What's in view for toddlers? Using a head-camera to study visual experience

Hanaka Yoshida

University of Houston

Linda B. Smith

Indiana University

Abstract

The paper reports two experiments using a new method to study 18- to 24-month-olds' visual experiences as they interact with objects. Experiment 1 presents evidence on the coupling of head and eye-movements and thus the validity of the head-camera view of the infant's visual field in the geometry of the task context. Experiment 2 demonstrates the use of this method in the naturalistic context of toy play with a parent. The results point to the embodied nature of toddler's attentional strategies and to the importance of hands and hand actions in their visual experience of objects. The head camera thus appears to be a promising method that despite some limitations will yield new insights about the ecology and content of young children's experiences.

What is in view for toddlers: Using a head-camera to study visual experience *"cognition depends on the kinds of experiences that come from having a body with particular perceptual and motor capabilities"* (Infancy vol.1 (1) pp.5 Presidential Addess; Thelen, 2000)

Developmental psychologists have long been interested in the nature of the input, in the everyday experiences that characterize early childhood and the role of those experiences in cognitive development. One method used by many (including the present authors) is to record via a video camera the child's naturalistic interactions with toys and social partners in an effort to understand the structure and regularities inherent in these everyday interactions. There is, however, a potentially fundamental problem with this approach that is illustrated in Figure 1. The camera records a "third-person" view (Figure 1a), the view of an outside observer. This view is not the child's view, which might be more like that shown in Figure 1b. The third-person camera provides a fixed and broad view that is not at the body scale of a young child, and when coded by adults for the information present, is potentially biased by the adult's conceptualization of the structure of the larger scene. The purpose of this paper is to present preliminary results on a new method, that although far from perfect, may enable researchers to gain new insights into the contents of visual experience from the young learner's point of view.

This goal is consistent with a collection honoring themes in Esther Thelen's work because the input from the child's point of view is intimately tied to the child's body and to movement. The role of the body, action, and self-generated experiences is not a well-studied domain within cognitive development, although there are increasing calls for and evidence showing the value of such an approach (e.g., Thelen, 2000; Smith & Gasser, 2005; Riser, Lockman, & Nelson, 2005) and ample demonstrations of the importance of self-generated experiences in perceptual learning (e.g., Adolph & Berger, 2006; Berthenthal, Campos, and Kermoian, 1994; Gibson, 1969; Lockman, 1990; Ruff & Rothbart, 1996). The present goal is to specifically develop a method to

study the first-person view –a view intimately tied to self-action– in a developmental period (18 to 24 months) and task context (toy play) important to language and category learning. In the ideal, we would like a dynamic record of visual experience through the developing child's eyes. The method we have developed offers an approximation to this view: a lightweight camera worn low on the forehead of the child. This camera records the visual environment from the direction and body scale of the child and also provides a dynamic view of the available visual information that is aligned with the child's momentary activity.

Insert Figure 1 about here

Several other laboratories are also working on the development of head camera systems. A symposium at the 2006 meeting of the International Society on Infant Studies was specifically devoted to this topic (Fiser, Aslin, Lathrop, Rothkopf & Markant, 2006; Garciaguirre & Adolph, 2006; Yoshida & Smith, 2006; see also von Hofsten, 2006). Some researchers are particularly interested in head cameras as adjuncts or alternatives to eye-trackers (e.g., Fiser et al, 2006; von Hofsten, 2006). As a recent special issue of *Infancy* on an eye-tracking attests, infant eye movements are a promising new dependent measure with which to study such phenomena as categorization (McMurray & Aslin, 2004), predicting the movement of an object behind an occluder (GredebŠck & von Hofsten, 2004), object perception (Johnson, Slemmer & Amso, 2004), and face perception (Hunnius & Geuze, 2004). Such studies demonstrate how moment-to-moment eye movements are reliable indices of on-line processing. However, those eye-tracking studies all share four well-recognized limitations: (1) although some body movement can be tolerated, infants are required to sit relatively still and (2) passively watch (3) a 2-dimensional visual display that is (4) not selected through the subject's own actions but is rather chosen by the experimenter. These particular methods thus do not provide information about the typical contents of everyday experiences nor the fine-grained dynamics of eye movements in the service

of action. Accordingly, there is considerable interest in developing new methods such as head-mounted eye-trackers (e.g., Corbetta, Williams, & Snapp-Childs, 2007) as well as possibly head-mounted cameras.

Our interest in a head-mounted camera does not derive from an interest in eye-tracking per se nor specifically in the moment-to-moment dynamics of visual attention. Rather our long term goal in developing this method is a description of the visual experiences that may underlie object and action categories – for example, the surfaces of objects and object parts that are in view, the relation of objects and their surfaces to the child and to other objects as children act on those objects and as mature partners demonstrate the functions and uses of objects. Ultimately, we also want to study the co-occurrences of words with those experiences. Our conjecture is that early meanings – for categories, for nouns, for relational words – may be grounded, and thus more transparent, in these first person experiences.

The relation between our more limited goal and the much more ambitious one of tracking eye movements in an active moving child may be clarified by considering Figure 1 again. Figure 1a, the viewpoint of an outside observer, is the current standard for studying children's learning about objects and their names. Figure 1b is the head camera view at the same moment in time. The third field shows what might be viewed as the ideal, the precise points of visual fixation within the (dynamic) visual field of the child, data currently achievable only with a head-mounted eye-tracker (e.g., Corbetta, 2007). Our specific and limited goal in this paper is to validate a method that provides the middle panel, a first-person view of objects and actions on those objects.

The critical problem in achieving this goal is still knowing just where the eyes are directed. In the present procedure, when the camera is placed on the child's head, it is adjusted so that the focus of the child's attention is in the center of the head-camera field (and we know from pilot

testing that the camera does not slip). However, the head camera's view changes with changes in head position, not with changes in eye-gaze direction. Thus, if the direction of eye gaze shifts without a head movement, the head-camera view and the child's view might not be aligned. Analogously, if the eyes remained fixed on an object while the head moves, the head camera field and the visual field would again not be perfectly aligned. Thus, if head and eye movements are independent, the head camera view and the child's view may at best overlap imperfectly, or worse, the child's visual field could, in principle, fall outside of the head camera field. The experimental goal of studying characteristic object views and actions on objects may well be met even if there are misalignments. That is, some discrepancies in the head camera and true visual field might be tolerated, if the overlap between them is sufficient. Accordingly, the main purpose of Experiment 1 is to provide evidence for this overlap. Then in Experiment 2, we report the results of a first study using the head camera in the context of naturalistic toy play.

