

# Journal of Visualized Experiments

## A view of their own: Capturing the egocentric view of infants and toddlers with head-mounted cameras --Manuscript Draft--

Article Type:	Invited Methods Article - JoVE Produced Video
Manuscript Number:	JoVE58445R1
Full Title:	A view of their own: Capturing the egocentric view of infants and toddlers with head-mounted cameras
Keywords:	egocentric vision; head-mounted cameras; visual scenes; Development; infant; toddler
Corresponding Author:	Jeremy Isaac Borjon Indiana University Bloomington Bloomington, UNITED STATES
Corresponding Author's Institution:	Indiana University Bloomington
Corresponding Author E-Mail:	jborjon@iu.edu
Order of Authors:	Jeremy Isaac Borjon Sara Elizabeth Schroer Sven Bambach Lauren Krogh Slone Drew Hamilton Abney David J Crandall Linda B Smith
Additional Information:	
Question	Response
Please indicate whether this article will be Standard Access or Open Access.	Standard Access (US\$2,400)
Please indicate the <b>city, state/province, and country</b> where this article will be <b>filmed</b> . Please do not use abbreviations.	1101 E 10th St, Bloomington, IN 47405



DEPARTMENT OF  
PSYCHOLOGICAL AND  
BRAIN SCIENCES

---

INDIANA UNIVERSITY  
Bloomington

May 9<sup>th</sup>, 2018

Dear Editor,

We submit the attached manuscript entitled “A view of their own: Capturing the egocentric view of infants and toddlers with head-mounted cameras” for consideration in the *Journal of Visualized Experiments* as an invited submission. The manuscript is an original work, is not currently being considered elsewhere for publication, and will not be submitted elsewhere prior to your editorial decision. Jeremy I. Borjon will be the corresponding author for this manuscript. Their contact information is listed below.

We believe this work is worthy of consideration for the *Journal of Visualized Experiments*. It provides an important framework for capturing the egocentric views of infants using head-mounted cameras and provides a novel contribution to work in developmental psychology seeking to better understand the visual ecology of the developing child.

Sincerely,

Jeremy I. Borjon  
Sara E. Schroer  
Sven Bambach  
Lauren K. Slone  
Drew H. Abney  
David J. Crandall  
Chen Yu  
Linda B. Smith

Address correspondences to:  
Department of Psychological and Brain Sciences  
Indiana University  
1101 E 10th Street  
Bloomington, IN 47405-7007  
jborjon@iu.edu

**TITLE:**

Capturing the Egocentric View of Infants and Toddlers with Head-Mounted Cameras

**AUTHORS & AFFILIATIONS:**

Jeremy I. Borjon<sup>1\*</sup>, Sara E. Schroer<sup>1\*</sup>, Sven Bambach<sup>2</sup>, Lauren K. Slone<sup>1</sup>, Drew H. Abney<sup>1</sup>, David J. Crandall<sup>2</sup>, Linda B. Smith<sup>1</sup>

<sup>1</sup> Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA

<sup>2</sup> School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA

\*These authors contributed equally to this work.

Corresponding Author:

Jeremy I. Borjon

Email Address: jborjon@iu.edu

Email Addresses of Co-authors:

Jeremy I. Borjon [jborjon@iu.edu]

Sara E. Schroer [seschroe@iu.edu]

Sven Bambach [sbambach@indiana.edu]

Lauren K. Slone [leslone@indiana.edu]

Drew H. Abney [dhabney@indiana.edu]

David J. Crandall [djcran@indiana.edu]

Linda B. Smith [smith4@indiana.edu]

**KEYWORDS:**

Egocentric vision, head-mounted cameras, visual scenes, development, infant, toddler

**SUMMARY:**

Infants and toddlers view the world in a fundamentally different way from their parents. Head-mounted cameras provide a tractable mechanism to understand the infant visual environment. This protocol provides guiding principles for experiments in the home or laboratory to capture the egocentric view of toddlers and infants.

**LONG ABSTRACT:**

Infants and toddlers view the world, at a basic sensory level, in a fundamentally different way from their parents. This is largely due to biological constraints: infants possess different body proportions than their parents and the ability to control their own head movements is less developed. Such constraints limit the visual input available. This protocol aims to provide guiding principles for researchers using head-mounted cameras to understand the changing visual input experienced by the developing infant. Successful use of this protocol will allow researchers to design and execute studies of the developing child's visual environment set in the home or laboratory. From this method, researchers can compile an aggregate view of all the possible items in a child's field of view. This method does not directly measure exactly what the child is looking at. By combining this approach with machine learning, computer vision algorithms, and

hand-coding, researchers can produce a high-density dataset to illustrate the changing visual ecology of the developing infant.

## **INTRODUCTION:**

For decades, psychologists have sought to understand the environment of the developing infant, which William James famously described as a “blooming, buzzing confusion<sup>1</sup>.” The everyday experiences of the infant are typically studied by filming naturalistic play with social partners from a third-person perspective. These views from the side or above typically show cluttered environments and a daunting number of potential referents for any new word an infant hears<sup>2</sup>. To an outside observer, James’s description is apt, but this stationary, third-person perspective is not the way an infant sees the world. An infant is closer to the ground and can move through their world, bringing objects closer for visual exploration. A third-person view of a parent-infant interaction is illustrated in **Figure 1**. Highlighted are the fundamental differences between their perspectives. Perhaps, the input that infants receive is not nearly as chaotic as anticipated by parents and researchers. The goal of methods with head-mounted cameras is to capture the infant experience from a first-person view in order to understand the visual environment available to them throughout development.

Head-mounted cameras, worn on a hat or headband, provide a window into the moment-to-moment visual experiences of the developing infant. From this perspective, the study of the structure and regularities in the infant’s environment becomes apparent. Head-mounted cameras have revealed infants’ visual experiences to be largely dominated by hands, both their own and their social partner’s, and that face-looks, once considered imperative for establishing joint attention, are much scarcer than anticipated<sup>3</sup>. Head-mounted cameras have also shown that infants and their caregivers create moments when objects are visually dominant and centered in the infant’s field of view (FOV), reducing the uncertainty inherent to object-label mapping<sup>4</sup>.

