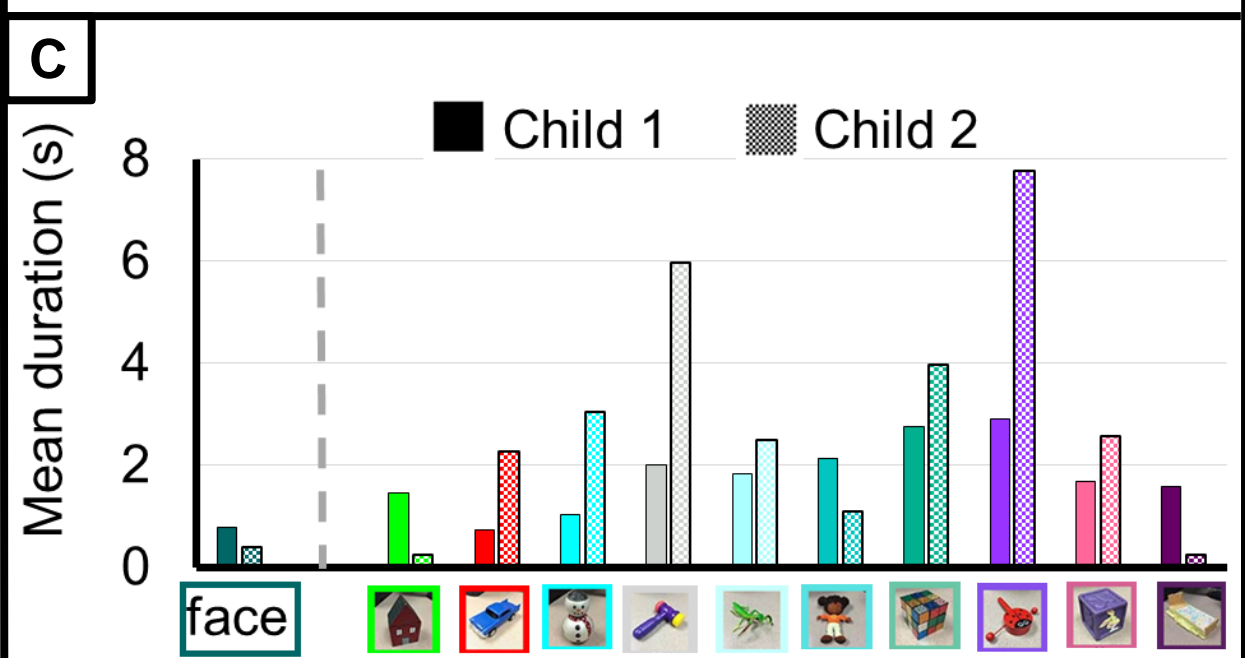
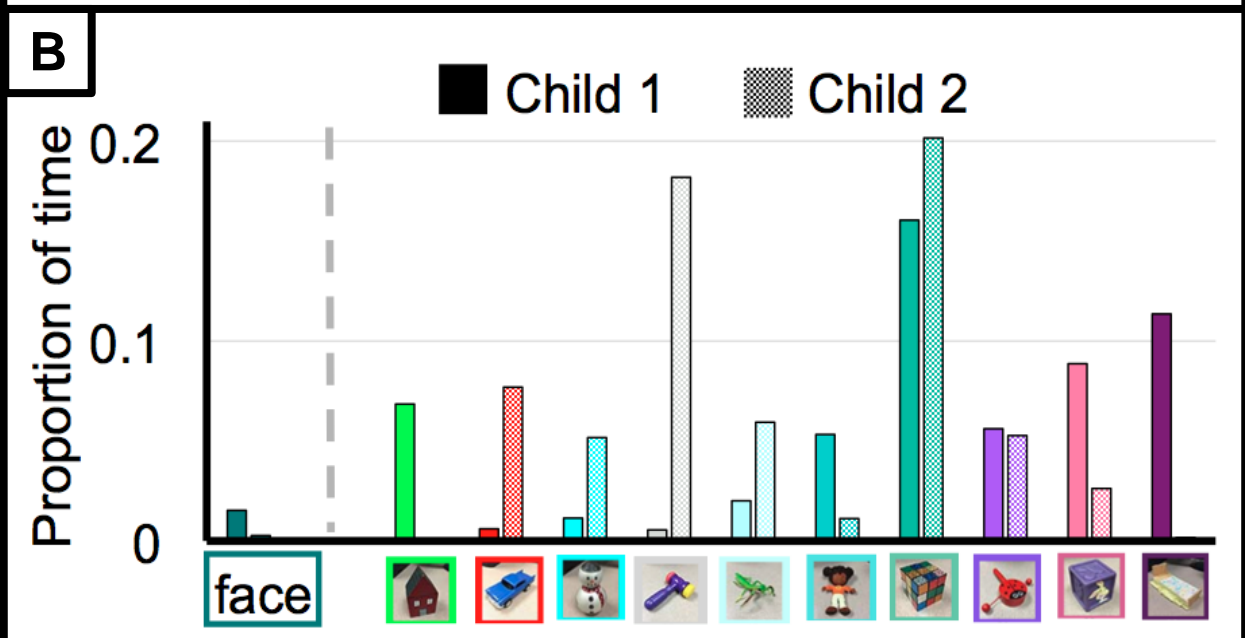
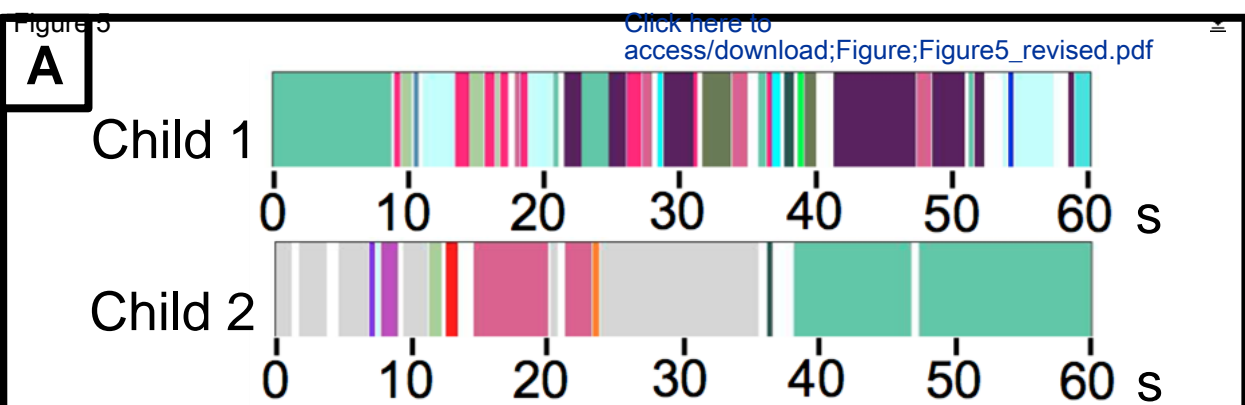
RF:000000 SF:039805 f/s:14.73 00:22:06 X: 0001 Y: 0001