Experiment 1

Close coordination of head and eye movements has been documented in 2-to 4- month old infants tracking of moving objects (e.g., Daniel & Lee, 1990; von Hofsten & Roseander, 1996). Further, in studies of 9-to-10 month olds, Savelsbergh (1997) and Smith, Thelen, Titzer & McLin (1999) found that a visual target to the periphery elicited a directional shift not just of eye gaze but also the whole body. In a study of reach-to grasp movements in 2 to 8-year olds, Schneiberg, Sveistrup, McFayden, McKinley, & Levin (1984) also report a tightly coupled system of reaching movements that involve joint shifts in eye, head and trunk as well as arms. Research also shows that in adults, head movements are an integrated part of visually-guided reaching (Biguer, Jeannerod & Prablanc, 1982; Biguer, Prablanc, & Jeannerod, 1984), although eye movements usually precede coupled head movements by several hundred milliseconds (e.g., Jeannerod, Paulignan, & Weiss, 1998). In brief, if head and eye movements are coupled for toddlers when

they are manipulating toys, then there may be sufficient overlap between the head camera images and the child's visual field to make the head-camera useful.

Experiment 1 was designed to assess the correspondence between the head camera view and the direction of eye gaze in a calibration task with geometry similar to that of tabletop toy play. We focused on this body-world geometry because the task is one that is highly relevant in the everyday learning of children and also because it constrains attention and action in ways that may support the alignment of head camera field and visual field. In particular, preliminary pilot studies suggest that coupling of toddler head and eye movements are stronger when attention shifts are primarily in the horizontal rather than vertical direction and when the attended objects are in reaching distance rather than viewed from far away. The table-top context supports these goals.

In the experimental task, we directed children's attention to locations on the table and simultaneously recorded from the head camera and from a second camera directed at the child's eyes. These were then independently coded – frame-by-frame --to determine the correspondence in judged direction of attention from the head camera view and from the direction of eye gaze. Because active reaching might be important to the coupling of head and eye direction, the task was conducted under instructions that encouraged merely looking and also under instructions that encouraged reaching to the cued locations.

Method

Participants. A total of 12 infants, 18 to 24 months (mean age 21.6 months) were recruited. Ten infants (5 male, 5 female) (83%) tolerated the head-camera and contributed data, and five each were randomly assigned into two conditions: the Looking condition and the Looking and Reaching condition. The two infants who did not contribute the data did not let the experimenters place the head camera on their head.

Head Camera. The camera itself was a Watec (WAT-230A) miniature color camera weighing approximately 30 g and 36 X 30 X 15 mm. The focal length of the lens is f3.8mm (F2.0). The number of effective pixels are 512 (H) X 492 (V) (NTSC). The resolution (horizontal) is 350 lines. The camera's visual field is 90 degrees and thus about half the total visual field of the infants (Mayer & Fulton, 1993).

The camera was sewn into a headband that could be placed on the forehead such that the mounted mini-camera is close to the child's eyes. The vertical angle (elevation) of the camera is adjustable. The child also wears a lightweight vest with a microphone, batteries for the mini-camera, video, and audio transmitters. Through wireless data communication, the transmitters send visual and audio data to the base – a briefcase with audio receiver, visual receiver and DV recorder. The wireless transmitter and receiver operate on a 900MHz frequency allowing the broadcast of clear pictures up to 700 feet. In the present situation, the transmitter and receiver were separated by about 50 feet.

An additional third-person digital video camera was placed opposite from the child and directed at and zoomed in on the child's face and eyes with the goal of recording eye movements so as to provide an independent measure of eye gaze direction. The two cameras were synchronized by an Edriol (model V-4) 4 channel video mixer. A small video monitor in the experimental room (out of the subject's view) displayed the head camera image to the experimenter.

The method for placing the head camera on the child was derived from strategies used to attach electrodes for conventional clinical EEG recordings from infants and toddlers (see, Candy, Skoczenski & Norcia, 2001). As the child was seated at the table, a highly engaging pop-up toy was presented by the first experimenter. As the child pressed buttons to cause small animals to pop up, a second experimenter placed the headband on the child's head in one movement. The

first experimenter gently prevented (by placing her hands above the child's hands) any movement of the child's hands to the head and encouraged button pressing. As the child remained engaged in playing with the toy, the second experimenter gently adjusted the angle of the head camera so that the child's hand at the moment of a button press was in the center of the head-camera field (as viewed on the video monitor). The second experimenter also ensured that the third person camera was directly pointed at the child's face with the eyes (and direction of eye gaze) in clear view. Once the lightweight headband was placed, play with the pop-up toy continued until it appeared the child was no longer aware of the headband.

Task Procedure. The child sat at a small table, with a 120 X 75 cm surface top with the elongated axis parallel to the child's frontal plane. In the Looking condition, 3 different colored stickers were placed on a 120 x 75 cm board centered on the table, with the stickers at the two ends and at mid point. On each trial, the experimenter pointed to one sticker and encouraged the child to look at that sticker, saying, "Look, look at this. See this" while pointing and tapping the sticker. There were 6 pointing trials, 2 at each location, each lasting 5 seconds. The order of the pointing locations was randomly determined. In the Looking and Reaching condition, the stickers were replaced with small (5 cm^3) 3-dimensional toys fixed to the locations and the child was encouraged to reach to them, "Can you pet the doggie? Get the dog." All other aspects of the procedure were the same as in the Looking condition.