Head-mounted cameras capture the infants’ first-person view based on head movements. This view is not perfectly synchronous with, or representative of, infant eye movements, which can only be captured in conjunction with an eye-tracker. For instance, a shift of only the eyes while keeping the head stationary, or a shift of the head while keeping the eyes fixed on an object, will create a misalignment between the infants’ actual FOV and the one captured by the head camera. Nonetheless, during toy play, infants typically center the objects they are attending to, aligning their head, eyes, and the location of the toy with their body’s midline<sup>5</sup>. Misalignments are rare and are typically created by momentary delays between an eye shift and the accompanying head turn<sup>3</sup>. Therefore, head-cameras are not well suited to capturing the rapid dynamics of shifts in attention. The strength of head-mounted cameras lies in capturing the everyday visual environment, revealing the visual content available to infants.

The following protocol and representative results will demonstrate how head-mounted cameras can be used to study the visual environment of infants and toddlers.

## **PROTOCOL:**



The following procedure to collect data on infant and toddler's visual experiences in the laboratory and at home was approved by the Indiana University Institutional Review Board. Informed consent was obtained from the infant's caregiver.

## 1. Choose a Head Camera

**Note:** There are numerous small, lightweight, and portable cameras readily available for purchase (**Figure 2**).

1.1. Choose a head camera that is unobtrusive and will not influence the scenes being recorded.

1.1.1. Mount the camera onto a snug hat or headband using a temporary adhesive or securely mount the camera onto a small plastic plate attached to the headband. Position the camera at the child's brow (**Figure 2B**, left).

1.1.1.1. Depending on the shape of the camera, mount the camera by sewing small cloth loops into a headband or hat (**Figure 2B**, center and right).

1.1.1.2. Ensure that the hat or headband are adjustable to achieve a snug and comfortable fit on the child's head (**Figure 2A**).

**Note:** A camera located directly between the eyes of the infant is theoretically ideal, however this is not quite possible with current technology. Small cameras that can be placed low on the forehead are becoming increasingly available.

1.2. If cameras are being sent home with the parents, ensure that parents can use them without any technical help.

1.2.1. Before the parents leave the laboratory, train them on how to operate the head camera.

1.2.2. Send the parents home with the camera to use, the head-band or hat, and a hand-out describing the procedure to turn on the camera and record data.

1.3. Ensure that the cameras are light enough for infants to wear and forget about.

**Note:** The recommended weight is less than 30 g<sup>3</sup>. Any chosen camera should not heat up during use and must be durable under heavy use.

1.4. If the experimental setup requires the children to be freely-moving, store recorded video to an internal storage card. Otherwise, use cameras tethered to the data collection computer.

1.5. Use a camera with a high-quality lens to better leverage recent advances in machine learning algorithms which segment visual scenes into regions of interest.

1.5.1. If visual scenes will be manually coded by researchers, use a camera that can capture images at sufficient quality for manual inspection.

1.6. Ensure that any battery-powered camera is capable of recording for the desired amount of time.

1.7. Shortly before beginning the experiment or sending the camera home with the parents, test the head camera to ensure it is working appropriately.

## 2. Data Collection in the Laboratory

**Note:** Head-mounted cameras can be easily added to most experiments.

2.1. Have 2-3 experimenters place the camera onto the child's head: one experimenter places the head camera, one monitors the head-camera view, and, if needed, one distracts the child.

2.1.1. Ask the parent to ensure that the child remains calm and to distract the child during the placement process.

2.2. Perform the camera placement in three steps as follows.

2.2.1. Desensitize the infant to hand actions near their head.

2.2.1.1. Ask the parent to lightly touch or stroke the infant's head and hair several times.

2.2.1.2. Ask the experimenter placing the hat on the infant's head to do the same as in 2.2.1.1.

2.2.2. Have the experimenter place the head-mounted camera when the child is distracted.

2.2.2.1. Use push-button toys to keep the child's hands busy.

2.2.2.2. Have the distracting experimenter or parent help at this stage by gently pushing the child's hands toward the engaging toy so that the child's hands do not go to the head.

2.2.3. Tighten the hat on the child's head and adjust the head camera when the child is engaged with the toy.

2.2.3.1. Adjust the camera so that when the infant holds an object in front of his/her face, the object is centered in the head camera FOV.

2.2.3.2. If the child is sitting, adjust the camera so that it captures most of the child's lap when the child looks down.

2.3. After placing the camera on the child's head, ask the experimenter to leave the room and begin the recording.

2.4. In the event that a camera is moved out of place or removed, re-enter the room to correct the camera.

2.4.1. Terminate the experiment if the child does not tolerate the camera being reapplied.

**Note:** For recording natural environments in the home, first fit a hat and camera to the individual infant and show parents how to position the camera. The design and fit of the camera must ensure that parents will be able to put the hat on their child without technical help.

### 3. Data Collection for the Parent-Infant Study

Note: The following representative method for head-cameras uses naturalistic toy play in the lab to demonstrate the type of analyses that can be conducted on the egocentric views of infants and their parents (**Figure 3A**).

3.1. Outfit the parent and child with head-mounted cameras, as described in 2.1 and 2.2.

3.2. Use head cameras to capture videos with a resolution of 720 x 1280 pixels at 30 frame/s. Proceed as described in steps 2.3 and 2.4.

3.2.1. Subsample the video stream at one frame every 5 s.

3.2.2. Manually draw bounding boxes around each toy (**Figure 3B**) within view using commercial software or a program developed in-house (see **Figure 3C** for a sample frame).

3.2.2.1. If only part of a toy is visible due to occlusion by other toys or truncation on the edge of the frame, only draw bounding boxes when a toy is easily identifiable and include all visible parts of the toy.