C



RF:027003 SF:025224 f/s:6.23 00:14:00 X: 0001 Y: 0001





Name of Material/ Equipment	Company	Catalog Number	Comments/Description
Head-mounted eye tracker	Pupil Labs	World Camera and Eye Camera	



1 Alewife Center #200
Cambridge, MA 02140
tel. 617.945.9051
www.jove.com

ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article: Gaze in action: Head-mounted eye tracking of children's dynamic visual attention during naturalistic behavior

Author(s): Lauren K Slone, Drew H. Abney, Jeremy I. Borjon, Chi-hsin Chen, Daniel Pearcy, Catalina Suarez-Rivera, Yayun Zhang, Linda B Smith, Chen Yu

Item 1 (check one box): The Author elects to have the Materials be made available (as described at <http://www.jove.com/author>) via: ☒ Standard Access ☐ Open Access

Item 2 (check one box):

- ☒ The Author is NOT a United States government employee.
- ☐ The Author is a United States government employee and the Materials were prepared in the course of his or her duties as a United States government employee.
- ☐ The Author is a United States government employee but the Materials were NOT prepared in the course of his or her duties as a United States government employee.

ARTICLE AND VIDEO LICENSE AGREEMENT

1. **Defined Terms.** As used in this Article and Video License Agreement, the following terms shall have the following meanings: “**Agreement**” means this Article and Video License Agreement; “**Article**” means the article specified on the last page of this Agreement, including any associated materials such as texts, figures, tables, artwork, abstracts, or summaries contained therein; “**Author**” means the author who is a signatory to this Agreement; “**Collective Work**” means a work, such as a periodical issue, anthology or encyclopedia, in which the Materials in their entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole; “**CRC License**” means the Creative Commons Attribution-Non Commercial-No Derivs 3.0 Unported Agreement, the terms and conditions of which can be found at: <http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode>; “**Derivative Work**” means a work based upon the Materials or upon the Materials and other pre-existing works, such as a translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Materials may be recast, transformed, or adapted; “**Institution**” means the institution, listed on the last page of this Agreement, by which the Author was employed at the time of the creation of the Materials; “**JoVE**” means MyJoVE Corporation, a Massachusetts corporation and the publisher of *The Journal of Visualized Experiments*; “**Materials**” means the Article and / or the Video; “**Parties**” means the Author and JoVE; “**Video**” means any video(s) made by the Author, alone or in conjunction with any other parties, or by JoVE or its affiliates or agents, individually or in collaboration with the Author or any other parties, incorporating all or any portion of the Article, and in which the Author may or may not appear.