Coding and reliability. The video from the head-camera was scored –frame-by-frame – by two trained coders using the open-source MacSHAPA coding system (Sanderson et al, 1993). When viewing the head camera video, the coder's task for each frame was to determine the specific sticker or object in view. This resulted in a categorical judgment of whether the child was looking at the left target, middle target, right target or in some other direction that did not include any of the targets (e.g., looks to mother, floor, door etc). The video from the third-person camera

was also scored—frame-by-frame—by two different coders for direction of the child's eye gaze. Coders made the categorical judgments of whether the direction of eye-gaze was left, center, right or not possible to judge. Reliability was determined by having two coders independently code 25% of the frames selected from both cameras; the coders agreed in their categorical judgments on more than 94% of these frames, Cohen's Kappa = .886, $p < .001$.

Results

Although we assigned children to either the Looking condition or Looking and Reaching condition, all but one child reached on at least some trials and all children in both conditions did not reach to the target on at least some trials. Accordingly, we analyzed the data by both conditions (Looking versus Looking and Reaching instructions) and by whether individual looks (during the experimenter pointing event) were or were not accompanied by a reach.

Looks were defined from the third person camera image. A look was defined as a continuous series of 10 or more frames (333 msec) in which the direction of an eye gaze was judged to be on the same target. By this measure, children in the Looking and Reaching Condition had more individual looks to the target (mean = 17.6) than did children in the Looking condition (mean =8.8); however, this difference was not reliable ($t(8) < 1.00$). The mean durations of each look also did not differ between conditions, 2.28 seconds for looks in the Looking and Reaching condition and 2.34 seconds for looks in the Looking condition. In brief, children in the Looking and Reaching condition tended to make more looks to the target being pointed to by the experimenter than children in the Looking only condition (who looked away from the target and table somewhat more often), but the average duration of the looks were the same.

The key question is whether the view captured by the head camera is sufficiently well aligned with the direction of eye gaze that it may provide a useful measure of what the child sees. Accordingly, for each look, we measured the correspondence of the head camera to the direction

of that look by determining the overlap in terms of number of frames in the same direction for the head camera and the judged eye-gaze direction from the third-person camera. Across all infants and trials, there were 9107 frames that were included in looks as defined above; 87% of these (7923 frames) were focused on the same target as indicated by the independent judgment of the corresponding head camera view (Kappa = .846, P < 001). This indicates that, at least in this task context, 18- to 24-month-old infants typically turn their heads as well as their eyes to view an object. Thus, within the constraints of this task, the head camera seems likely to provide a reasonably accurate view of what is in the child's visual field.

Table 1 provides data for each of the 10 infants when looks occurred with and without reaches. As can be seen, the correspondence in judged direction of eye and head from the two cameras is high for all children except S2 in the Looking condition who sat very still through out the entire experiment and attempted to reach only once to the target. Across all children, the 13% of frames during looks in which judged eye and head direction did not correspond reflect the fact that eye shifts led head turns by, on average, 412 milliseconds. Although the data in Table 1 show that the lag of head turns behind eye shifts is slightly less for looks accompanied by reaches than for looks without reaches, there are no reliable effects of Condition (Looking versus Looking and Reaching, $F(1,8) < 1.00$) or of looks with and without reaches ($F(1,8) < 1.00$).

The main result, then, is this: The view of the head camera coincides with the direction of eye gaze but also systematically lags behind the shift in eye gaze, by slightly less than half a second. This correspondence occurs in a task setting in which children look at a constrained set of locations on a table and when they may reach to objects at those locations, and thus could be limited to this task context. Nonetheless, the degree of correspondence between head and eye direction appears sufficiently high that within these limitations, the images from the head camera may provide useful information about the child's perspective.

Experiment 2

The purpose of this experiment was to use the head camera in the naturalistic task of toy play with a similar geometry to that in Experiment 1. Mother and infants sit at a small table with a variety of toys with which to play. Mothers are asked to talk about the toys and to engage their children's attention as they naturally would in play.

Method

Participants. The 5 participants were 18-month-old (+/- 2 weeks) infants and one of their parents. Two additional infants rejected the head camera, yielding a success rate of 71%.

Procedure. The infant and the parent sat at a small (60 cm by 60 cm) table as shown in Figure 1a. The procedure used in Experiment 1 was again used to place and adjust the head camera on the child. A second camera—the third-person camera—recorded a broad view of the task context as shown in Figure 1a that included the tabletop, the child, and the parent.

The parent was given a box of 16 toys and told to select toys from the box in order to engage their child in play. Because we were particularly interested in the potential limits of the head camera to capture useful information about different kinds of objects and activities, parents were instructed to bring new toys into view periodically. Parents were told that multiple toys could be on the tabletop and used together. After fitting the child with the headband and giving instructions, the experimenters left the testing room. The entire play session, depending on the child's interest, lasted a minimum of 6 minutes and a maximum of 9 minutes.

Coding. Two coders coded the video from the head camera, frame-by-frame, for content. The nonmutually-exclusive coding categories included the individual toys (16 unique possibilities), the parent's face, the parent's hands, the child's hands, and whether the parent's or child's hand were in contact with a toy, and if so, which one. All objects in view -- those at the periphery and the center of the visual field (and accordingly at the periphery and the center of the

child's visual field) -- were counted if a sufficient proportion of the object was in view that the observer could recognize the specific toy, otherwise the object was not counted in view. Less than 1.5% of all head camera views were views away from the tabletop and/or parent (e.g., at the floor, door, wall). These frames are excluded in all reported analyses. Two coders also coded the specific objects that were on the table from the third person camera. In a separate pass, they also recorded the parent's actions from the video of the third-person view, noting the start and stop frame of any hand contact with a toy by the parent. For reliability, two coders independently coded a sample of 25% of all frames (13,581 frames). Agreement on the objects, hands, and face in view on the head camera was high, Kappa = .908, $p < .0001$. Agreement on the objects on the table from the third person view was 100%. The two coders' decisions about the start and stop times of parent hand contact with an object were within two frames of each other on 93% of the judgments.

Results

Our goal is to provide the first description of the head-camera images from toddlers with respect to the third person camera images in an effort to demonstrate the unique potential of this method for studying children's experiences with objects.