3.2.2.2. For example, if only the doll's leg is visible, draw a bounding box around its leg. If an object occludes half of the doll, leaving the hair and legs visible, then draw a box that includes both hair and legs.

### REPRESENTATIVE RESULTS:

One simple, yet informative, analysis is to count the number of objects in view at each point in time. Since a head camera produces data at approximately 30 Hz (30 images/s), down-sampling the data to 1 image every 5 s helps to produce a more manageable dataset while maintaining a

resolution appropriate for understanding the types of scenes children see. Prior research has demonstrated that visual scenes are slow-changing in infants<sup>3</sup>. A custom script was used to draw bounding boxes around the toys in view. **Figure 4** shows representative results for 1 parent-infant dyad. An independent t-test comparing the number of scenes with a given number of objects between the parent (**Figure 4A**) and the child (**Figure 4B**) revealed this child had a greater number of scenes with fewer objects in view compared to the parent ( $t(78) = 4.58, p < 0.001$ ).

Another informative analysis is to calculate how visually large the objects are in each view. The proportion of the screen taken up by each object in view can be calculated and analyzed. For both parent and child, there is a negative correlation between the number of objects in view and the visual sizes of the objects in that view (**Figure 4C**, Spearman correlation  $r = -0.19, p < 0.001$  and **Figure 4D**, Spearman correlation  $r = -0.23, p < 0.001$ ). That is, if there are more objects in view, each object takes up less of the screen than if there are fewer objects in view. For this dyad, the child captured more scenes with less than 10 objects in view and the parent exhibits a larger number of objects in view. Similar results have been previously reported in the literature<sup>3-5,8-13</sup>.

#### FIGURE AND TABLE LEGENDS:

**Figure 1: An illustrative schematic demonstrating the different views of a parent and their child during play.**

**Figure 2: Examples of head-mounted cameras and their attachments. (A)** Infants and toddlers wearing head-mounted cameras at home and in the lab. **(B)** Examples of ways to attach head cameras to headbands (left, middle) and hats (right).

**Figure 3: The 24 toys used in the representative method. (A, left)** Representative frame from the head camera of a child illustrating a smaller number of objects in view and at visually larger sizes. **(A, right)** A representative frame from the head camera of a parent illustrating their typical view: many objects at visually smaller sizes. **(B)** Toys are consistent in size, ranging from 2-7 inches on the longest dimension, and between 2-3 inches on the shorter dimensions. **(C)** Boxes are drawn around each toy, or part of toy that is visible and identifiable, using an in-house graphical user interface.

**Figure 4: Representative results from a single dyad participating in toy play.** Histograms grouping the number of scenes based on the number of objects in view for the parent **(A)** and child **(B)**. The proportion of the screen taken up by each object in view *versus* the number of objects in view for the parent **(C)** and the child **(D)**. The black line is the line of best fit.

#### DISCUSSION:

This paper outlines the basics for applying head-mounted cameras to infants to capture their egocentric visual scene. Commercially available head cameras are sufficient for the vast majority of studies. Small, lightweight, and portable cameras should be incorporated into a soft fabric hat or headband and applied to the child's head. Once successfully designed and implemented, a variety of experiments can be run, both in laboratory settings as well as in the home

environment. From the videos gathered, aggregate data about the developing infant's visual ecology can be compiled and analyzed.

The most critical step with this method is the application of the head camera onto the child. If done incorrectly, the head camera will be poorly placed and data quality will be diminished or unusable. An incorrect placement could also spur the child to reject the camera and halt the experiment. We will briefly discuss suggestions to ensure success with the application of the head camera. Cameras should be placed on the infant in one move without hesitation. If the researcher is apprehensive to place a camera on the child's head, or if multiple attempts are made, the likelihood of refusal becomes much higher. Experimenters should practice placing hats and camera devices on willing toddlers or mannequins beforehand. When placing the camera, it must be placed low enough on the forehead to ensure a clear view of the scene in front of the face. Slightly angling the camera downwards will guarantee view of the infant's hands during active manipulation. The camera should also be stable and secure on the infant's head. Stable cameras mean stable and clear images. If the headwear jiggles, toddlers can notice this and pull the camera off. For children under 18 months of age, anything drawing attention to the gear increases refusals. This includes having the infant handle the equipment or talking about it before placing it on the child. For children over 18 months of age, talking about the camera beforehand and asking the child's permission to put it on may be more effective. With a trained researcher, success rates in placing head cameras on infants, without the infant fussing out of the experiment, can reach around 75%.

When sending a head-camera home with the families, take considerable time to design the cap/headband and camera placement. The way a parent places a camera on their child's head will not always be at the same level of precision as a trained researcher. Ensure the cap is easy to apply by the parents and ensure the research questions do not demand exacting specifications. If experimental needs require precise placement of the camera on the head, consider running the study in a laboratory setting instead of at home.

Head-cameras will have limitations in what they can capture. Given the location of the camera on the head of the infant, the horizontal views of the infant as they move their head from left to right will be extensively captured. Vertical displacements of the camera, when the infant looks up and down, will be unable to capture the very extremes of the visual scene. This is especially true if the camera is angled slightly downwards on the head of the infant, in order to capture the infant's hands.

Head-mounted cameras have revealed that children have a view of their own. On a fundamental level, toddlers and infants view the world differently than their parents. Toddlers shape their visual experience with their hands: holding and manipulating objects close to their face<sup>4-5,8</sup>. Given the very short arms of toddlers, the object is held close and appears large in the field of view. Such scenes with a clear focal object are often long-lasting, around four seconds in duration, and coincide with the lessening of head movements by the infant<sup>4</sup>. It is important to note, however, that head-mounted cameras do not provide any information as to where the participant is looking. Instead, this protocol can quantitatively describe the range of visual scenes available to

children. Across an entire play session, there is a high probability that the child's eyes are typically centered in the middle of the visual scene in front of the child<sup>5</sup>. Head cameras allow us to investigate the aggregate of the scenes available to a child. For example, how often is a face available for them to look at? How persistent are these scenes with faces? How often are children viewing cluttered (a pile of toys on the floor) *versus* uncluttered (the ceiling or a blank wall) scenes? This egocentric view methodology is best suited for data at the macro-scale, 100 million images collected over days. If the research question needs more fine-grained resolution than these aggregate-level questions, head-mounted eye-tracking may be better suited to capture the exact dynamics of infant vision.

