Objects in the Center: How the Infant's Body Constrains Infant Scenes

Sven Bambach[1,3], Linda B. Smith[1,2], David J. Crandall[1,3], Chen Yu[1,2]
[1]Cognitive Science Program, [2]Department of Psychological and Brain Sciences, [3]School of Informatics and Computing
Indiana University
Bloomington, IN 47405, USA
{sbambach, smith4, djcran, chenyu}@indiana.edu

***Abstract*—During early visual development, the infant's body and actions both create and constrain the experiences on which the visual system grows. Evidence on early motor development suggests a bias for acting on objects with the eyes, head, trunk, hands, and object aligned at midline. Because these sensory-motor bodies structure visual input, they may also play a role in the development of visual attention: attended objects are at the center of a head- and body-centered scene. In this study, we designed a table-top object exploration task, in which infants and parents were presented with novel objects for joint play and examination. Using a head-mounted eye-tracking system, we measured each infant's point of gaze relative to the head when attending objects. With an additional overhead camera, we recorded the position of each object relative to the infant's body. We show that even during free toy play, infants tend to bring attended objects towards their body's midline and attend objects with head and eyes aligned, systematically creating images with the attended object at center.**

I. Introduction

Berkeley [1] proposed that the spatial aspects of vision would be explained by the body's morphology and its propensity for action. Berkeley's idea may be best understood as it is in developmental robotics, in terms of how the baby's body and actions constrain and create the experiences on which the visual system grows. The dependence of sensory experience on the body's morphology has been a core idea in robotics [2] and developmental robotics especially [3], [4].

A. *The Body's Midline and Motor Development*

The human body is a complex system with many degrees of freedom. As infants discover new motor skills within this complex system, stability and coordination pose profound problems [5]. Early in development, infants solve this problem by limiting degrees of freedom for movement to action at or near the body's midline. For example, when first beginning to reach, infants reach with both hands and do not cross the midline, and do so more accurately and smoothly for objects presented at midline to the body versus objects presented laterally [6], [7]. Infants as old as 12 and 18 months who have multiple targets in view show strong biases to reach to the object that is nearest to midline [8]. Holding objects at midline stabilizes and aligns the trunk and head so that infants who are just beginning to sit, sit longer when holding an object with two hands at midline [9] and infants who are just beginning to walk, walk more stably and take more steps when holding an object at midline [10]. All these phenomena are generally understood in terms of constraining the degrees of freedom – by aligning eyes, head, hands and trunk – within a complex motor system to create stabilities and controlled action.

These same motor processes may also play a role in the development of sustained visual attention. When infants, toddlers, and children look at objects in the world, they typically do so with aligned head and eyes (see [11] for a review). Although heads and eyes can be directed in different directions, for toddlers they are usually aligned in the context of freely moving bodies acting in a 3-dimensional world [12], [13]. Other evidence suggests a possible role for aligned hands in sustained attention as well, with episodes of long duration of looking at objects being characterized by an infant-held object at midline with both eyes and head directed to the object [14].

In brief, the evidence on early motor development suggests a bias for viewing and acting on objects with the eyes, head, trunk, hands, and object aligned at midline; this bias decreases degrees of freedom and supports the dynamic stability of the system and controlled action. This bias – emergent from the body's morphology – must also structure visual scenes, placing objects at the center of a head- and body-centered scene, and if so, this regularity should be pervasive. This is the hypothesis that we test in the present study. In the General Discussion, we consider how this regularity in the infant's visual experience may be foundational in the development of the human visual system.

B. *Rationale for the Present Approach*

To test this hypothesis, we used a table-top object exploration task in which infant and a parent were presented with three novel objects for play and examination. The parent sat across the table from the infant and was only told to help their child engage with the objects. The objects, in principle, could be looked at and reached for at any location on the table. We chose this object play context with a mature partner because it is the natural context in which infants explore objects and shift action and attention among those objects, thus providing many opportunities for coordinating (or not coordinating) eyes, heads, and hands, and creating scenes in which the attended object is at the body's midline. The participants were recruited from two age groups that spanned a period of from lesser (9 to 12 month olds) to greater (15 to 19 month olds) object