2. **Background.** The Author, who is the author of the Article, in order to ensure the dissemination and protection of the Article, desires to have the JoVE publish the Article and create and transmit videos based on the Article. In furtherance of such goals, the Parties desire to memorialize in this Agreement the respective rights of each Party in and to the Article and the Video.

3. **Grant of Rights in Article.** In consideration of JoVE agreeing to publish the Article, the Author hereby grants to JoVE, subject to **Sections 4 and 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Article in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Article into other languages, create adaptations, summaries or extracts of the Article or other Derivative Works (including, without limitation, the Video) or Collective Works based on all or any portion of the Article and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. If the “Open Access” box has been checked in **Item 1** above, JoVE and the Author hereby grant to the public all such rights in the Article as provided in, but subject to all limitations and requirements set forth in, the CRC License.

ARTICLE AND VIDEO LICENSE AGREEMENT

4. Retention of Rights in Article. Notwithstanding the exclusive license granted to JoVE in **Section 3** above, the Author shall, with respect to the Article, retain the non-exclusive right to use all or part of the Article for the non-commercial purpose of giving lectures, presentations or teaching classes, and to post a copy of the Article on the Institution's website or the Author's personal website, in each case provided that a link to the Article on the JoVE website is provided and notice of JoVE's copyright in the Article is included. All non-copyright intellectual property rights in and to the Article, such as patent rights, shall remain with the Author.

5. Grant of Rights in Video – Standard Access. This **Section 5** applies if the "Standard Access" box has been checked in **Item 1** above or if no box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby acknowledges and agrees that, Subject to **Section 7** below, JoVE is and shall be the sole and exclusive owner of all rights of any nature, including, without limitation, all copyrights, in and to the Video. To the extent that, by law, the Author is deemed, now or at any time in the future, to have any rights of any nature in or to the Video, the Author hereby disclaims all such rights and transfers all such rights to JoVE.

6. Grant of Rights in Video – Open Access. This **Section 6** applies only if the "Open Access" box has been checked in **Item 1** above. In consideration of JoVE agreeing to produce, display or otherwise assist with the Video, the Author hereby grants to JoVE, subject to **Section 7** below, the exclusive, royalty-free, perpetual (for the full term of copyright in the Article, including any extensions thereto) license (a) to publish, reproduce, distribute, display and store the Video in all forms, formats and media whether now known or hereafter developed (including without limitation in print, digital and electronic form) throughout the world, (b) to translate the Video into other languages, create adaptations, summaries or extracts of the Video or other Derivative Works or Collective Works based on all or any portion of the Video and exercise all of the rights set forth in (a) above in such translations, adaptations, summaries, extracts, Derivative Works or Collective Works and (c) to license others to do any or all of the above. The foregoing rights may be exercised in all media and formats, whether now known or hereafter devised, and include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. For any Video to which this Section 6 is applicable, JoVE and the Author hereby grant to the public all such rights in the Video as provided in, but subject to all limitations and requirements set forth in, the CRC License.

7. Government Employees. If the Author is a United States government employee and the Article was prepared in the course of his or her duties as a United States government employee, as indicated in **Item 2** above, and any of the licenses or grants granted by the Author hereunder exceed the scope of the 17 U.S.C. 403, then the rights granted hereunder shall be limited to the maximum rights permitted under such

statute. In such case, all provisions contained herein that are not in conflict with such statute shall remain in full force and effect, and all provisions contained herein that do so conflict shall be deemed to be amended so as to provide to JoVE the maximum rights permissible within such statute.

8. Likeness, Privacy, Personality. The Author hereby grants JoVE the right to use the Author's name, voice, likeness, picture, photograph, image, biography and performance in any way, commercial or otherwise, in connection with the Materials and the sale, promotion and distribution thereof. The Author hereby waives any and all rights he or she may have, relating to his or her appearance in the Video or otherwise relating to the Materials, under all applicable privacy, likeness, personality or similar laws.