Just as the toddler and adult have different visual experiences, the visual experience of infants and toddlers is not developmentally static. As children grow, the available visual scenes change dramatically, and there is developmental structure in the people and objects visually available to infants of different ages. For instance, when infants are very young, their visual environment is dense with the faces of a very small number of people<sup>10</sup>. From this non-uniform sampling of a few faces, infants can extrapolate and learn to recognize and discriminate between faces they encounter. At around 8-10 months of age, infants are beginning to sit steadily, to crawl, and to play with objects, but their manual skills are still quite limited when compared to older infants. As a result, these infants experience a higher frequency of visual scenes with few objects in view, compared to older infants. Nevertheless, mealtime scenes from these same 8- to 10-month-olds also reveal times of clutter<sup>13</sup>, with each mealtime scene containing many different objects. Despite this clutter, there is a predictable structure to the objects in view: a very small set of objects appear repeatedly. These repeated objects belong to categories encompassing the first words learned by infants<sup>13</sup>. Thus, although it may be easy to look upon children's environment and argue their world is a "blooming, buzzing confusion," head-camera data showing infants' egocentric views reveal that predictable statistical regularities exist in their field of view to dampen the din and confusion.

#### ACKNOWLEDGMENTS:

The authors thank Dr. Chen Yu for his guidance in the creation of this manuscript and for the data used in the Representative Results section. We thank the participating families that agreed to be used in the figures and filming of the protocol as well as Lydia Hoffstaetter for her careful reading of this manuscript. This research was supported by the National Institutes of Health grants T32HD007475-22 (J.I.B., D.H.A.), R01 HD074601 (S.B.), R01 HD028675 (S.B., L.B.S.), and F32HD093280 (L.K.S.). National Science Foundation grants BCS-1523982 (S.B., L.B.S) and CAREER IIS-1253549 (S.B., D.J.C.), the National Science Foundation Graduate Research Fellowship Program #1342962 (S.E.S.), and by Indiana University through the Emerging Area of Research Initiative – Learning: Brains, Machines, and Children (J.I.B., S.B., L.B.S.).

#### DISCLOSURES:

The authors declare no conflicts of interest.

#### REFERENCES:

1. James, W. *The Principles of Psychology*. Henry Holt and Co. New York. (1890).

350 2. Quine, W.V.O. *Word and object: An inquiry into the linguistic mechanisms of objective*  
351 *reference*. The MIT Press. Cambridge, MA. (1960).

352 3. Yoshida, H., & Smith, L. B. What's in view for toddlers? Using a head camera to study visual  
353 experience. *Infancy*. **13**(3), 229-248. (2008).

354 4. Yu, C., & Smith, L. B. Embodied attention and word learning by toddlers. *Cognition*. **125**(2),  
355 244-262. (2012).

356 5. Bambach, S., Smith, L. B., Crandall, D. J., & Yu, C. Objects in the center: How the infant's body  
357 constrains infant scenes. In *Joint IEEE International Conference on Development and Learning and*  
358 *Epigenetic Robotics 2016*, 132-137. (2016).

359 6. Adolph, K. E., Gilmore, R. O., Freeman, C., Sanderson, P., & Millman, D. Toward open behavioral  
360 science. *Psychological Inquiry*. **23**(3), 244–247. (2012).

361 7. Sanderson, P. M., Scott, J. J. P., Johnston, T., Mainzer, J., Wantanbe, L. M., & James, J. M.  
362 MacSHAPA and the enterprise of exploratory sequential data analysis (ESDA). *International*  
363 *Journal of Human-Computer Studies*. **41**(5), 633–681. (1994).

364 8. Pereira, A. F., Smith, L. B., & Yu, C. A bottom-up view of toddler word learning. *Psychonomic*  
365 *Bulletin & Review*. **21**(1), 178–185. (2014).

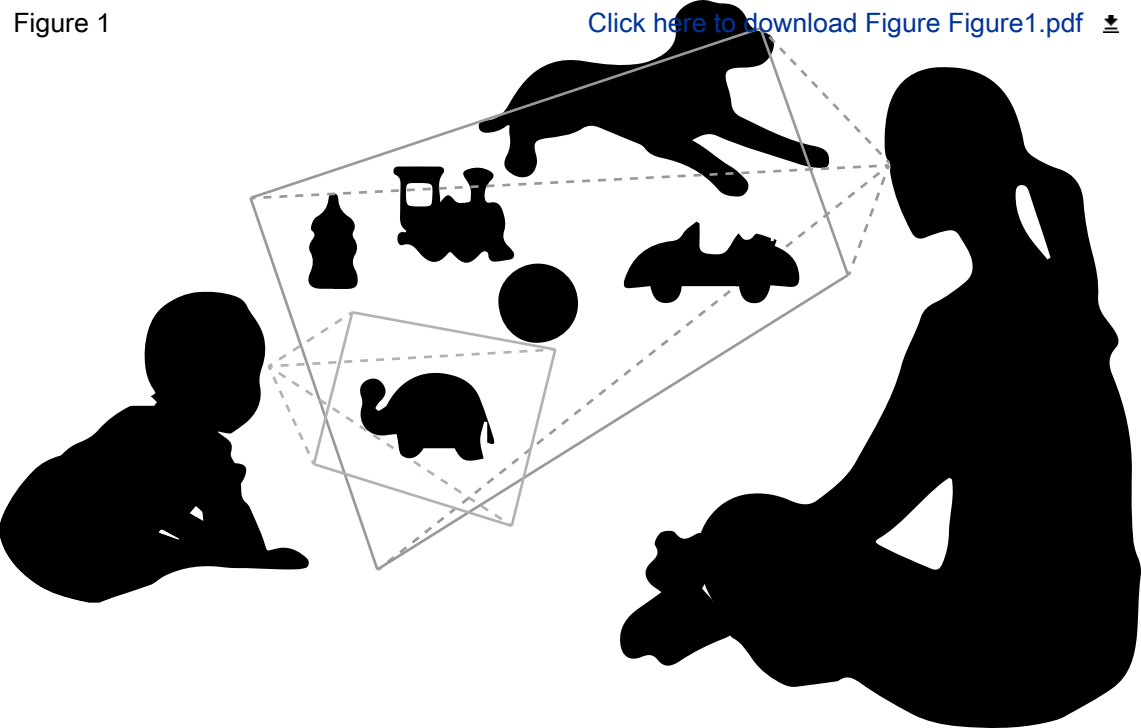
366 9. Yu, C. & Smith, L.B. Joint Attention without Gaze Following: Human Infants and Their Parents  
367 Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLOS ONE*. **8**(11),  
368 e79659. (2013).