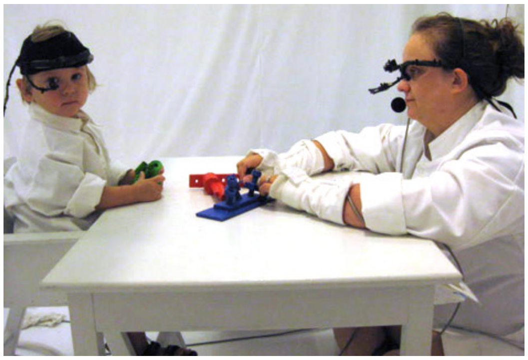

Fig. 1. *Experimental setup.* Child and parent play with a set of 3 toys at a table. Both wear head-mounted eye tracking systems (although we do not consider the data from the parent in this work). An additional camera (not seen in the figure) is mounted over the table and records the bird's-eye view of the interaction.

manipulation skills. Parents participated in the play session to sustain infant engagement with the objects. Their participation also offers the opportunity to compare midline scene biases in parent-held versus infant-held objects. Infants wore a head-mounted eye-tracker with a head-centered scene camera. The principal dependent measures were the spatial distributions of object-directed gaze in the head-camera image and the spatial distributions of those objects on the tabletop.

II. Method

A. Participants

The participants were 17 infants (7 male) between 9 and 12 months of age (mean age 10.9 mo, SD 1.5 mo) and 16 infants (11 male) between 15 and 19 months of age (mean age 17.3 mo, SD 1.6 mo).

B. Experimental Setup

Infant and parent sat across from each other at a white table with dimensions $43 \times 91 \times 63$ cm. The infant sat on a chair that was centered at the table's midpoint. The chair, as shown in Figure 1, had sides that supported the trunk and constrained but not did not completely eliminate trunk movement, such that objects at midline to the body should be at the center of the table. Heads and eyes were free to orient in any direction, as evident in the figure. The parent sat on the floor directly across from the infant such that the distances between the center of the table to the parents and childs eyes were each each around 45 cm. In order to create a visually clean setup and to allow for easy coding of objects, everything in the room was white, and parents and infants wore white jackets (see Figure 1).

C. Stimuli

Each participant played with six unique toy objects selected from a pool of 15 novel objects (on average, about $9.5 \times 6.5 \times 5.0$ cm), created in the laboratory. Within a set of three objects, each individual object was a single uniform but different colorr – blue, red or green.

D. Head-Mounted Eye Tracking

The infant wore a head-mounted eye tracking system (*Positive Science LLC*), which included an infrared camera, mounted on the head and pointed to the right eye of the infant, that recorded eye images, and a scene camera that captured the infant's field of view. The scene camera was wide-angled (108° diag. FOV), providing a broad view to approximate the full visual field [15]. Both cameras recorded data at 30 Hz.

E. Overhead Camera

An additional camera was mounted above the table center, perpendicular to the table's surface and including the whole table top in view, to provide a clean overhead view of the interaction. The overhead camera also recorded data at 30 Hz and was synchronized with the head-mounted cameras.

F. Procedure

A team of two experimenters placed the eye-tracking system on the toddler and performed a calibration procedure (see [15] for details). Parents were instructed to play with their child naturally with three toys at a time for a total of four trials, each lasting around 1.5 minutes. If necessary, the eye-tracking system was re-calibrated between trials.

G. Data Processing and Coding

1) Head-mounted eye tracking data: Based on the eye tracking data, we annotated moments in which children overtly attended one of the three toy objects. Coding was done by trained human coders naïve to the specific hypotheses. Frame by frame, they coded when the cross-hair indicating gaze fell on any part of the three objects (see Figure 3 (a) for an example frame, and [16] for more details).

Due to the geometry of the eye tracking system, the recorded gaze, although accurate, may have a constant offset with respect to the scene camera center. Thus, after the coding process, gaze data was shifted such that the mean gaze position of each trial was at the center of the scene camera. On average, gaze was shifted $30px$ (6% of vertical camera resolution) up and $35px$ (5% of horizontal resolution) to the left to adjust for the imperfect camera alignment.

2) Overhead camera data: We extracted the position and shape of each of the three objects in view of the overhead camera, which directly corresponded to their position on the table (see Figure 2 (a) for an example frame). Objects were detected automatically based on optimized color thresholds (see [16] for details).

3) Held object observation data: Using both the first-person video and the overhead video for reference, human coders determined frame by frame if child or parent were holding one of the three objects. Holding was defined as any hand contact with an object.