9. Author Warranties. The Author represents and warrants that the Article is original, that it has not been published, that the copyright interest is owned by the Author (or, if more than one author is listed at the beginning of this Agreement, by such authors collectively) and has not been assigned, licensed, or otherwise transferred to any other party. The Author represents and warrants that the author(s) listed at the top of this Agreement are the only authors of the Materials. If more than one author is listed at the top of this Agreement and if any such author has not entered into a separate Article and Video License Agreement with JoVE relating to the Materials, the Author represents and warrants that the Author has been authorized by each of the other such authors to execute this Agreement on his or her behalf and to bind him or her with respect to the terms of this Agreement as if each of them had been a party hereto as an Author. The Author warrants that the use, reproduction, distribution, public or private performance or display, and/or modification of all or any portion of the Materials does not and will not violate, infringe and/or misappropriate the patent, trademark, intellectual property or other rights of any third party. The Author represents and warrants that it has and will continue to comply with all government, institutional and other regulations, including, without limitation all institutional, laboratory, hospital, ethical, human and animal treatment, privacy, and all other rules, regulations, laws, procedures or guidelines, applicable to the Materials, and that all research involving human and animal subjects has been approved by the Author's relevant institutional review board.

10. JoVE Discretion. If the Author requests the assistance of JoVE in producing the Video in the Author's facility, the Author shall ensure that the presence of JoVE employees, agents or independent contractors is in accordance with the relevant regulations of the Author's institution. If more than one author is listed at the beginning of this Agreement, JoVE may, in its sole discretion, elect not take any action with respect to the Article until such time as it has received complete, executed Article and Video License Agreements from each such author. JoVE reserves the right, in its absolute and sole discretion and without giving any reason therefore, to accept or decline any work submitted to JoVE. JoVE and its employees, agents and independent contractors shall have

ARTICLE AND VIDEO LICENSE AGREEMENT

full, unfettered access to the facilities of the Author or of the Author's institution as necessary to make the Video, whether actually published or not. JoVE has sole discretion as to the method of making and publishing the Materials, including, without limitation, to all decisions regarding editing, lighting, filming, timing of publication, if any, length, quality, content and the like.

11. **Indemnification.** The Author agrees to indemnify JoVE and/or its successors and assigns from and against any and all claims, costs, and expenses, including attorney's fees, arising out of any breach of any warranty or other representations contained herein. The Author further agrees to indemnify and hold harmless JoVE from and against any and all claims, costs, and expenses, including attorney's fees, resulting from the breach by the Author of any representation or warranty contained herein or from allegations or instances of violation of intellectual property rights, damage to the Author's or the Author's institution's facilities, fraud, libel, defamation, research, equipment, experiments, property damage, personal injury, violations of institutional, laboratory, hospital, ethical, human and animal treatment, privacy or other rules, regulations, laws, procedures or guidelines, liabilities and other losses or damages related in any way to the submission of work to JoVE, making of videos by JoVE, or publication in JoVE or elsewhere by JoVE. The Author shall be responsible for, and shall hold JoVE harmless from, damages caused by lack of sterilization, lack of cleanliness or by contamination due to the making of a video by JoVE its employees, agents or independent contractors. All sterilization, cleanliness or decontamination procedures shall be solely the responsibility of the Author and shall be undertaken at the Author's

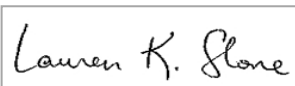
expense. All indemnifications provided herein shall include JoVE's attorney's fees and costs related to said losses or damages. Such indemnification and holding harmless shall include such losses or damages incurred by, or in connection with, acts or omissions of JoVE, its employees, agents or independent contractors.

12. **Fees.** To cover the cost incurred for publication, JoVE must receive payment before production and publication the Materials. Payment is due in 21 days of invoice. Should the Materials not be published due to an editorial or production decision, these funds will be returned to the Author. Withdrawal by the Author of any submitted Materials after final peer review approval will result in a US\$1,200 fee to cover pre-production expenses incurred by JoVE. If payment is not received by the completion of filming, production and publication of the Materials will be suspended until payment is received.

13. **Transfer, Governing Law.** This Agreement may be assigned by JoVE and shall inure to the benefits of any of JoVE's successors and assignees. This Agreement shall be governed and construed by the internal laws of the Commonwealth of Massachusetts without giving effect to any conflict of law provision thereunder. This Agreement may be executed in counterparts, each of which shall be deemed an original, but all of which together shall be deemed to be one and the same agreement. A signed copy of this Agreement delivered by facsimile, e-mail or other means of electronic transmission shall be deemed to have the same legal effect as delivery of an original signed copy of this Agreement.