Number of objects in view. How different is the child's active view from the traditional static 3rd person view used to study parent and child interaction? If the child sits back in the chair and does not move (acting like a stable tripod for the head camera), then the entire table is in head camera view. However, the children were free to move themselves and their head close to the table and the toys; they were free to manually move objects close to the face and eyes, and they were interacting with social partners who might also make some objects more dominant in the child's view, Thus, if the child were not moving, the two views – from the third-person camera and from the head camera -- though recording the scene from different perspectives, should

contain roughly the same amount of information. Differences in these views provides one source of information on how the child's own actions select or constrain the available visual information. Accordingly, we compared images from the head camera and from the third-person camera for the number of objects in view as the parent brought objects onto the table and the child moved and reached to interact with them.

Figure 2 shows, for each subject, the proportion of frames with 0, 1, 2, 3, 4, 5 or more than 5 toys in view on the head and third person images. The third-person view provides the objective reality by showing the number of objects that parents have placed on the small table at any one time. Although parents differ in their tendency to put out many objects at the same time -- with some crowding the table with more and some with less -- for all children there were typically more than 4 toys on the table at any one time (mean = 4.95 across children). The head camera provides the view from the child's perspective, and from this view there is a dramatic reduction (mean = 1.39) with respect to the number of toys in view from the third person camera ($\underline{t}(4) = 7.94$, $\underline{p}<.001$). Figure 1a and 1b show a highly typical situation—3 objects in view from the third-person view but only one in the head camera view. Again, this reduction is not due to the small field of the head camera which is 90 degrees, nor is it a necessary consequence of the child's size or the position of the child's chair with respect to the table. As shown in Figure 3, when sitting back in the chair and not interacting with objects, the child can easily view the entire tabletop and its contents. There are at least three reasons for the reduction of the number of objects in the head-camera image with respect to those on the table: (1) in general, at any one moment, one object is closer to the child than all other objects, thus blocking the view of the other objects; (2) because the child actively brings the object of attention close to his or her face by manually moving the object to the body's midpoint and also by moving the head forward; and (3) parents often put one (and rarely two) objects directly in front of the child's face.

Figures 2 and 3 about here

This finding that the child's visual world during manual play with objects is dominated by one object at a time –even in a task context highly cluttered with interesting things– is potentially meaningful. It suggests that in active interactions with objects, sustained attention may be as much about actively bringing objects into the right body region as it is about bringing the eyes to those objects. This in turn raises interesting questions for future work about how, in contexts of manual interaction with objects, attention shifts from a momentarily dominant object to another and how social partners may direct and guide attention in those contexts. It seems possible in light of these data that tasks involving the active manipulation and manual movement of objects of interest may present the learner with a different kind of attentional task –and potentially different kinds of solutions– than ones involving the more passive viewing of objects (see Kidwell & Zimmerman, 2007, for relevant evidence). Indeed, although much research on joint attention has focused on eye-gaze following (see MacPherson & Moore, 2007, for review), there is increasing interest in the whole-body nature of attentional interactions in social settings (e.g., Kidwell & Zimmerman, 2007; Lindblom & Ziemke, 2006; Moll & Tomasello, 2007). Finally, the finding that there is typically one dominant object in the head camera image also boosts confidence that the head camera image is capturing child-relevant aspects of the scene.

Hands. Table 2 shows a further summary of the main contents of the head camera view for the five subjects. We specifically counted the frames in which the parent's face was in view (PF), in which the parent's hand was on an object (PHO), in which the child's hand was on an object (CHO), and in which there was an object with no hands in view (O). These are exhaustive but not mutually-exclusive categories (though in practice they virtually are). The single-most prevalent image is the child's own hands playing with a toy; 51% of the head

camera frames include the child's hands acting on a toy. The child's own hands on an object were the most frequent head camera view for all children, and more frequent than the second most frequent view—the parent's hands on an object ($t(4) -= 2.54$, $p < .06$). Again, in the context of toy play, visual experience seems not to be about vision alone but rather about seeing and doing. Thus, Table 2 highlights another staple of visual experience that may have been overlooked in analyses of the third person views: hands and hand actions. Overall, someone's hands, either the child's or the parent's, were in view and dynamically acting on an object in over 80% of the frames.

The parent's face. The limited appearance of the parent's face in the head-camera view was surprising. For the 18-month old children in this table top task of toy play, an image of the parent's face was in the head camera view on average less than 25 seconds of the about 7 minute play session. There are at least several reasons to be cautious about any interpretation of this result. First, the task of a brief period of toy play is mostly about doing things with toys, and thus parents and children were both jointly focused on the toys and may have directed each other's attention through manual action on the attended object. Thus, it seems likely that the dominance of objects and hands– rather than faces and eyes– may be highly specific to this type of activity. Second, the head camera view will not show brief glances to the parent's face (particularly if they involve vertical eye-movements) that are unaccompanied by head movements, and these may well have occurred. One observation hints that such brief glances could be a regular part of children's manual actions on objects. Figure 4 shows a brief timeline of four kinds of images for one subject: (1) the parent's face and (2,3, and 4) the child's own hand on each of three different objects that were on the table. For very brief durations but systematically –throughout the child's play– the head camera image shifted from toys to the mother's face. Although these looks were very brief, the mother's face was being

monitored. Thus, it is quite possible that there were more such looks but without corresponding head turns. An intriguing question for future work is how such brief glances (with or without head turns) may be integrated into children's object play and whether, in fact, they follow a rhythm (of checking up on the parent) that is independent of the child's momentary toy play goals as has been suggested by some researchers (Jones and Hong, 2001).

Attention shifting. The final set of analyses examined precursors to shifts in the head camera view as potential influences on the child's shifts in attention. Since the head camera view typically contains one or two objects, the presumed object(s) to which the child was attending, we defined "object shifts" as the appearance of an object in the head-camera view that was not in the just previous frame that included objects. By this system, a shift from an image of object A to the mother's face and then back to object A did not count as an "object shift", but a shift from a head-camera image with object A in it to one with object B or to one with object A and object B did (whether an image of the parent's face, for example, intervened or not). These "object shifts" include shifts due to both changes in the head-camera direction and also to new objects being manually brought into view by the child or by the parent. We then examined the contents of the preceding 30 head camera frames and 30 preceding third-person frames to determine what events might have precipitated the head camera shift. Three categories of preceding events were suggested: (1) looks to the parent's face (coded from the head camera), (2) parent hand actions (coded from the third person camera), and (3) spontaneous shifts (that is, with no obvious preceding visual event).