369 10. Jayaraman, S., Fausey, C. M., & Smith, L. B. The faces in infant-perspective scenes change over  
370 the first year of life. *PLOS ONE*. **10**(5), e0123780. (2015).

371 11. Fausey, C. M., Jayaraman, S., & Smith, L. B. From faces to hands: Changing visual input in the  
372 first two years. *Cognition*. **152**, 101-107. (2016).

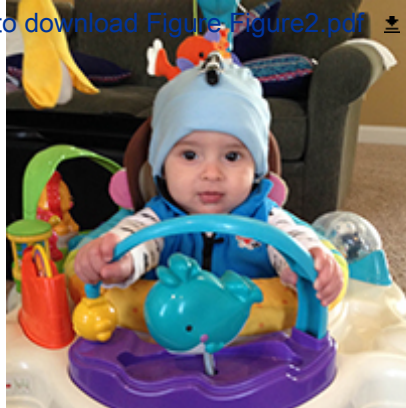
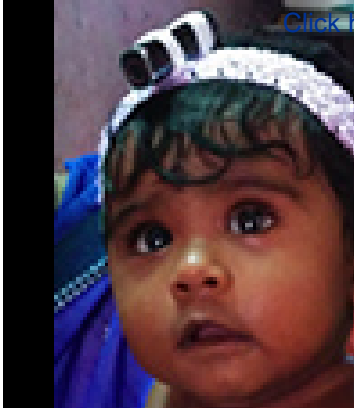
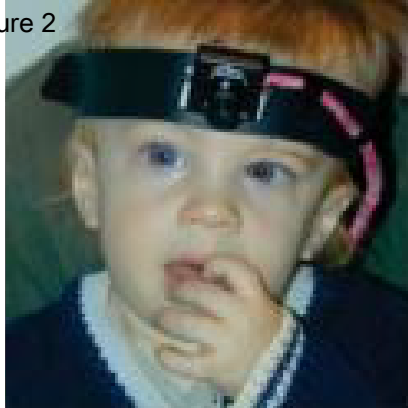
373 12. Jayaraman, S., Fausey, C. M., & Smith, L. B. Why are faces denser in the visual experiences of  
374 younger than older infants? *Developmental Psychology*. **53**(1), 38. (2017).

375 13. Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. Real-world visual statistics and infants'  
376 first-learned object names. *Philosophical Transactions of the Royal Society B, Biological Sciences*.  
377 **372**, 20160055. (2017).



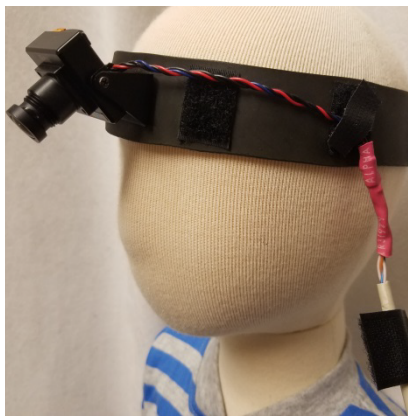


A. Figure 2



[Click here to download Figure2.pdf](#)

B.