III. Results

By hypothesis, toddlers bring attended objects to the midline of their body and attend to them with eyes and head aligned. Two frames of reference are relevant to this hypothesis: (1) the

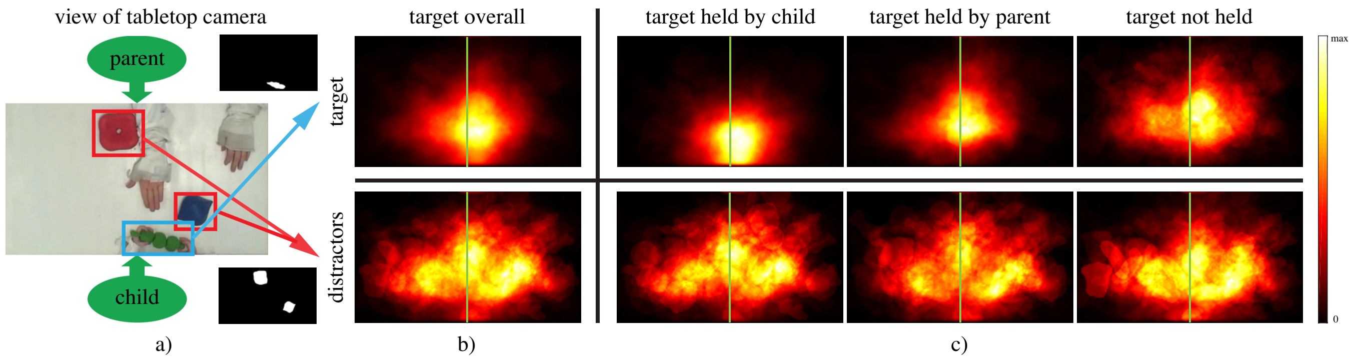

Fig. 2. *Locations of toy objects on the table.* a) Example of a frame captured by the overhead camera, with the child sitting south and the parent sitting north of the table. Object shapes of the target (blue) and the distractors (red) were extracted automatically. b) Heatmap visualization of targets (top row) and distractors (bottom row) across the table. Heatmaps were created by accumulating object shapes across all subjects. c) The same visualization with data split based upon who is holding the target object.

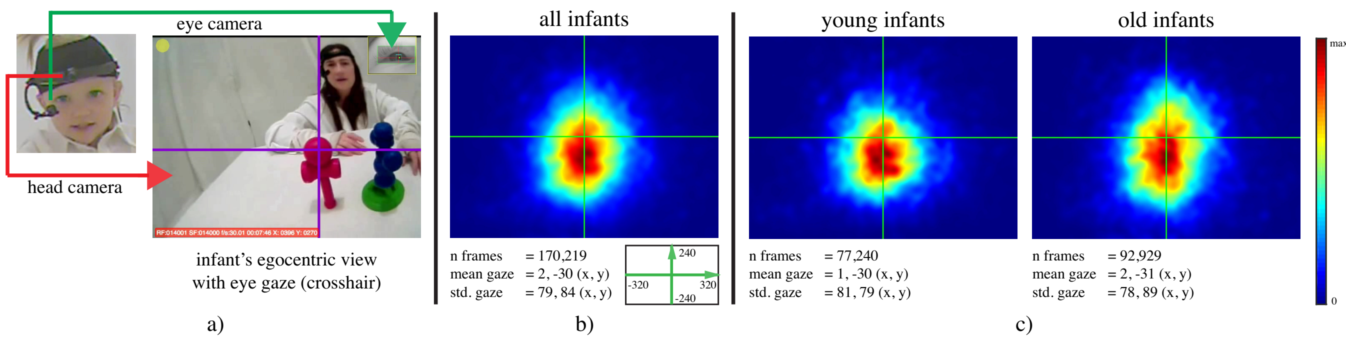

Fig. 3. *Toddler eye gaze within the head-centered reference frame.* a) Example of a frame captured by the toddler's head camera. Overlaid is a cross-hair indicating the xy-position of the toddler's gaze. b) Heatmap visualization of toddler gaze when targeting a toy object, accumulated across all subjects. c) The same visualization with gaze data split into age groups. For visualization purposes, each map was slightly smoothed with a Gaussian kernel ($\sigma = 8px$).

location of objects (attended and not-attended) in the physical world and in relation to the child's body, and (2) the location of the child's gaze within the head camera image. If infants actively orient objects at their body's midline when attending to them, the physical location of attended and unattended objects should differ. Whereas unattended objects (potential but not current targets of attention) should be scattered all over the table, attended objects (objects to which the child's gaze is directed at that moment) should be centered on the table, at the midline of the child's body.