As is evident in Table 3, parent hand actions appear to play a significant role in organizing shifts among different in-view objects. On average, over 60% of all shifts of the head camera to a new object were immediately preceded by a hand action by the parent on the object to which the child then shifted attention. Again, these results suggest the perhaps significant role

of hand actions in orchestrating attention at least in the context of tabletop play with a mature social partner.

In summary, Experiment 2 demonstrates the usability of a head camera with toddlers in one task context and also suggests the character of the new insights that may emerge from the use of this technology: (1) the role of the body in attention, specifically in selecting and positioning single objects for attention and (2) the role of hand actions– the child's own and those of others– as generators of visual experience, conduits of meaning, and organizers of attention. Although the importance of hand actions have been recognized by others (e.g., Ruff & Lawson, 1990; Woodward, 2003), their structure as they act on objects from the first-person view may prove particularly informative.

General Discussion

The contribution of this paper is largely methodological -- a demonstration of a new method that promises, quite literally, a new way of viewing the experiences of children. The two experiments show that a head-camera attached to a headband worn low on the fore head is both tolerated by 18- to 24-month-olds and capable of capturing the first person visual field. Because shifts in eye-gaze and head direction are coupled for toddlers in the task context of tabletop toy play, there is considerable overlap between the head-camera images and the child's visual field. Because young children position objects with respect to their bodies so that one object dominates the viewing field at any one time, the head camera seems likely to capture the objects –and their views– that are of interest to children. However, because eye-shifts do systematically precede head shifts, and may occur without them in some important contexts such as monitoring the parent, and because the head camera field is relatively large, it may have limited use by itself in measuring the fine-grained temporal or spatial dynamics of children's attention.

Even with these limitations, the potential for this method appears considerable.

Attempting to "see" the learning environment from the child's perspective is likely to reveal structure and regularities different from those apparent in the more usual third person view. The dominance of hands –and hand actions– in the images from the head camera and the role of parent's hand movements in organizing the child's attention underscores this point. The child's everyday visual experience must include hours upon hours, day after day, of watching hands and their actions. This massive visual experience of hand actions may be part of the explanation of infants' early understanding of the causal implications of hand actions (Baldwin, 1993; Rothblat, 1987; Sommerville & Woodward, 2005; Woodward, 1998; Woodward, 2003). This fact may also be relevant to the intimate link between hand gestures and conceptual content in language (Goldin-Meadow, 2003a) and to the spontaneous invention by some children of hand gestures as a means of communicating (Goldin-Meadow, 2003b). The head-camera method offers a new technique to study these visual properties of hand actions and hand shapes from the first-person view.

Direct access to the first-person view should benefit research programs in other domains as well. One of these domains is the study of social cues in early language learning. By one conceptualization, the child's first task in learning language is mapping heard word forms to potential referents in the sensory stream (Gentner, 1982). As has been well argued (Quine, 1960; Snedeker, 2004), this is a difficult task; the sensory input at any moment offers an infinite number of referents and potential meanings. Recent research documents the powerful role of social-interactional cues in guiding infants' in-the-moment attention to the intended referent (Baldwin, 1993; Baldwin, 1996; Bloom, 2000; Tomasello, 2000, 2001, Tomasello & Akhtar, 1995; von Hoften, Dahlstrom & Frediksson, 2005; Woodward & Guajardo, 2002). Several researchers have argued for an analysis of social cues in terms of learned correlations among a variety of bodily indicators of a speaker's attention, including head direction, hand movements,

and posture (e.g., Newtson, et al, 1987; Smith, 2000b; Yu, Ballard, and Aslin, 2005). The head-camera could aid in the real time analyses of the rich field of bodily cues that regulate and inform learning in a social context.

An additional research domain in which the head camera may prove useful is the study of perception and action in the context of object exploration and symbolic play. For example, contemporary research in cognitive neuroscience indicates a strong link between visual object recognition and motor areas (e.g., Ernst, Banks & Bulthoff, 2000; James, Humphrey & Goodale, 2001). Further, action on objects has been shown to change perceived shape in young children (Smith, 2005). The developmental study of how hand actions inform perception and parsing of object shape could lead to profound new insights about the multi-modal nature of object recognition. New evidence also suggests that watching hand actions may be central to the development of the mirror neuron system in human infants (Falck-Ytter, Gredeback, & von Hofsten, 2006). Also pertinent to these issues is Ruff's (1986; 1989) landmark work on infants' manual exploration of objects. These studies present clear examples of how the information in the visual learning environment is structured by the child's own action. As infants finger, rotate, and bang objects they generate rich multimodal information and dynamically changing visual views of objects (see also, Bushnell, 1982; 1985; 1989). The dynamic structure of these self-generated views of objects is highly relevant to children's emerging object categories and their dynamic structure might be particularly well studied through the first person view.

Conclusion

Visual experience has a perspective, a spatially circumscribed view of the world. The motivation behind developing a head camera is the idea that the view, the perspective of the learner, determines the structure of the learning task for the learner and thus may matter deeply as a force for developmental change. The learner's view is also always tied to the learner's body

and its momentary disposition in space, which also matter deeply in defining the learning task, possible solutions to that task, and developmental process. The present experiments demonstrate the validity of using a head camera to study visual experience from the perspective of the developing child. Young children tolerate this device (indeed, once it's placed most seem to forget about it). Further, and more critically, at least in the geometrically constrained task of tabletop toy play, the head camera appears to substantially captures the child's visual field. There is much we do not know about children's experiences and their role in development. The head camera offers one new method through which we may discover new and important regularities in children's experiences.

Author Note

This work was supported by NIH grant R21EY017843-01. We thank Megumi Kuwabara, Charlotte Wozniak, and Elizabeth Hanibal for their assistance in data collection and coding.

Corresponding author: L. Smith, Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405 (smith4@indiana.edu)

References

Adolph, K. E. (1995). A psychophysical assessment of toddlers' ability to cope with slopes. Journal of Experimental Psychology: Human Perception and Performance, *21*, 734-750.

