Supplemental Experimental Procedures

Participants. The final sample consisted of 36 (19 male infants) parent-infant dyads with the infants ranging in age from 11 to 13 months (mean = 12.52, SD= 1.15); 10 additional dyads began the study but the infants refused to wear the measuring equipment throughout the entire procedure. The sample of infants was broadly representative of Monroe County, Indiana: 84 % European American, 5% African American, 5% Asian American, 2% Latino, 4% Other) consisting of predominantly working- and middle-class families.

Stimuli. There were 6 unique novel "toys" constructed in the laboratory and pilot-tested to be interesting and engaging to infants. Each novel toy was a complex object made from multiple and often moveable parts and were of similar size, on average, 288 cm^3. These were organized into two sets of three so that each object in the set had a unique uniform color.

Experimental setup. As shown in Figure 1, parents and infants sat across from each other at a small table (61cm × 91cm × 64cm). Parents sat on the floor such that their eyes and heads were at approximately the same distance from the tabletop as those of the infants, a posture that parents reported to be natural and comfortable. Both participants wore head-mounted eye trackers (positive science, LLC http://www.positivescience.com/). The positive science eye-tracker was designed for use with infants and was designed to be attached to head so as to be stable on the head (even in self-locomoting infants and infants, see [S2, S3]. The tracking system has been widely and successfully used in both infant and adult research [S1, S4-S8]. Both parent and infant eye-tracking systems include an infrared camera – mounted on the head and pointed to the right eye of the participant – that records eye images, and a scene camera that captures the events from the participant's perspective. The scene camera's visual field is 108 degrees, providing a broad view to approximate the full visual field [S9]. Each eye tracking system recorded both the egocentric-view video and eye-in-head position (x and y) in the captured scene at a sampling rate of 30 Hz.

In addition to head-mounted eye tracking, three additional video cameras were used to record the interaction from three different viewpoints that was independent of participants' movements: a bird's-eye camera mounted on the top of the interaction tabletop, a camera pointing to the infant, and a camera pointing to the parent. In total, 7 video streams were recorded in a geovision video capture card (Model 1480B) which automatically synchronized multiple video streams. Synchronization was verified by using a standard camera flash procedure. An experimenter triggered a camera flash both at the beginning and end of each interaction which was captured in one frame in all cameras. Before data processing, coders compared the frame numbers across all the video streams to confirm synchronization. The flash-marked frames were used for resynchronization, if necessary.

Placing the head gear and eye tracker calibration. Prior to entering the testing room, in the waiting area, the first experimenter desensitized the infant to touches to the head and hair by lightly touching the hair several times when the attention and interest of the infant was directed to a toy. Both the parent and the infant entered the experimental room, and a second experimenter and the parent engaged the infant with an enticing toy with buttons to push that make animals pop up. The infant's head gear was placed while the infant was engaged with the toy. This was done in one movement and care was taken by the experimenter to ensure that the infant remained engaged with the toy and that the infant's hands didn't go to the head gear. The first experimenter then adjusted the scene camera to ensure that the button being pushed by the infant was in the center of the scene camera. We have used this procedure in multiple head-camera and head-mounted eye-tracking experiments [S10-S14] with an overall 70% success rate.

Instructions and procedure. Parents were told that the goal of the experiment was to study how parents and infants interacted with objects during play and therefore they were asked to engage their infants with the toys and to do so as naturally as possible. Each of the two sets of toys was played with twice for 1.5 min, resulting in 6 minutes of play data from each dyad. Order of sets (ABAB or BABA) was counterbalanced across dyads.