A. *Object Locations on the Table*

Figure 2 (b) shows heat maps of toy locations on the table, created by accumulating the extracted object shapes from the tabletop camera across all frames and all subjects (i.e. both age groups). The top row shows only the objects that were directly attended by the child (target) while the bottom row shows the two unattended objects (distractors) for the same moments in time. As is apparent, the unattended toys were located all over the table (to the right, at midline, and to left of the child) but the attended toys were physically located in the center of the table. On a per subject level, we entered the mean absolute lateral distance from object centroid to table midline into a 2 (age: young/old) $\times$ 2 (object: target/distractor) analysis of variance and found a main effect of object ($F(1, 60) = 51.28, p < 0.001$), where target objects were closer to the table midline than distractors. No other effects approached significance.

We further divided the data depending on whether the target object was held by the child, the parent, or neither (top row of Figure 2c). A per subject 2 (age: young/old) $\times$ 3 (holding: child/parent/neither) analysis of variance on the absolute lateral distance from the target centroid to the table midline found a main effect of holding ($F(2, 88) = 8.32, p < 0.001$), where targets that were not held were further away from the table center than held targets, with no difference between child holding and parent holding. However, targets that were not held were still closer to the center than their corresponding distractors (t-test; $t(29) = 3.93, p < 0.001$).

In summary, for both younger and older infants, attended objects were located close to the toddler's body midline when held by the child, held by the parent, and to a lesser degree when not held by anyone.

B. *Overall Eye-Head Alignment*

The midline bias also predicts that infant gaze will be close to the center of the head-camera image; that is, that objects will be attended with heads and eyes aligned. Figure 3 (b) shows the distribution of gaze within the head-camera image for moments when gaze was directed at a toy object, accumulated across all subjects. Indeed, most of the eye gaze is very close around the head-camera (i.e. field of view) center, with little horizontal or vertical variance ($\sigma_x = 79px$

or 25% of horizontal FOV, $\sigma_y = 84px$ or 35% of vertical FOV). We further investigated, on a per subject level, whether the mean of the gaze distribution significantly differed from the center of the field of view, and found no difference in x-position but a significant difference in y-position (t-test; $t(32) = -10.36, p < 0.001$), where the eye gaze mean was slightly below the FOV center.

Figure 3 (c) splits the data between younger and older infants. We compared both groups in terms of per subject horizontal/vertical variance and horizontal/vertical mean, and found that older infants tend to have a larger vertical variance (t-test; $t(31) = -2.55, p = 0.016$), with no other significant differences.

In summary, infant gaze tends to be very close to the center of the field of view when attending a target object, with a mean slightly below the center. This small shift is likely due toy objects being located on the table top below the toddler. Older infants are more likely to have a greater vertical (but not horizontal) variance in gaze, meaning that they still attend objects at body (or head) midline.

C. Eye-Head Alignment based on Holding Objects and Object Location

We further investigate how the toddler's eye-head alignment is affected by toddler or parent holding the target object, as well as by the distance between the target and table center (and by proxy the toddler's body midline). Figure 4 divides the toddler's eye gaze (when targeting a toy) based on who is holding the target. Additionally, we divided the table into three equally sized regions (left/center/right) and further split the gaze data based on the table region where the centroid of the target was located.

On a per subject basis, we performed a 3 (holding: child/parent/neither) $\times$ 3 (target position: left/center/right) analysis of variance on the vertical/horizontal gaze center as well as the vertical/horizontal gaze variance. We found a significant effect of holding, where the infants' gaze tends to be lower when infants are holding the toy ($F(2, 266) = 43.74, p < 0.001$). We further found an effect of target position, where the gaze center tends to be more to the left/right if the target is on the left/right side of the table ($F(2, 266) = 194.86, p < 0.001$). However, this horizontal offset is rather small, such that toddlers likely also aligned their heads towards the off-center targets.