Adolph, K. E. & Berger, S. A. (2006). Motor development. In W. Damon & R. Lerner (Series Eds.) & D. Kuhn & R. S. Siegler (Vol. Eds.), Handbook of child psychology: Vol 2: Cognition, perception, and language (6th ed.) New York: Wiley, pp. 161-213.

Baldwin, D. (1993). Early referential understanding: Infant's ability to recognize referential acts for what they are. Developmental Psychology, *29*, 832-843.

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., & Tidball, G. (1996). Infant's reliance on a social criterion for establishing word-object relations. Child Development, *67*, 3135-3153.

Baldwin, D. A., & Moses, L. J. (1994). Early understanding of referential intent and attentional focus: Evidence from language and emotion. In C. Lewis & P. Mitchell (Eds.), Children's early understanding of the mind (pp. 133-156). Hillsdale, NJ: Erlbaum.

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. Behavioral and Brain Sciences, *20* (4), 723-767.

Bertenthal, B. I., Campos, J. J. and Kermoian, R. (1994) An Epigenetic Perspective on the Development of Self-Produced Locomotion and Its Consequences. *Current Directions in Psychological Science, 3* (5) 145-140

Biguer, B., Jeannerod, M., & Prablanc, C. (1982). The coordination of eye, head, and arm movements during reaching at a single visual target. Experimental Brain Research, *46* (2), 301-304.

Biguer, B., Prablanc, C., & Jeannerod, M. (1984). The contribution of coordinated eye and head movements in hand pointing accuracy. Experimental Brain Research, *55* (3), 462-469.

Bloom, P. (2000). How children learn the meanings of words. Cambridge, MA: The MIT Press.

Bornstein, M. H., & Sigman, M. D. (1986). Continuity in mental development from infancy. Child Development, *57* (2), 251-274.

Bushnell, E.W. and Boudreau, J.P. (1993) Motor development and the mind: The potential role of motor abilities as determina of aspects of perceptual development. Child Development, *64*, 1005-1021.

Butterworth (1991) The ontogeny and phylogeny of joint visual attention In Natural theories of mind: Evolution, development, and simulation of everyday mind reading, A. Whitten (Eds.), pp. 223-232, Oxford, England; Blackwell.

Candy, T. R., Skoczenski, A. M., & Norcia, A. M. (2001). Normalization models applied to orientation masking in the human infant. Journal of Neuroscience, *21*(12), 4530-4541.

Clark, H. H. (1996). Common ground. In Using language. pp. 92-103. Cambridge: Cambridge University Press.

Corbetta, D., Williams, J., & Snapp-Childs, W. (2007) Object Scanning and its Impact on Reaching in 6-to-10 Months Old Infants. Presented at the meetings of the Society for Research in Child Development, Boston.

Daniel, B. M., & Lee, D. N. (1990). Development of looking with head and eyes. Journal of Experimental Child Psychology, *50* (2), 200-216.

Ersnt, M. O., Banks, M. S., & Bulthoff, H. H. (2000). Touch can change visual slant perception. Nature Neuroscience, *3* (1), 69-73.

Falck-Ytter, T., Gredeback, G. & von Hofsten, C. (2006). Infants predict other people's action goals. Nature Neuroscience, 9, 878 – 879.

Fiser, J., Aslin, R., Lathrop, A., Rothkopf, C., and Markant, J. (2006) An infants' eye view of the world: Implications for learning in natural contexts, International Conference on Infant Studies, Kyoto.

Frick, J. E., Colombo, J., & Saxon, T. F. (1999). Individual and developmental differences in disengagement of fixation in early infancy. Child Development, *70* (3), 537-548.

Garciaguirre, J. & Adolph, K. (2006) Infants' Everyday Locomotor Experience: A Walking and Falling Marathon . International Conference on Infant Studies, Kyoto.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj II (Ed.), Language Development (Vol. 2). Hillsdale, NJ: Erlbaum.

Gibson, E. J. (1969). Principles of perceptual learning and development. Appleton-Century-Crofts, East Norwalk, CT: US.

Goldin-Meadow, S. (2003a). Hearing gesture: How our hands help us think. Cambridge, MA: Harvard University Press.

Goldin-Meadow, S. (2003b). The resilience of language: What gesture creation in deaf children can tell us about how all children learn language. New York: Psychology Press.

Goldsmith, H. H., & Rothbart, M. K. (1991). Contemporary instruments for assessing early temperament by questionnaire and in the laboratory. In J. Strelau & A. Angleitner (Eds.), Explorations in temperament: International perspectives on theory and measurement. Perspectives on individual differences (Vol. xvii, pp. 249-272). New York, NY, US:

Plenum Press.

Gredeback, G., & von Hofsten, C. (2004). Infants' evolving representation of moving objects between 6 and 12 months of age. Infancy, *6*, 165-184.

Gre'zes, J. & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation and verb generation of action: A meta-analysis. Human Brain Mapping, *12*, 1–19.

Hains, S. M. J., & Muir, D. W. 11996). Infant sensitivity to adult eye direction. Child Development, *67*, 1941-1951.

Hood, B. M. (1995). Gravity rules for 2 to 4-year-olds. Cognitive Development, *10* (4), 577-598(522).

Hunnius, S., & Geuze, R. H. (2004). Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: A longitudinal study. Infancy, *6*, 231-255.

James, K. H., Humphrey, G. K., & Goodale, M. A. (2001). Manipulating and recognizing virtual objects: Where the action is. Canadian Journal of Experimental Psychology, *55* (2), 111-120.

Jankowski, J. J., Rose, S. A., & Feldman, J. F. (2001). Modifying the distribution of attention in infants. Child Development, *72* (2), 339-351.

Jeannerod, M., Paulignan, Y., & Weiss, P. (1998). Grasping an object: One movement, several components. Novartis Foundation Symposium, 218, 5-16; discussion 16:20.

Johnson, S. P., Slemmer, J. A., & Amso, D. (2004). Where infants look determines how they see: Eye movements and object perception performance in 3-month-olds. Infancy, *6*, 185-201.

Jones, S.S., & Hong, H.W. (2001). Onset of voluntary communication: smiling looks to mother. Infancy, *2*, 353–370.

Kidwell, M., & Zimmerman, D. H. (2007). Joint attention as action. Journal of Pragmatics, *39*(3), 592-611.