A signed copy of this document must be sent with all new submissions. Only one Agreement required per submission.

CORRESPONDING AUTHOR:

Name:	Lauren Slone	
Department:	Psychological and Brain Sciences	
Institution:	Indiana University	
Article Title:	Gaze in action: Head-mounted eye tracking of children's dynamic visual attention during naturalistic behavior	
Signature:		Date: May 18, 2018

Please submit a signed and dated copy of this license by one of the following three methods:

- 1) Upload a scanned copy of the document as a pdf on the JoVE submission site;
- 2) Fax the document to +1.866.381.2236;
- 3) Mail the document to JoVE / Attn: JoVE Editorial / 1 Alewife Center #200 / Cambridge, MA 02139

For questions, please email submissions@jove.com or call +1.617.945.9051

Detailed Responses to Reviewer Comments

Editorial comments:

Changes to be made by the Author(s):

1. Please take this opportunity to thoroughly proofread the manuscript to ensure that there are no spelling or grammar issues.

Response: We have thoroughly proofread the manuscript for both spelling and grammar.

2. Figure 5c: Please change “sec” to “s”.

Response: “Sec” in Figure 5 has been changed to “s”.

3. Please rephrase the Introduction to include a clear statement of the overall goal of this method.

Response: The introduction has been modified to include the following statement of the overall goal: “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 developments⁴⁻⁸.”

4. Please revise the protocol text to avoid the use of any personal pronouns (e.g., “we”, “you”, “our” etc.).

Response: The one instance of “our” in “our laboratory” has been revised to read “a laboratory”. No other use of “we”, “you”, or “our” was found.

5. Please revise the protocol so that all text in the protocol section is written in the imperative tense as if telling someone how to do the technique (e.g., “Do this,” “Ensure that,” etc.). The actions should be described in the imperative tense in complete sentences wherever possible. Avoid usage of phrases such as “could be,” “should be,” and “would be” throughout the Protocol. Any text that cannot be written in the imperative tense may be added as a “Note.” However, notes should be concise and used sparingly.

Response: The protocol has been revised to only include statements written in the imperative tense or in a “Note”.

6. The Protocol should contain only action items that direct the reader to do something. Please move the discussion about the protocol to the Discussion.

Response: We have moved portions of sections 1.2, 3.1, 3.2.1, 3.3.2.1, 3.3.2.3, 3.3.2.3.1, and 4.1 of the original manuscript to the Discussion section of the revised manuscript.

7. There is a 2.75 page limit for filmable content. Please highlight 2.75 pages or less of the Protocol (including headings and spacing) that identifies the essential steps of the protocol for the video, i.e., the steps that should be visualized to tell the most cohesive

story of the Protocol. Remember that non-highlighted Protocol steps will remain in the manuscript, and therefore will still be available to the reader.

Response: Approximately 2.5 pages of the Protocol, including headings and spacing, is highlighted for filming.

8. Please ensure that the highlighted steps form a cohesive narrative with a logical flow from one highlighted step to the next. Please highlight complete sentences (not parts of sentences). Please ensure that the highlighted part of the step includes at least one action that is written in imperative tense. Please do not highlight any steps describing euthanasia.

Response: The highlighted portions of the revised manuscript are complete sentences that include at least one action written in imperative tense and that form a cohesive narrative. There are no steps describing euthanasia.

9. Please include all relevant details that are required to perform the step in the highlighting. For example: If step 2.5 is highlighted for filming and the details of how to perform the step are given in steps 2.5.1 and 2.5.2, then the sub-steps where the details are provided must be highlighted.

Response: Relevant sub-steps of the protocol are now highlighted.

10. Discussion: Please also discuss critical steps within the protocol.

Response: As noted in our response to comment 6 above, discussion about the protocol has been moved to the Discussion. Moreover, we now include in the Discussion notes on the most critical steps in the protocol, particularly several aspects of calibration.

11. References: Please do not abbreviate journal titles. Please include volume and issue numbers for all references.

Response: The references now include the full title of all journals, as well as volume and issue numbers.

Reviewers' comments:

Reviewer #1:

Manuscript Summary:

This manuscript outlines a protocol for conducting real-world studies with infants and toddlers and recording their eye movements via head-mounted eye trackers. These methods are of great interest to the developmental science community and the authors of this protocol have been pioneering in this work so I am certain this protocol will be well received. However, the success of their work somewhat overshadows some major complexities to gathering good data with this target developmental population which I believe must be emphasised in a rewrite before submission.