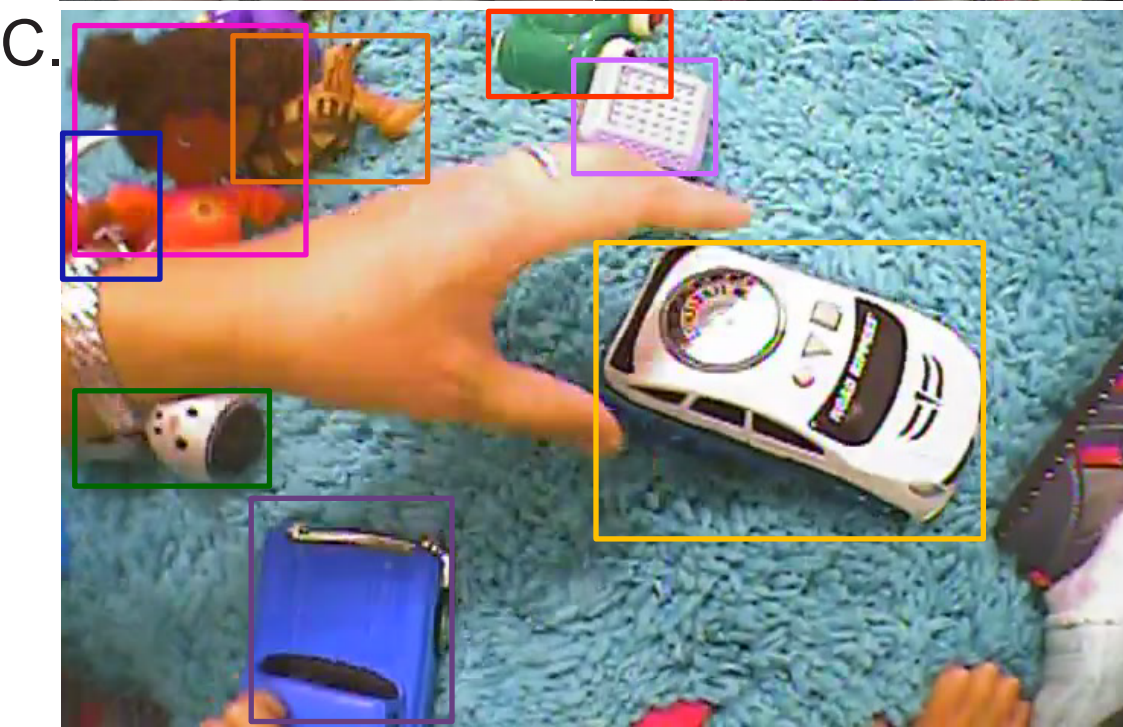
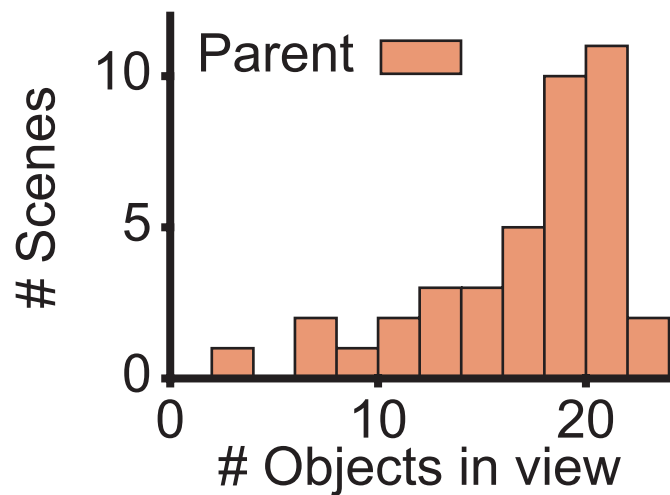


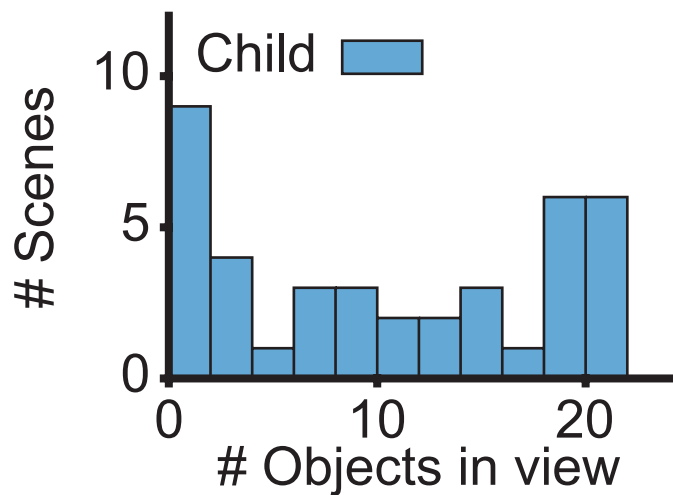
Figure 4

A.

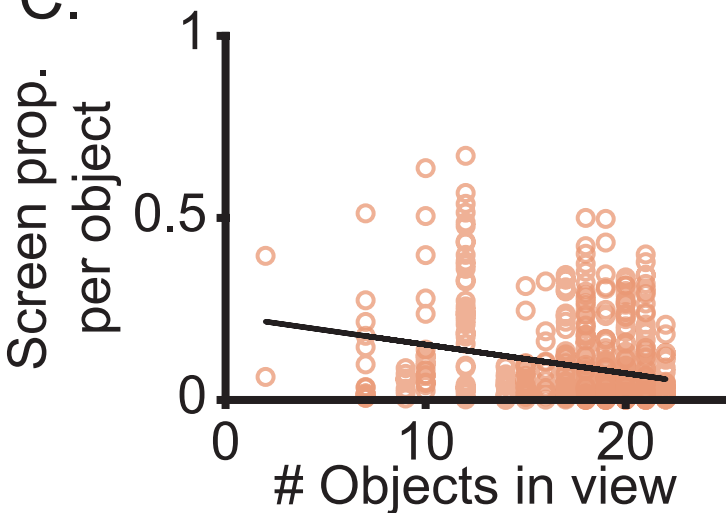


B.

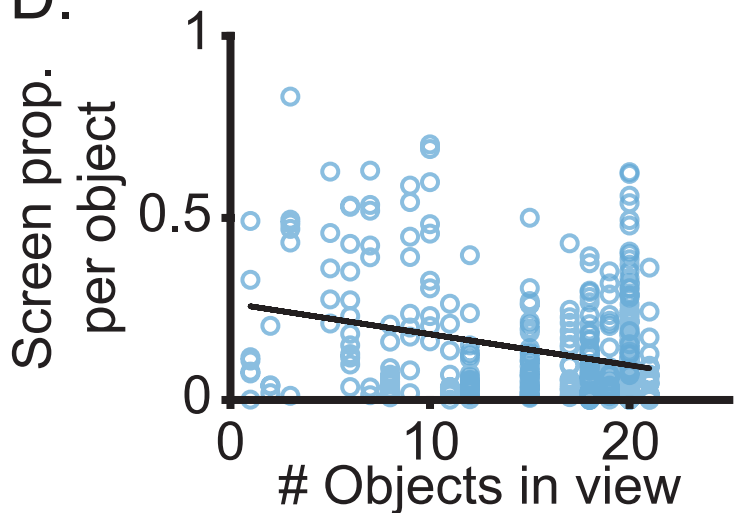
[Click here to download Figure Figure4\\_v2.pdf](#)



C.



D.