D. Bringing Targets towards the Body Midline

Figure 4 also shows that moments in which the target object is not centered on the table are rare (26%) and that the target is usually held by either parent or child (74%). We thus investigated whether off-center moments happen randomly or if toddlers (and maybe also parents) actively bring off-center targets towards the toddler's body midline. Figure 5 summarizes our results by plotting the proportion of instances in which the target object was centered on the table as a function of time. We defined being "centered" based on the same three-region split of the table as in Figure 4. The blue

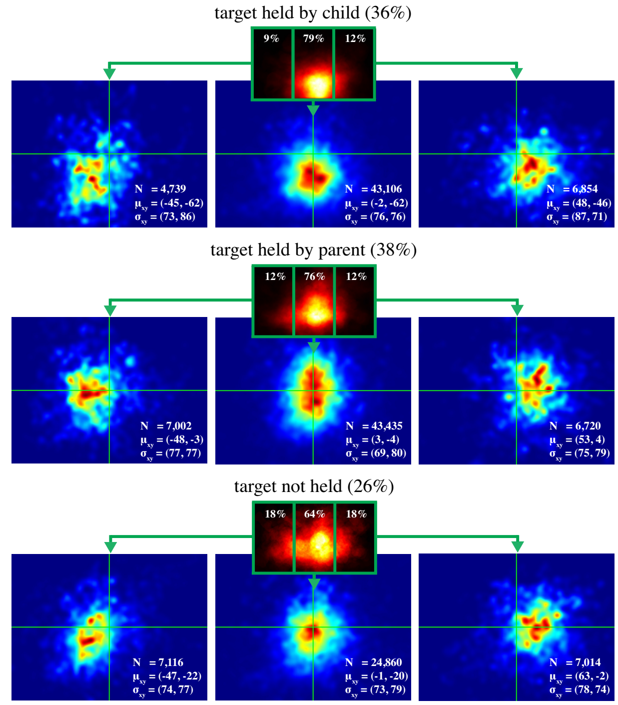

Fig. 4. *Toddler eye gaze across different conditions.* Heatmap visualizations (across both age groups), split based upon who is holding the target object (rows) and where on the table the target object is located (columns).

solid line shows targets held by the toddler which were off-center at the toddler's attentional onset. After 3 seconds, around 65% of those targets were brought to the table center by the toddler. Interestingly, parents (red solid line) were also likely to center target toys that were initially off-center (around 52% after 3 seconds). Conversely, targets that were already centered at the toddler's attentional onset (dotted lines) tend to stay centered around 90% of the time.

In summary, moments where attentional targets are away from the toddler's body midline tend to be rare and happen predominantly at the beginning of the attentional onset. Toddlers (and to a lesser extend parents) are likely to subsequently bring the target object towards the toddler's body midline.

IV. General Discussion

There are three central findings. First, both the younger and older infants were strongly biased to view objects with aligned heads and eyes, such that the attended object was at the center of the head-centered scene. Second, this bias was linked to manual actions, by both infant and parent, that located attended objects at the midline of the infant's body. Third, infants dynamically coordinate eyes, head, and body to bring the attended object to midline such that objects rapidly become centered after initial gaze is directed to the object.

A. Sustained Attention

Adults have great flexibility in both their body and the spatial orientation of visual attention. In laboratory visual attention tasks, adults can readily attend to one specific location (and rapidly detect objects at that location) without moving the eyes and while eye gaze is fixated elsewhere

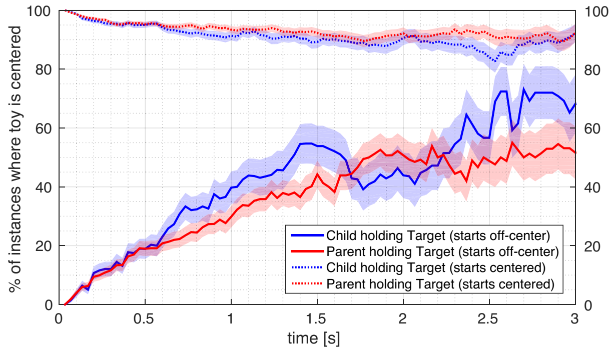

Fig. 5. *Proportion of instances where the toddler's visual target was centered on the table.* During the first 3 seconds after the attentional onset of the toddler, targets that were initially off-center (solid lines) are likely to be brought towards the table center. Conversely, targets that were already centered tend to stay centered (dotted lines). Shaded areas depict standard error. We plot the first 3 seconds after the attentional onset because uninterrupted attention to the object rarely lasted longer (mean duration 1.6s, SD 1.8s).