Major Concerns:

-There should be some comments/caveats on data quality. Just because you get a good eye image and the system calibrates does not mean the data is good enough quality for

your intended analysis. See Holmqvist et al (2011) and Saez De Urabain, Johnson and Smith (2014).

Response: This has been added to the Discussion as follows: “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 ²⁰).”

-It should be highlighted more that manual ROI coding requires a lot of time and resources as automated tools are unreliable. Authors should include a time estimate for cleaning each minute of raw gaze data and for hand-coding ROIs. Also include links to possible automation techniques if they believe these may be possible.

Response: Time estimates depend strongly upon how the coding is done. We now include the following statements in the discussion: “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.”

We are not aware of any reliable automation techniques and thus have not included any in the manuscript. Tools like the neural network YOLO (Redmon, Divvala, Girshick, & Farhadi, 2016) hold promise for automated object detection, but the output of such networks are bounding boxes, not object segmentation at the pixel level, and as mentioned in the Protocol we do not recommend trusting the crosshairs at the pixel level for various reasons, for instance there can be overlapping objects in the scene view or the pupil can be incorrectly located in the eye image.

-Lines 575-576: The authors point out that 70% of already collected data is usable/of high-quality, but given the focus on methods, the discussion should also point out the actual drop-out rate (i.e. how much usable/high-quality data is obtained of all participants who took part in the study, including those toddlers who did not start the study due to, e.g., fussiness as they did not tolerate the cap/ET gear). This should be noted as this may affect labs with limited resources or participant pools. Head-mounted eye tracking is specifically aversive to many young children whose resistance may not even allow the study to begin.

Response: We now realize that it was unclear to what the estimate of “70%” referred. We have revised this section to clarify that this number refers to the percentage to participants from whom we get sufficient high-quality data, as well as to note the causes for the failure to obtain good data from the remaining 30% of infants. Specifically, the revised manuscript reads: “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 participants¹⁸. 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...”

Minor Concerns:

Lines 34, 82, 546: The authors mention "infants". This protocol focuses on toddlers, and although the protocol is similar for infants, there are some additional challenges with infant testing that were not discussed in the manuscript, so the protocol should either mention this or focus on toddlers/young children only.

Response: We have maintained use of both of the terms “infant” and “toddler” because this protocol is designed for use with children aged 9-24 months, which we now state explicitly in the revised manuscript. We find that within this age range, there are large differences in attention and motor ability between individuals even of the same age, often larger even than mean differences between individuals of different ages, hence we have not specified particular ages for which the various methodological challenges might apply. Nevertheless, in addition to noting that the age of the participants may affect the length of the session, we have also added the following statements to the revised manuscript: “...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}”.

Lines 93-95, Eye-tracking equipment: To my knowledge, there is only one commercially available system for young children, namely Positive Science. The images in this manuscript suggest the authors have modified the Pupil Labs system to work with a custom made infant cap but this should be made clear so the readers do not leave with the false impression that such systems can be commonly bought off the shelf. This section and the following sections suggest that options are available on the market.

Response: The revised manuscript now clarifies this issue in several places. Protocol section 1.1 now states: “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.” More detail on how to do this is now provided in protocol section 1.1.3.1: “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.” Protocol section 2.2.2 has also been revised to read: “2.2.2 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."

Lines 112/208-209: How do you put the eye camera out of a toddler's view? It will always be in the view and is a major challenge for HMET testing.

Response: These sections have been revised to specify that the cameras should not be *in the center of* the child's view.

Line 115: "more normal" Bit awkward.

Response: This sentence has been revised to read: "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."

Point 2.3.1.1 - You should also mention the benefits of binocular eye tracking if the gaze is expected to shift in depth during the testing protocol. Monocular tracking will create an offset on the world camera relative to the actually fixated point if monocular gaze is used as the actual point is a consequence of vergence between the two eyes.

Response: The revised manuscript includes the following note after Protocol section 2.4.2.1, including references for more information on binocular eye tracking should the reader be interested: "Note: Binocular eye tracking is a developing technology^{13,14} that promises advances in tracking gaze in depth."

Calibration section, general comments:

Parts of the calibration section could be restructured in their order. Some parts are repetitive. The calibration section should be divided into what to do/consider during the study when getting points, and what to do/consider when calibrating in the software.