Name of Material/ Equipment	Company	Catalog Number	Comments/Description
Head-camera	Looxcie	Looxcie 3	
Head-camera	Watec	WAT-230A	
Head-camera	Supercircuit	PC207XP	
Head-camera	KT&C	VSN500N	
Head-camera	SereneLife	HD Clip-On	
Head-camera	Conbrov	Pen TD88	
Head-camera	Mvowizon	Smiley Face Spy Button	
Head-camera	Narrative	Clip 2	
Head-camera	MeCam	DM06	





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

## ARTICLE AND VIDEO LICENSE AGREEMENT

Title of Article:

A view of their own: Capturing the egocentric view of infants and toddlers with head-mounted cameras

Author(s):

Jeremy I. Borjon, Sara E. Schroer, Sven Bambach, Lauren K. Stone, Drew H. Abney, David J. Crandall, Chen Yu, Linda B. Smith

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:	Jeremy I. Borjon	
Department:	Psychological and Brain Sciences	
Institution:	Indiana University, Bloomington	
Article Title:	A view of their own: Capturing the egocentric view of infants and toddlers with head-mounted cameras	
Signature:		Date: 5/9/18

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](mailto:submissions@jove.com) or call +1.617.945.9051



DEPARTMENT OF  
PSYCHOLOGICAL AND  
BRAIN SCIENCES

INDIANA UNIVERSITY  
Bloomington

June 28<sup>th</sup>, 2018

Dear Editor and Reviewers,

Thank you for your careful consideration of our manuscript. We have revised the draft according to your comments and will address them in the pages below.

Further, per the author's request, we have removed Dr. Chen Yu from the list of authors for this manuscript.

Sincerely,

Jeremy I. Borjon  
Sara E. Schroer  
Sven Bambach  
Lauren K. Slone  
Drew H. Abney  
David J. Crandall  
Linda B. Smith

Address correspondences to:  
Department of Psychological and Brain Sciences  
Indiana University  
1101 E 10th Street  
Bloomington, IN 47405-7007  
[jborjon@iu.edu](mailto:jborjon@iu.edu)



### **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. The JoVE editor will not copy-edit your manuscript and any errors in the submitted revision may be present in the published version.

*This version has been thoroughly proofread for spelling and grammar issues.*

2. Figure 4: Please capitalize the first word in the labels on the x and y axis (i.e., Screen, Objects).

*The change has been made.*

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

*We reworked the introduction to make the overall goal of this method clearer, specifically lines 62-64. "The goal of this method is to capture the infant experience from a first-person view in order to understand the visual environment available to them throughout development."*

4. Please remove all commercial language from your manuscript and use generic terms instead. All commercial products should be sufficiently referenced in the Table of Materials and Reagents. For example: Looxcie 3, Watec (WAT-230A), Supercircuits (PC207XP), KT&C model VSN500N, SereneLife HD Clip-On, Conbrov Pen TD88, Mvowizon Smiley Face Spy Button, the Narrative Clip, McCam, LENA™, LENA Foundation, etc.

*All commercial language has been removed from the manuscript.*

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. Please include all safety procedures, etc. Please revise 1.1-1.8, 2.1, 2.2, 3.1-3.3, as well as their sub-steps accordingly.

*We revised the protocol to ensure it is in the imperative tense and removed all "should, could and would" statements.*

6. 2.4 (including its sub-steps): The Protocol should contain only action items that direct the reader to do something. Please move the discussion about the protocol to the Discussion.

*We have moved this portion of the Protocol to the Discussion.*

7. Lines 249-253: Please move details of the methodology to the Protocol.

*We moved Lines 249-253 to the Protocol.*

8. Line 266: Please include a space between all numbers and their corresponding units: 30 Hz, etc.

*A space has been included in line 266 and we confirmed the rest of the manuscript is to the same standard.*

9. As we are a methods journal, please revise the Discussion to explicitly cover the following in detail in 3-6 paragraphs with citations:

- a) Critical steps within the protocol
- b) Any modifications and troubleshooting of the technique
- c) Any limitations of the technique
- d) The significance with respect to existing methods
- e) Any future applications of the technique

*The discussion has been revised to explicitly cover the steps mentioned.*

10. Please remove the embedded Table of Materials from the manuscript. The uploaded xls/xlsx file is sufficient.

*The embedded Table of Materials has been removed.*

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

*We changed the journal titles to their full length.*

### **Reviewers' comments:**

#### **Reviewer #1:**

Manuscript Summary:

This paper describes recommended equipment and protocol for recording the egocentric view of toddlers and infants by using a head-mounted camera. The protocol should be helpful for other researchers to design and conduct experiments to study the development of child's visual environment set out of a laboratory setting.

Overall, the paper is well written and provides useful information for those who plan to carry out experiments to analyze egocentric views of infants and toddlers with head-mounted cameras. I only have minor comments to be addressed before the paper is published.

1. A part of the manuscript is highlighted (L99-L104, for instance). What does this mean?

*The highlighted portions indicate which parts of the manuscript will be filmed by JoVE's production team.*

2. The reported protocol includes different types of information. Some are guidelines for choosing equipment and designing experiments (e.g., 1.1.1 and 1.1.2), while others are more like observations and know-how from the authors' own experiments (e.g., 1.7 and 2.1). It would be better if the description of the protocol is reorganized such that guidelines and observations can be more easily understand without confusion.

*Thank you for this suggestion. We went ahead and moved any discussion of the method to the Discussion section.*

3. Some items such as 3.1, 3.2.2, and 3.3 seem to be obvious and may be omitted.

*We have omitted the items noted. We further removed the Representative Analysis section from the manuscript and replaced it with a Representative Method.*

4. In L110, "56.7 grams of less": The number is very specific. How did you get this?

*This number was taken from our own internal documents. We decided to include a reference for the recommended weight of a head camera and changed the recommended weight to 30 grams, which is in line with the reference provided.*

## **Reviewer #2:**

This paper provides a set of guidelines for researchers interested in collecting first-person video data with infants in lab and home settings. These techniques are difficult and the authors are pioneers in this field. Overall this paper could make a solid contribution to the field, opening this exciting area up for more research.

Intro -  
No issues.