Kleinke, C. (1986). Gaze and eye contact: A research review. Psychological Bulletin, *100*, 78-100.

Lawson, K. R. R., Holly A. (2004). Early focused attention predicts outcome for children born prematurely. Journal of Developmental & Behavioral Pediatrics, *25* (6), 399-406.

Lindblom, J., & Ziemke, T. (2006). The social body in motion: Cognitive development in infants and androids. Connection Science, *18*(4), 333-346.

Lockman, J. J. (1990). Perceptuomotor coordination in infancy. In C. A. Hauert (Ed.), Developmental psychology: Cognitive, perceptuo-motor and neuro-psychological perspectives. Amsterdam: North-Holland Elsevier.

MacPherson, A. C., & Moore, C. (. (2007). Attentional control by gaze cues in infancy. In R. Flom, K. Lee & D. Muir (Eds.), Gaze-following: Its development and significance. (pp. 53-75). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Markova, G., & Legerstee, M. (2006). Contingency, imitation, and affect sharing: foundations of infants' social awareness. Developmental Psychology, *42* (1), 132-141.

Mayer DL, Fulton AB. Development of the human visual field. In: Simons K, ed. Early Visual Development, Normal and Abnormal. New York: Oxford University Press, 1993:117–29.

McMurray, B., & Aslin, R. N. (2004). Anticipatory eye movements reveal infants' auditory and visual categories. Infancy, *6*, 203-229.

Moll, H., & Tomasello, M. (2007). How 14-and 18-month-olds know what others have experienced. Developmental psychology, *43*(2), 309-317.

Newell, K. M. (1978). Some issues on action plans. In G. E. Stelmach (Ed.), Information processing in motor control learning (pp. 41-54). New York: Academic Press.

Newtson, D., Hairfield, J., Bloomingdale, J., & Cutino, S. (1987). The structure of action and interaction. Special Issue: Cognition and action. Social Cognition, *5*, 191-237.

Quine, W. (1960). Word and object. Cambridge, MA: MIT Press.

Rieser, J. J., Lockman, J. J., & Nelson, C. A. (2005). Action as an organizer of learning and development: Volume 33 The Minnesota symposia on child psychology Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Richards, J. E., & Turner, E. D. (2001). Extended visual fixation and distractibility in children from six to twenty-four months of age. Child Development, *72* (4), 963-972.

Roitblat, H. L. (1987). Introduction to comparative cognition. New York: Freeman.

Ruff, H.A. (1986). Components of attention during infant's. manipulative exploration. *Child Development, 57* 105-114.

Ruff, H.A. (1989). Infants' manipulative exploration of objects: Effects of age and object characteristics. *Developmental Psychology,20*, 9-20.

Ruff, H. A., & Lawson, K. R. (1990). Development of sustained, focused attention in young children during free play. *Developmental psychology, 26* (1), 85-93. Ruff, H. A., & Rothbart, M. K. (1996). Attention in early development: Themes and variations. New York, NY.

Rutter, D. R. (1984). Looking and seeing: The role of visual communication in social interaction. New York: Wiley.

Savelsbergh, G., von Hofsten, C., & Jonsson, B. (1997). The coupling of head, reach and grasp

movement in nine months old infant apprehension. Scandinavian journal of psychology, *38* (4), 325-333.

Sanderson, P.M., Scott, J.J.P., Johnston, T., Mainzer, . Wantanbe, L.M.< & ames, .M. (1994) MacSHAPA and the enterprise of Exploratory Sequential Data Analysis. International Journal of Human Computer Studies, *41,* 633-681. www.openshapa.org

Schneiberg, S., Sveistrup, H., McFadyen, B., McKinley, P., & Levin, M. F. (2002). The development of coordination for reach-to-grasp movements in children. Experimental Brain Research, *146* (2), 142-154.

Smith, L.B., & Gasser, M. (2005). The development of embodied cognition: six lessons from babies. Artificial Life, *11*, 13–30.

Smith, L. B., Thelen, E., Titzer, R., & McLin, D. (1999). Knowing in the context of acting: The task dynamics of the a-not-b error. Psychological Review, *106* (2), 235-260.

Snedeker, J., & Gleitman, L. (2004). Why it is hard to label our concepts. In Hall & S.Waxman (Eds.), Weaving a lexicon (pp. 257-294). Cambridge,MA: MIT Press.

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. Cognition, *96*, B1ÐB11.

Thelen, E. (2000). Motor development as foundation and future of... Mechanisms of categorization in infancy. Infancy, *1*, (1), 59-76.

Tomasello, M. (2000). The social-pragmatic theory of word learning. Pragmatics, *10*, 401-14.

Tomasello, M. (2001). Perceiving intentions and learning words in the second year of life. In M. Bowerman & S. Levinson (Eds.), Language Acquisition and Conceptual Development. (pp. 111-128): Cambridge University Press.

Tomasello, M., & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. Cognitive Development, *10*, 201-224.

von Hofsten, C., & Rosander, K. (1996). The development of gaze control and predictive tracking in young infants. Vision Research, *36*(1), 81-96.

Von Hofsten, C. (2006) An action perspective on early cognitive development , Presented at the meetings of the International Conference of Infant Studies, Kyoto.

Von Hofsten, C., Dahlstrom, E. & Fredriksson, Y. (2005). 12-month-old infants perception of attention direction in static video images. Infancy, 8, 217-231.

van Hof-van Duin & Mohn (1986) The development of visual acuity in normal fullterm and preterm infants. Vision Research, *26* (6) 909-16.

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1–34.

Woodward, A. L. (2003). Infants' developing understanding of the link between looker and object. Developmental Science, *6* (3), 297-311.

Woodward, A., & Guajardo, J. (2002). Infants understanding of the point gesture as an object-directed action. Cognitive Development *17*, 1061-1084.

Yoshida, H. & Smith, L.B. (2006) From the first-person view: Joint attention is through the hands not eyes. International Conference on Infant Studies, Kyoto.

Yu, C., Ballard, D.H., & Aslin, R.N. (2005). The role of embodied intention in early lexical acquisition. Cognitive Science, *29* (6), 961–1005.