Response: We have chosen to organize the Protocol sequentially rather than topically and have therefore left the section on what to do during the study to obtain calibration points as part of Protocol section 2, as this step is part of the data collection process, rather than combining it with section 3 on performing the calibration in the software. We have, however, added the italicized words in the section titles pasted below to make it more clear when each of these steps take place: section 2.3 is now titled "Obtain Points *During the Study* for Offline Calibration" and section 3 is titled "Calibrate the ET Data *using Calibration Software*" (italics not used in the manuscript). We have also revised and/or deleted portions of both Protocol sections 2.3 and 3 to reduce redundancy and reordered portions of section 3 for better flow.

Example for repetition: Section 3.2.1 (line 279) is repeated in 2.4 (lines 252); Section 3.3.2.1. mentions accurate pupil tracking as a requirement when previous sections already mention it (e.g. 3.3.1.).

Response: These Protocol sections have all been revised to reduce redundancy. Note, however, that we have kept these sections separate rather than combining them because, despite referencing similar information, they instruct the reader to do different

things at different stages. For instance, section 2.5 (2.4 in the original manuscript) instructs the reader to take note during the study of the timing of any changes in the ET's position, whereas section 3.1.1 (3.2.1 in the original manuscript) instructs the reader to later create separate calibrations for those sections of the recordings. To better integrate these sections, we have added references between the sections, for instance we have added "(see Section 3.1.1)" to Protocol section 2.5.

Example for structure: Section 3.2.2. (line 288) could go in the previous section for "obtaining points for offline calibration".

Response: As noted in our response to the previous comment by the Reviewer, we have opted not to combine these sections because, despite referencing similar information, they instruct the reader to do different things at different stages.

Calibration section, specific comments:

Line 232, Calibration: It could be recommended to use a few more points than needed for offline calibration (if this is feasible for the study). This ensures that if a calibration point turns out unusable later, an alternative one can be used. Also, it should be noted not to go too fast between points (e.g. the laser should not be immediately moved to the next location once the researcher sees the child focused on the point). This makes it easier to detect the moment of fixation later, when calibrating.

Response: Both of these recommendations have been added to Protocol section 2.4.1, which now reads: "2.4.1. Alternate between drawing attention to different locations that require large angular displacements of the eye. 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. 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. If feasible, obtain more calibration points than needed in case some turn out to be unusable later."

Line 249, adding calibration points: If feasible, calibration points could be added during the recording if the session is longer.

Response: We have added this to Protocol section 2.4.3, which now reads: "2.4.3 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."

Line 302: "use pupil only" - this can, however, often lead to inaccurate tracking and occasionally result in displacements

Response: In cases where the pupil and corneal reflection are both reliably and consistently detected, using both leads to higher accuracy compared with using the pupil only. Nevertheless, we have found in practice it is much easier to position the eye camera to detect the pupil, and much more difficult to detect both the pupil and corneal reflection consistently when we run a study for much longer than 5 minutes, a relatively long period of time for infants/toddlers. Thus, we have kept this point in the manuscript, but qualified

it with the phrase in parenthesis as follows: “3.2.1 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).”

Line 360, assessing quality: "indicate known gaze positions, such as the dots produced by a laser pointer during calibration" - however, if calibration points themselves are used for calibration, then these points cannot be also used for checking data quality, unless post-hoc calibration points are provided.

Response: This is a great point, which we have now added to Protocol section 3.3 as follows: “3.3 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.”

ROI coding

Lines 390-392: Consider rewriting for clarity.

Response: This section has been heavily revised for clarity and moved to the discussion. We have also added an example to illustrate our point, as follows: “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...”

Line 423, Use the POG track as a guide, not as ground-truth: This section should come before the previous section so that the previous section is more readily understandable in terms of why there is a need to use the pupil as information (since the crosshair is available).

Response: This section now comes before the section on using pupil movement, as recommended.

Point 4.1.- This point is important but will only be understood if "fixation" is properly defined and the issues associated with classifying it in mobile data are expanded. Please do so with reference to Holmqvist et al (2011) and Smith & Saez De Urabain (2017)

Response: We have moved this information to the Discussion section to allow for elaboration, including a definition of fixations and an example to illustrate the complexity of the issue. This section is pasted below for your reference. We also refer the reader to the citations the Reviewer mentioned, but have chosen not to go into more detail in the manuscript because our focus is on coding ROI rather than fixations.

Paragraph 5 of the Discussion now reads: “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 ^{15,22}...”

Additionally, because fixation detection is so difficult in head-mounted eye tracking and we do not want readers to get caught up in the issue of fixations versus saccades, we have also removed Protocol section 4.2.2.2, pasted here, as it mentions saccades and is not critical to the protocol: “4.2.2.2 Code POG during saccades – the several frames it may take for the eye to shift from visually attending one object to attending another object – as a separate ROI category rather than as the ROI the POG track may happen to indicate on that particular frame. Use contextual information from the preceding and following frames to identify these transitional eye movements.”