Data processing. During post-processing and before coding, the quality of the eye tracking video (with eye images superimposed) for each infant and parent was checked to ensure the quality of calibration at the end as well as the beginning of the session. Re-calibration would be conducted if necessary. ROI coding was done by human coders. These coders were highly trained and code these variables for many different experiments and projects. They were naïve to the specific hypotheses or experimental questions of this study. The four regions-of-interest (ROIs) were defined in the head-camera videos: the three toy objects and the partner's face. To determine gaze that fell within these ROIs, coders watched the first-person view video with a cross-hair indicating gaze direction, frame-by-frame, and annotated when the cross-hair fell on a pixel identified as any part of the four ROIs. Because the experimental room is white and all participants wore white clothing that covers all but faces and hands, and because the three toys in play were three different primary colors that are different from skin tones, it was straightforward for coders to identify the three object and face regions in view. In addition, using the eye tracking software, we rendered eye images via picture-in-picture superimposed at the upper-right corner of a scene frame (see Figure 1), which allowed coders to constantly use them as a reference to verify reliability of cross-hair indicating gaze direction in view. If coders detected from an eye image that the eye tracking software failed to detect the pupil correctly due to image quality or eye blinks, coders disregarded that frame for any ROIs because the cross-hair was incorrect. Thus, we measured infants' and parents' visual attention in terms of gaze directed at any of the three objects or the partner's face. In implementation, coders went through each video four times wherein one of the four ROIs was focused in each round and they needed to make a yes/no decision (whether the cross-hair was on the ROI) based on the overlap of the cross-hair with the ROI. In previous studies with the same setup, we've also developed an image processing algorithm to automatically separate the three objects in play from each other and the background (see [S13 S14] for details). We've applied automatic object detection to this dataset and calculated object sizes in view. We found that on average, each object took 3.25% of the scene image in the infant's view and 1.82% in the parent's view. Thus, relatively large objects in view with the clean background made ROI coding highly reliable when compared with coding ROIs from more naturalistic and complex visual scenes. From gaze ROI coding, each dyad provided two gaze data streams containing the four ROIs as shown in Figures 1 and 2. A second coder independently coded a randomly selected 10% of the frames with the inter-coder reliability ranged from 82% to 95% (Cohen's kappa = 0.81).

Supplemental References

S1. Franchak, J.M., Kretch, K.S., Soska, K.C., and Adolph, K.E. (2011). Head-Mounted Eye Tracking: A New Method to Describe Infant Looking. Child development *82*, 1738-1750.

S2. Franchak, J.M., Kretch, K.S., Soska, K.C., Babcock, J.S., and Adolph, K.E. (2010). Head-mounted eye-tracking of infants' natural interactions: a new method. In Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications. (ACM), pp. 21-27.

S3. Franchak, J.M., and Adolph, K.E. (2010). Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults. Vision research *50*, 2766-2774.

S4. Evans, K.M., Jacobs, R.A., Tarduno, J.A., and Pelz, J.B. (2012). Collecting and analyzing eye tracking data in outdoor environments. Journal of Eye Movement Research *5*, 6.

S5. Baschnagel, J.S. (2013). Using mobile eye-tracking to assess attention to smoking cues in a naturalized environment. Addictive behaviors *38*, 2837-2840.

S6. Kretch, K.S., Franchak, J.M., and Adolph, K.E. (2014). Crawling and walking infants see the world differently. Child development *85*, 1503-1518.

S7. Macdonald, R.G., and Tatler, B.W. (2013). Do as eye say: Gaze cueing and language in a real-world social interaction. Journal of vision *13*, 6.

S8. Maldarelli, J.E., Kahrs, B.A., Hunt, S.C., and Lockman, J.J. (2015). Development of early handwriting: Visual-motor control during letter copying. Developmental Psychology *51*, 879-888.

S9. Smith, L., Yu, C., Yoshida, H., and Fausey, C.M. (2014). Contributions of Head-Mounted Cameras to Studying the Visual Environments of Infants and Young Children. Journal of Cognition and Development.

S10. Yu, C., and Smith, L.B. (2013). Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. PLOS ONE *8*.

S11. Smith, L.B., Yu, C., and Pereira, A.F. (2011). Not your mother's view: The dynamics of toddler visual experience. Developmental Science *14*, 9-17.

S12. Pereira, A.F., Smith, L.B., and Yu, C. (2014). A Bottom-up View of Toddler Word Learning. Psychological Bulletin & Review, 1-8.

S13. Yu, C., and Smith, L.B. (2012). Embodied Attention and Word Learning by Toddlers. Cognition *125*, 244-262.

S14. Yu, C., Smith, L.B., Shen, H., Pereira, A., and Smith, T. (2009). Active Information Selection: Visual Attention Through the Hands. IEEE Transactions on Autonomous Mental Development *2*, 141–151.