(e.g. [17], [18]). Thus spatial attention in adults is internal and does not require moving the sensors toward the attended object. However, adult attention is also tied to the body. Adults typically orient eye gaze to the attended location. Moreover, eye movements [19], [20], head movements [21], and hand movements [22]–[24] bias visual attention in the direction of the movement. Visual attention thus appears coupled to mechanisms of directional action, perhaps because, more often than not, we direct attention in preparation for action. Moreover, other adult studies show that aligned heads and eyes are better for visual processing by multiple measures than misaligned heads and eyes (e.g., [25]–[27]). This alignment may help stabilize visual attention on a visual target. This has been suggested to be a factor in infant attention [28], [29]. For example, in a study of 1- to 4-year-olds, Ruff & Capozzoli [29] noted that concentrated and sustained attention during play was associated with minimal extraneous bodily activity and a posture that centered the object and brought it closer to the face and eyes.

Recent findings from head camera studies also suggest that effective toddler attention is associated with a still head, holding an object, and the centering of the object with respect to the head and the head-centered visual scene [16], [30]. In these studies, parents named objects as infants and parents played with them, and then, after play infants were tested to determine which object names were learned. Head-mounted camera images were selected that surrounded parent naming moments and these images were categorized in terms of whether the subsequent testing showed that the object had been learned or not. The results showed that the stability (over 5 to 7 seconds) of the child view of the named object at midline predicted object name learning. In brief, these ideas support the proposal that the sustained alignment of the eyes, head, body, and attended object at midline support infant visual attention and learning. Thus, the coordinated head-eye-object alignment seen in the present study may be the bodily underpinning of effective visual attention and its development.

B. *The Scene Statistics that build the Attentional System*

The human visual system appears to build itself at every level by incorporating the statistical regularities in visual experiences into the visual circuitry. Infants and toddlers actively create scenes in which attended and handled objects are at midline. These early pervasive regularities may explain basic properties of human adult vision. For example, adults are biased to fixate and begin viewing a scene (presented on a screen) at the center of the image and they process scene-center information more rapidly (e.g., [31]–[34]). This center-of-scene bias also extends to adult gaze in natural everyday contexts in which adults freely move while performing goal-directed actions [35]. Adults also show a near-hand bias, attending to objects near their own hands and near the hands of others. In a variety of detection and discrimination studies, adults show faster and more accurate performance when the visual target is near their own hands [36], and in several recent studies also near the hand of another in joint action tasks [37]. The center-scene and near-hand biases are so strong in adults that adult gaze within a head centered scene can be predicted with minimal error from just the scene's properties of proximity of objects to hands and center [35]. It seems highly likely that the coordinated and centered nature of heads, hands, and eyes in young infants creates the statistical regularities from which the scene-center and near-hand biases in adults emerge.

The eyes-head-hands biases may also play a role in other early developmental achievements. For example, the extant evidence shows that very young infants follow another's gaze in highly restricted viewing contexts (e.g., [38], [39]), but also shows that the spatial resolution of gaze following is often not sufficient for navigating real-time social interactions in more spatially complex social settings (e.g., [15], [40]–[42]). Critically, the spatial complexity of social interactions explodes as infants become more physically active and transition from social interactions dominated by face-to-face play to social interactions that are dominated by shared engagement with objects [43]. In one study using simultaneous head-mounted eye trackers worn by toddlers and parents, Yu and Smith [15] found that one-year-old infants coordinated their own gaze with that of the parent, not by following parent eye-gaze, but by fixating on and following parent hand movements to objects (to which parent eye gaze was also dynamically coordinated). Computational modelers have further proposed that hand-following – with its superior spatial precision – may tune and refine gaze following (e.g., [15], [44], [45]), which in principle could enable gaze skills to increasingly meet the challenge of complex interactions with objects.

C. *Conclusion*

In conclusion, the regularities of infant ego-centric scenes are likely to have an outsized influence on the development of the human visual system. The properties of those scenes depend on infant sensory-motor systems. Here we show that the system is strongly biased to actively create scenes in which the attended object is at the center of that scene, and during which eyes, head, and the location of the object relative to the body are aligned and centered.

Acknowledgments

We thank S. Foster, M. Elston, C. Wozniak, and S. Elmlinger for collecting the data. This research was supported by NSF IIS-1253549, NIH R01 HD074601, and NIH R21 EY017843. SB was supported by a Paul Purdom Fellowship.

References

[1] G. Berkeley, *An essay towards a new theory of vision*. A. Rhames, 1709.

[2] R. Pfeifer and J. Bongard, *How the body shapes the way we think: a new view of intelligence*. MIT press, 2006.