New references:

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

Saez de Urabain, I.R. and Johnson, Mark H. and Smith, Tim J. (2014) GraFIX: a semiautomatic approach for parsing low- and high-quality eye-tracking data. Behavior Research Methods 47 (1), pp. 53-72. ISSN 1554-3528.

Smith, Tim J. and Saez de Urabain, I.R. (2017) Eye tracking. In: Hopkins, B. and Geangu, E. and Linkenauger, S. (eds.) Cambridge Encyclopedia of Child Development. Cambridge, UK: Cambridge University Press, pp. 97-101. ISBN 9781107103412.

Response: Thank you for providing these full references.

Reviewer #2:

Manuscript Summary:

This protocol provides guiding principles and practical recommendations for researchers using head-mounted trackers in the laboratory and in naturalistic settings. It also includes

the kind of fine-grained data generated from this type of eye-tracking equipment in free-play.

Major Concerns:

Introduction could be improved in documenting research showing how head-mounted eye tracking could provide unique information about infant cognitive, language and motor development. It is unclear what UNIQUE information the ET system brings to infant cognition. In other words, is there any evidence that infants who pay more attention to a certain object learn the word for it earlier? Categorize these kind of objects earlier?

Response: We have revised the final paragraph of the introduction to note some of the unique contributions of head-mounted eye tracking. Specifically, this paragraph now reads: "...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 developments⁴⁻⁸. The use of these techniques has transformed understandings of joint attention⁷⁻⁹, sustained attention¹⁰, changing visual experiences with age and motor development^{4,6,11}, and the role of visual experiences in word learning¹². 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."

We also describe in the Discussion how head-mounted eye tracking provides greater opportunity for capturing individual differences in children's visual attention and has yielded new insights into infants' looking to parents' faces, compared to screen-based eye tracking research. That section is pasted here for your reference: "...[head-mounted eye tracking] 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^{4,5,7-10}."

Fig 5 shows some data on 2 different children. How old were the children? Where was the parent located? The low LT at the face will be directly affected by these variables. There is huge variability in the number of faces infants see during the first few months of life, for example.

Response: The children were both approximately 18 months of age, which is now noted in the results section of the manuscript. We agree that the position of the parent relative to the child will affect amount of looking to the parent's face and for that reason parents were instructed to play freely and were not told to stay in a particular position

relative to their child. Child 1's parent chose to sit at a 90 degree angle to the right and in front of the child. Child 2's parent chose to sit to the left of the child, turned at a 45 degree angle from the child.

Not quite clear how Fig 5A is a stream of data points.

Response: This is now clarified in the Legend for Figure 5, which reads: "(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..."

Discussion: "this method is an alternative.." to what? Babies also have control over where and how long they focus their visual attention in traditional free play sessions. They typically code how long they manually explore objects.

Response: The revised sentence now reads "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."

Minor Concerns:

No mention of Fig 2DEF in text.

Response: Figure 2C-F is now referenced in Protocol section 2.3.2.1.

Please clarify what you mean by cap (no visor?) vs bandana.

Response: The first time we mention a "cap" in Protocol section 1.1 we now direct the reader to Figures 1 and 2 for a visual example of what we are referring to. We have also revised Protocol section 2.2.2, which gives other examples of what we mean by "cap": "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."

Any data on tolerance of the system as a function of age? The pupil camera must be distracting starting at a certain age.

Response: Please see our Response to a similar comment from Reviewer 1 above, which we also paste here for easy reference: "We have maintained use of both of the terms "infant" and "toddler" because this protocol is designed for use with children aged 9-24 months, which we now state explicitly in the revised manuscript. We find that within this age range, there are large differences in attention and motor ability between individuals even of the same age, often larger even than mean differences between individuals of different ages, hence we have not specified particular ages for which the various methodological challenges might apply. Nevertheless, in addition to noting that the age of the participants may affect the length of the session, we have also added the following statements to the revised manuscript: "...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}."

How long should the parent have the child wear a cap before testing?

Response: As stated in Protocol section 2.2.1, the goal of having the child wear a cap before testing is to "get them accustomed to having something on their head". We intentionally do not specify a particular duration of time for this as children will take different amounts of exposure to a cap to become comfortable with it, and because this step will often be constrained by how far in advance an appointment can be scheduled. Moreover, leaving the recommendation open to the parent's discretion helps to ensure that the parent does not feel burdened by this recommendation.