Guidelines (general) -

The formatting of the guidelines is not intuitive (the list is hierarchical but visually appears "flat") - will this be resolved in the final draft?

*The final formatting of the document will be completed by JoVE. For the purposes of the final submitted drafts, this format is in line with their requirements.*

Why does this section start with a statement about the IRB?

*We have removed the statement about the IRB from the final manuscript.*

Why is under the heading "Protocol" (rather than, e.g. guidelines?) The three sections ( camera selection, data collection, and data analysis) provide guidelines for many studies and are thus not a protocol per se. The statements here should be integrated with the section on "representative results". Additionally, the section on representative results should be broken down into methods and then results.

*Thank you for these recommendations. We considerably rearranged the Protocol. We removed the Representative Analysis section and replaced it with Representative Methods. Statements that were more indicative of a discussion of the Protocol were moved to the Discussion.*

A number of the guidelines simply state a recommendation without enough description of how to actually follow the recommendation. For example, 1.2 (examples of what kinds of issues might come up, or more generally, some more general description of what camera features will be "easy enough" vs too hard) , 1.4 - add that they should be tested? Or are there some technical specifications to look for ?

*We reworded Section 1 to be more specific in recommendations for the head-camera. We did our best to avoid "judgement-call" language that was not explained more thoroughly. We also added a section to remind the experimenters to test the cameras before the experiment begins.*

Also: 1.1.1: Could the authors provide some specific examples /considerations for how cameras can be secured. Describing the way the four cameras in figure 2 were attached (and lessons learned) would be plenty

*We have revised Figure 2 to more clearly show the ways in which head cameras can be secured. The top row of Figure 2 demonstrates actual use of the cameras while the bottom row is a close-up shot of the camera on a mannequin.*

2.4.5. and more specifically - 2.4.5.2: It's not clear what the authors are referring to re: building redundancy into the cap or how this could be achieved.

*We have removed the vague language from this section and replaced it with more direct language.*

Comments on specific sections:

Section 1 (camera selection). I read through the section on head cameras surprised that there was no list of recommended cameras. I think point 1.8 should be brought up at the beginning of this section or there should be a mention of it at the beginning, e.g.: There are numerous small, lightweight and portable cameras readily available for purchase, and we provide a list of recommendations below.

*Per JoVE standards, all commercial language has been removed from the manuscript. Specific product models have been included in the Table of Materials.*

Under point 1, figure 2 doesn't seem super useful as is - cameras are all very small and its hardly possible to make out the differences. Could the images be cropped and expanded so that more details on the cameras and head-mounts are made visible?

*Thank you for this suggestion. We have revised Figure 2 to address your concerns. It now more clearly shows the way the head-cameras have been attached.*

1.8. Can the authors provide a description of the shared features of the listed cameras and what differentiates them from cameras not being recommended to use? E.g. is it just their size? Have these been tested for durability under heavy use? are they easy for parents to use? etc  
Battery requirements are not mentioned. Any considerations here?

*Thank you for suggesting to include battery requirements. We have added a section (1.7) to speak directly to this issue. Unfortunately, a thorough comparison of different camera models is outside the scope of this manuscript.*

Section 2 (data collection)

2.2 refers to three experimenters but 2.3.1.1. refers to a parent. Are there four adults putting the cap on? If no, then 2.2 should be rephrased to "adults"

*We revised Section 2.1-2.2 and their subsections to make this process clearer. The parent's role in this process is mainly to maintain their child's temperament and distract the child. The majority of the work should be done by one or two experimenters. The third experimenter is monitoring the image of the head camera.*

The angle of the camera is brought up multiple times 2.3.3.1 and then 2.4.2 - can the authors make a single heading in which to describe motivations/ goals and tradeoffs about angles (viewing hands, down vs. up, etc)

*As the angle of the camera is more a discussion point of the method, we moved these sections to the Discussion section.*

Section 3 (data analysis)

This section is a bit weak compared to the others. E.g. the authors mention computer vision techniques in the intro but not here. Can there be a few points about this added here, including examples of what has already been done in the literature and what is likely/unlikely to be possible given the nature of first person video?

*We have removed the Data Analysis section from this manuscript and replaced it with a Representative Method section. This new section gives a better idea of the process by which to collect head camera data in a laboratory setting.*

Add 3.1.1 - move the descriptions of downsampling and considerations for downsampling (lines 266-269) here

*We have removed this section from the Revised Draft.*

Re: 3.2.1. - is there software available to do the bounding boxes? Could this be provided to the readers?

*There are multiple bounding box software solutions, some of which can be coded in programming languages such as Matlab. Per JoVE's policies we do not mention commercial products in the Revised Manuscript. We do not provide a software to the readers.*

3.3. "more detailed analyses" should be explicitly linked to the section which follows on representative results - otherwise the point does not make a useful recommendation

*We have removed the Data Analysis section from this manuscript and replaced it with a Representative Method section.*

Representative analyses:

Figure 4a - for the histograms how are they using a t-test? On the means? This statistic isn't clear

*We revised this section to be clearer. Lines 214-217: "An independent t-test comparing the number of scenes with a given number of objects between the child and the parent revealed this child had a greater number of scenes with fewer objects in view compared to their parent"*

Discussion:

The three Sentences about gaze vs. head angle (323-326) seem out of place in the discussion - it's already been made above, and the next sentence (Head cameras allow us to investigate the aggregate of the scenes available to a child) is clear enough.

*Thank you for the suggestion. We removed the sentences in question.*

Minor points

P3 line 82 - the second three here seems to be a typo: (only 13% of frames 3)

*Thank you for catching this typo. It has been fixed.*

P3 line 84: Run on sentence. Also missing a period. Thus, if the question of interest is the visual content available to infants (not the dynamics of momentary attention) head-mounted cameras do a good job capturing the visual environment most of the time, providing a window into their everyday experiences

*Thank you for catching this run-on. We have edited the lines appropriately.*

2.1.2 should have a citation.

*This section is no longer in the Manuscript.*