Table 1. Number of looks, proportion of corresponding frames (coded direction of eye gaze and direction of head camera image), and delay in direction shift (direction of eye gaze minus head camera) for children in the two Instruction conditions and for looks with and without reaches.

		Looks with reaches			Looks without reaches		
Instruction	Subject	Number Looks	Matching Frames	Delay (msec)	Number Looks	Matching Frames	Delay (msec)
Looking	1	11	.95	700	11	.94	833
And	2	24	.95	450	8	.98	200
Reaching	3	7	.91	200	9	.93	867
	4	8	.91	303	7	.93	1233
	5	7	.81	166	6	.79	466
Group mean		**11. 4**	**.91**	**363.8**	**8.2**	**.91**	**719.8**
Looking	6	3	1.00	0	7	.94	100
	7	1	1.00	0	6	.50	500
	8	0			6	.91	500
	9	5	.83	333	6	.93	566
	10	4	.89	100	6	.90	-133
Group mean		**2.6**	**.93**	**108.2**	**6.2**	**.84**	**306.6**
Experiment mean		**7.0**	**.92**	**250.2**	**7.2**	**.88**	**513.2**

Table 2. Time in seconds for 4 major contents of head camera images in Experiment 2 (excluded are times when the head camera is directed away from the table) for the five participants: The parent's face (PF), the parent's hand on an object (PHO), the child's hand on an object (CHO), or an object (or objects) without hand in contact (O). (These categories are not strictly mutually exclusive –though in practice they virtually are –and thus the sum of the proportions of total frames with these contents given at the bottom of the Table slightly exceeds 100%).

Subject	PF	PHO	CHO	O	Total duration
1	15.9	55.8	258.7	39.3	369.7
2	13.5	128.6	219.6	94.6	456.3
3	35.7	116.4	220.9	100.4	473.4
4	17.3	210.7	239.5	73.7	541.2
5	43.5	156.8	155.0	52.6	407.8
Mean time (secs)	25.2	133.6	218.7	72.1	449.7
Proportion frames	.06	.31	.51	.17	

Table 3. Number of changes among distinct object images in the head camera view and the proportion of these that were preceded by a look to the parents ‘ face (PF), by a parent hand action (PHA), or that were directly from one object to another, or apparently spontaneous (S) for the five participants in Experiment 2.

Subject	Number	PF	PHA	S
1	51	.21	.61	.18
2	48	.17	.44	.39
3	59	.14	.68	.18
4	62	.16	.60	.24
5	51	.04	.69	.27
Mean	54.2	.16	.60	.25

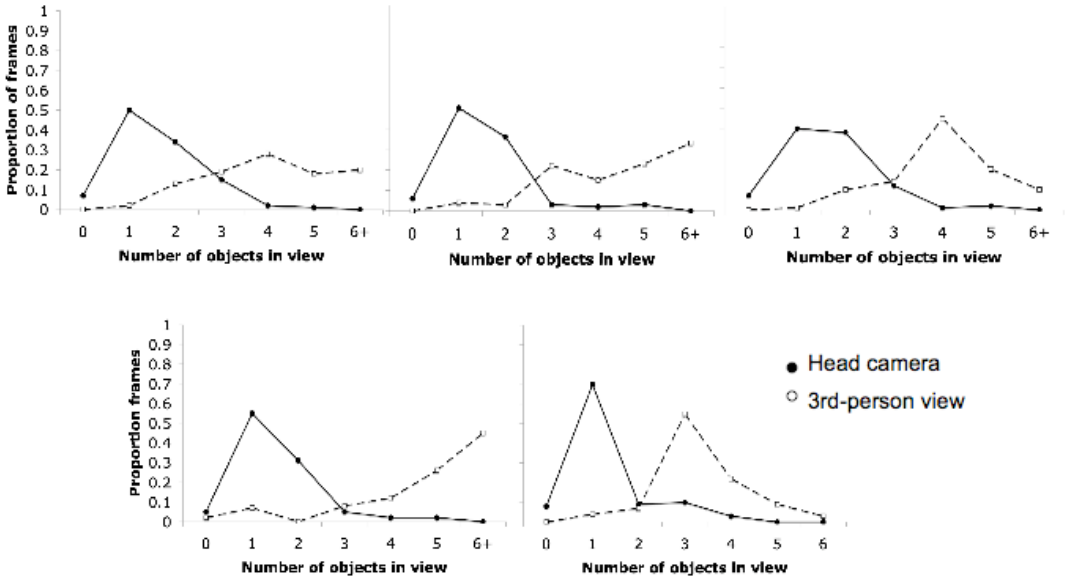
Figure captions

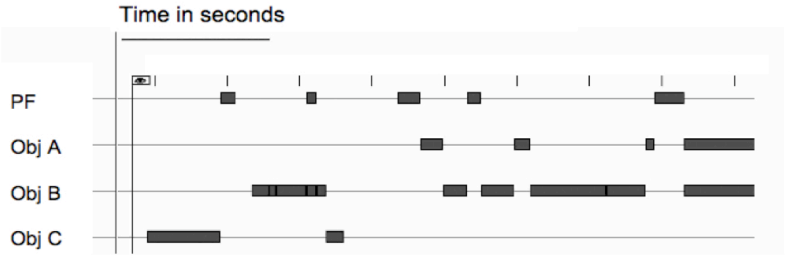
Figure 1. (a) Mother and child in table toy task of Experiment 2. (b) The head camera image at the same point in time. (c) An illustration of the unresolved issued of just where in the image the child is looking.

Figure 2. The proportion of frames in which 0, 1, 2, 3, 4, 5 and more than 6 objects were in view from the head camera (solid line) and from the third-person camera (dashed line) for each of the five children in Experiment 2.

Figure 3. A head camera view (a) and simultaneous third camera view (b).

Figure 4. A 50 second timeline of head-camera images for one child. Shown is the sequence and duration of 4 different images in the head camera: PF-Parent face, Obj. A, B or C, three different toy objects.


Proportion of frames
Number of objects in view
Number of objects in view
Number of objects in view
Proportion frames
Number of objects in view
Number of objects in view
Head camera
3rd-person view


Time in seconds
PF
Obj A
Obj B
Obj C