[3] A. F. Morse, V. L. Benitez, T. Belpaeme, A. Cangelosi, and L. B. Smith, "Posture affects how robots and infants map words to objects," *PloS one*, vol. 10, no. 3, p. e0116012, 2015.

[4] M. Lungarella, G. Metta, R. Pfeifer, and G. Sandini, "Developmental robotics: a survey," *Connection Science*, vol. 15, no. 4, pp. 151–190, 2003.

[5] E. Thelen, J. S. Kelso, and A. Fogel, "Self-organizing systems and infant motor development," *Developmental Review*, vol. 7, no. 1, pp. 39–65, 1987.

[6] J. S. Bruner and B. Koslowski, "Visually preadapted constituents of manipulatory action," *Perception*, vol. 1, no. 1, pp. 3–14, 1972.

[7] J. Fagard, E. Spelke, and C. von Hofsten, "Reaching and grasping a moving object in 6-, 8-, and 10-month-old infants: Laterality and performance," *Infant Behavior and Development*, vol. 32, no. 2, pp. 137–146, 2009.

[8] A. Sheya and L. B. Smith, "Changing priority maps in 12-to 18-month-olds: An emerging role for object properties," *Psychonomic bulletin & review*, vol. 17, no. 1, pp. 22–28, 2010.

[9] L. J. Claxton, J. M. Strasser, E. J. Leung, J. H. Ryu, and K. M. O'Brien, "Sitting infants alter the magnitude and structure of postural sway when performing a manual goal-directed task," *Developmental psychobiology*, vol. 56, no. 6, pp. 1416–1422, 2014.

[10] L. J. Claxton, D. K. Melzer, J. H. Ryu, and J. M. Haddad, "The control of posture in newly standing infants is task dependent," *Journal of Experimental Child Psychology*, vol. 113, no. 1, pp. 159–165, 2012.

[11] L. B. Smith, C. Yu, H. Yoshida, and C. M. Fausey, "Contributions of head-mounted cameras to studying the visual environments of infants and young children," *Journal of Cognition and Development*, vol. 16, no. 3, pp. 407–419, 2015.

[12] H. Yoshida and L. B. Smith, "What's in view for toddlers? using a head camera to study visual experience," *Infancy*, vol. 13, no. 3, pp. 229–248, 2008.

[13] S. Bambach, D. J. Crandall, and C. Yu, "Understanding embodied visual attention in child-parent interaction," in *IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL), 2013*. IEEE, 2013, pp. 1–6.

[14] H. A. Ruff and K. R. Lawson, "Development of sustained, focused attention in young children during free play." *Developmental psychology*, vol. 26, no. 1, p. 85, 1990.

[15] C. Yu and L. B. Smith, "Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination," *PloS one*, vol. 8, no. 11, p. e79659, 2013.

[16] ——, "Embodied attention and word learning by toddlers," *Cognition*, vol. 125, no. 2, pp. 244–262, 2012.

[17] J. R. Müller, M. G. Philiastides, and W. T. Newsome, "Microstimulation of the superior colliculus focuses attention without moving the eyes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 3, pp. 524–529, 2005.

[18] M. Shepherd, J. M. Findlay, and R. J. Hockey, "The relationship between eye movements and spatial attention," *The Quarterly Journal of Experimental Psychology*, vol. 38, no. 3, pp. 475–491, 1986.

[19] M.-H. Grosbras, A. R. Laird, and T. Paus, "Cortical regions involved in eye movements, shifts of attention, and gaze perception," *Human brain mapping*, vol. 25, no. 1, pp. 140–154, 2005.

[20] G. Rizzolatti, L. Riggio, I. Dascola, and C. Umiltá, "Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention," *Neuropsychologia*, vol. 25, no. 1, pp. 31–40, 1987.

[21] C. L. Colby and M. E. Goldberg, "Space and attention in parietal cortex," *Annual review of neuroscience*, vol. 22, no. 1, pp. 319–349, 1999.

[22] D. Hagler Jr, L. Riecke, and M. Sereno, "Pointing and saccades rely on common parietal and superior frontal visuospatial maps." *Neuroimage*, vol. 35, pp. 1562–1577, 2007.

[23] E. I. Knudsen, "Fundamental components of attention," *Annual Review of Neuroscience*, vol. 30, pp. 57–78, 2007.

[24] D. Thura, F. Hadj-Bouziane, M. Meunier, and D. Boussaoud, "Hand position modulates saccadic activity in the frontal eye field," *Behavioural brain research*, vol. 186, no. 1, pp. 148–153, 2008.

[25] W. Einhäuser, F. Schumann, S. Bardins, K. Bartl, G. Böning, E. Schneider, and P. König, "Human eye-head co-ordination in natural exploration," *Network: Computation in Neural Systems*, vol. 18, no. 3, pp. 267–297, 2007.

[26] L. Thaler and J. T. Todd, "The use of head/eye-centered, hand-centered and allocentric representations for visually guided hand movements and perceptual judgments," *Neuropsychologia*, vol. 47, no. 5, pp. 1227–1244, 2009.

[27] J. Jovancevic-Misic and M. Hayhoe, "Adaptive gaze control in natural environments," *The Journal of Neuroscience*, vol. 29, no. 19, pp. 6234–6238, 2009.

[28] K. N. Kannass and L. M. Oakes, "The development of attention and its relations to language in infancy and toddlerhood," *Journal of Cognition and Development*, vol. 9, no. 2, pp. 222–246, 2008.

[29] H. A. Ruff and M. C. Capozzoli, "Development of attention and distractibility in the first 4 years of life." *Developmental psychology*, vol. 39, no. 5, p. 877, 2003.

[30] A. F. Pereira, L. B. Smith, and C. Yu, "A bottom-up view of toddler word learning," *Psychonomic bulletin & review*, vol. 21, no. 1, pp. 178–185, 2014.

[31] G. T. Buswell, *How people look at pictures*. University of Chicago Press Chicago, 1935.

[32] S. Mannan, K. Ruddock, and D. Wooding, "Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-d images," *Spatial vision*, vol. 9, no. 3, pp. 363–386, 1995.

[33] D. Parkhurst, K. Law, and E. Niebur, "Modeling the role of salience in the allocation of overt visual attention," *Vision research*, vol. 42, no. 1, pp. 107–123, 2002.

[34] B. W. Tatler, R. J. Baddeley, and I. D. Gilchrist, "Visual correlates of fixation selection: effects of scale and time," *Vision research*, vol. 45, no. 5, pp. 643–659, 2005.

[35] Y. Li, A. Fathi, and J. Rehg, "Learning to predict gaze in egocentric video," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 3216–3223.

[36] C. C. Davoli, J. R. Brockmole, and A. Goujon, "A bias to detail: How hand position modulates visual learning and visual memory," *Memory & cognition*, vol. 40, no. 3, pp. 352–359, 2012.

[37] M. Costantini, G. Committeri, and C. Sinigaglia, "Ready both to your and to my hands: mapping the action space of others," *PloS one*, vol. 6, no. 4, p. e17923, 2011.

[38] T. Farroni, M. H. Johnson, M. Brockbank, and F. Simion, "Infants' use of gaze direction to cue attention: The importance of perceived motion," *Visual cognition*, vol. 7, no. 6, pp. 705–718, 2000.

[39] T. Farroni, S. Massaccesi, D. Pividori, and M. H. Johnson, "Gaze following in newborns," *Infancy*, vol. 5, no. 1, pp. 39–60, 2004.

[40] M. J. Doherty, J. R. Anderson, and L. Howieson, "The rapid development of explicit gaze judgment ability at 3years," *Journal of Experimental Child Psychology*, vol. 104, no. 3, pp. 296–312, 2009.

[41] J. M. Loomis, J. W. Kelly, M. Pusch, J. N. Bailenson, and A. C. Beall, "Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior," *Perception*, vol. 37, no. 9, pp. 1443–1457, 2008.

[42] M. D. Vida and D. Maurer, "The development of fine-grained sensitivity to eye contact after 6years of age," *Journal of Experimental Child Psychology*, vol. 112, no. 2, pp. 243–256, 2012.

[43] T. Striano and V. M. Reid, "Social cognition in the first year," *Trends in cognitive sciences*, vol. 10, no. 10, pp. 471–476, 2006.

[44] J. Triesch, C. Teuscher, G. O. Deák, and E. Carlson, "Gaze following: why (not) learn it?" *Developmental science*, vol. 9, no. 2, pp. 125–147, 2006.

[45] S. Ullman, D. Harari, and N. Dorfman, "From simple innate biases to complex visual concepts," *Proceedings of the National Academy of Sciences*, vol. 109, no. 44, pp. 18 215–18 220, 2012.