Developmental Science (2009), pp 1–9 DOI: 10.1111/j.1467-7687.2009.00947.x

PAPER

Not your mother's view: the dynamics of toddler visual experience

Linda B. Smith, Chen Yu and Alfredo F. Pereira

Department of Psychological and Brain Sciences, Indiana University, USA.

Abstract

Human toddlers learn about objects through second-by-second, minute-by-minute sensory-motor interactions. In an effort to understand how toddlers' bodily actions structure the visual learning environment, mini-video cameras were placed low on the foreheads of toddlers, and for comparison also on the foreheads of their parents, as they jointly played with toys. Analyses of the head camera views indicate visual experiences with profoundly different dynamic structures. The toddler view often consists of a single dominating object that is close to the sensors and thus that blocks the view of other objects such that individual objects go in and out of view. The adult view, in contrast, is broad and stable, with all potential targets continually in view. These differences may arise for several developmentally relevant reasons, including the small visuo-motor workspace of the toddler (short arms) and the engagement of the whole body when actively handling objects.

Introduction

Human toddlers are the most powerful learning devices known. In domains such as language, categories, object recognition, and naïve physics, very young children exhibit formidable learning skills unmatched by the most powerful artificial intelligence or advanced robots built to date (e.g. Smith & Gasser, 2005). Contemporary theories attempt to explain this prowess via domain specific learning mechanisms (e.g. Carey, 2009), powerful statistical learning (e.g. Xu & Tenenbaum, 2007) and the social contexts in which toddlers learn (e.g. Tomasello, 2007). However, there is much that is not known about the learning environment itself, and the data on which any of the proposed learning mechanisms must operate. One limitation on current understanding is that descriptions of the toddler learning environment are based on our adult point of view. Here we show that in at least one common everyday learning context, the dynamic structure of toddler visual experience differs fundamentally from that of adults, and does so in ways that may matter deeply for understanding learning.

The possibility of consequential differences in toddler and adult experience arises because of considerable differences in toddler and adult bodies. Learning is the accrued effects of momentary sensory-motor events that are tightly tied to the body's morphology and movements. In vision, the moment-to-moment pattern of stimulation depends on the orientation of eyes, head, and whole body with respect to the physical world and, critically, also on the movements of hands as they grasp, turn and move objects. All these movements, in turn, depend on the interests of the perceiver and will be influenced – moment by moment – by the perceiver's own actions and those of social partners as they cause objects to come into and go out of view. To the degree that toddlers' bodies, movements, and interests are not like those of adults, then the dynamic structure of toddler visual experience – the data on which learning depends – may differ significantly from that of adults.

The question of how toddlers' own actions determine their dynamic visual experience is particularly compelling in the context of growing evidence of tight links between visual and motor development (e.g. Bertenthal, Campos & Kermoian, 1994; James, In press; Soska, Adolph & Johnson, in press; Smith & Gasser, 2005; Thelen & Smith, 1994). Correlational studies show close dependencies between motor achievements and visual processing in individual infants and toddlers (e.g. Soska *et al.*, in press; Bertenthal *et al.*, 1994). Experimental studies show that enriched motor experiences can accelerate children's perceptual and cognitive development (e.g. Bertenthal, Campos, and Kermoian, 1994; Bojczyk & Corbetta, 2004; Needham, Barrett & Peterman, 2002; James, in press). Such findings point to the theoretical importance of a detailed understanding of active vision and the visual experiences created by body movements. At present, we know very little about the dynamic structure of developing infants' and children's visual

Address for correspondence: Linda B. Smith, 1101 East 10th Street, Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA; e-mail: smith4@indiana.edu

experiences in the natural everyday activities that provide the context for development.

The goal of this study is to describe toddlers' first-person visual experience in one everyday context; the study specifically asks how toddlers' own actions may play a role in selecting visual information. The world is highly cluttered with many potential targets of attention and learning. Processes that limit and reduce the available information and that select and focus attention are thus critical to learning. Accordingly, the experiment and analyses were designed to examine whether – and in what way – the child's dynamic view effectively selects and reduces available information. The task we chose was toy play with multiple available toys, on a table top, and with a mature social partner (the toddler's parent). We chose this task for three reasons. First, toy play with a social partner is common in lives of toddlers. Second, multiple toys and an engaged social partner create a naturally complex visual context with multiple visual targets and opportunities to shift attention. Third, we chose toy play on a tabletop because it has a constrained geometry (albeit, a natural one) which makes our method possible.

To capture the toddler's first-person view of events, we placed a tiny video-camera low on the toddler's forehead (and, for comparison, also one on the parent's forehead). Our specific goal was to describe the objects in view at any moment –and the changes in those views – as the toddlers moved their heads and hands (and the objects) in active play. The head camera provides a broad view that moves with head movements but not with eye movements. In a prior calibration study using the same table-top geometry, Yoshida and Smith (2008) independently measured eye gaze direction and found that the head camera view and eye gaze direction of toddlers were highly correlated in this context such that 90% of head camera frames coincided with independently coded directions of eye gaze. Although head and eye movements can be decoupled, the restricted geometry and the motor behavior of toddlers at play creates a context in which the head camera field is a good approximation of the contents of the toddler's first-person view.

Method

Participants

Ten children (half male, between 17 and 19 months of age) and their parent contributed data; four additional children were recruited but refused to wear the head camera.

Stimuli

Eighteen toys were organized into six sets of three. The toys were about 10 cm^3 in volume, and included dishes, animals, and various shaped blocks. All had simple shapes and a single main color.

Head cameras

The toddler and participating parent wore identical head cameras, each embedded in a sports headband. The cameras are Supercircuits (PC207XP) miniature color video cameras weighing approximately 20 g. The focal length of the lens is f3.6mm. The number of effective pixels are 512 (H) × 492 (V) (NTSC). The resolution (horizontal) is 350 lines. The camera's visual field is 70 degrees, this is a broad view but less than the full visual field (approximately 180^o). We consider implications with respect to the periphery in the General Discussion. The direction of the camera lens when embedded in the sports band was adjustable. Input power and video output went through a camera cable connected to a wall socket, via a pulley, so as not to hinder movement. The head cameras were connected via standard RCA cables to a digital video recorder card in a computer in an adjacent room.

Bird's-eye view camera

A high-resolution camera was also mounted right above the table with the table edges aligned to the edges of the bird-eye image. This view provided visual information that was independent of the gaze and head movements of the participants.

The experimental environment

Figure 1 shows the set-up and the two parent seating arrangements that were used. The table (61cm × 91cm × 64cm), walls and floor were white and participants wore white smocks leaving the toys, hands and faces as the only nonwhite objects in the images (this supports computer object recognition, see below). The child's seat was 32. 4 cm above the floor (average distance of eye to the center of the table: 43.2 cm). Parents participated in one of two sitting positions relevant to the comparison of parent and child head camera views. Half the parents sat naturally on a chair at the table. Since parents are taller than their toddlers, this means that parents' heads, eyes, and head cameras were higher above the table than the toddlers' heads, eyes and head cameras (average distance of eye to table center for parents in chairs: 68.6 cm). The remaining parents sat on the floor such that their eyes, heads and head cameras were at approximately the same distance from the tabletop as their toddler (average distance of eye to table center for parents sitting on the floor: 44.5 cm).

Procedure

White smocks were put on by both participants. The child was then seated and distracted with a push-button pop-up toy while a second experimenter (from behind) placed the headband low on the forehead. One experimenter then directed the child to push a button on a pop-up toy while the second experimenter adjusted the

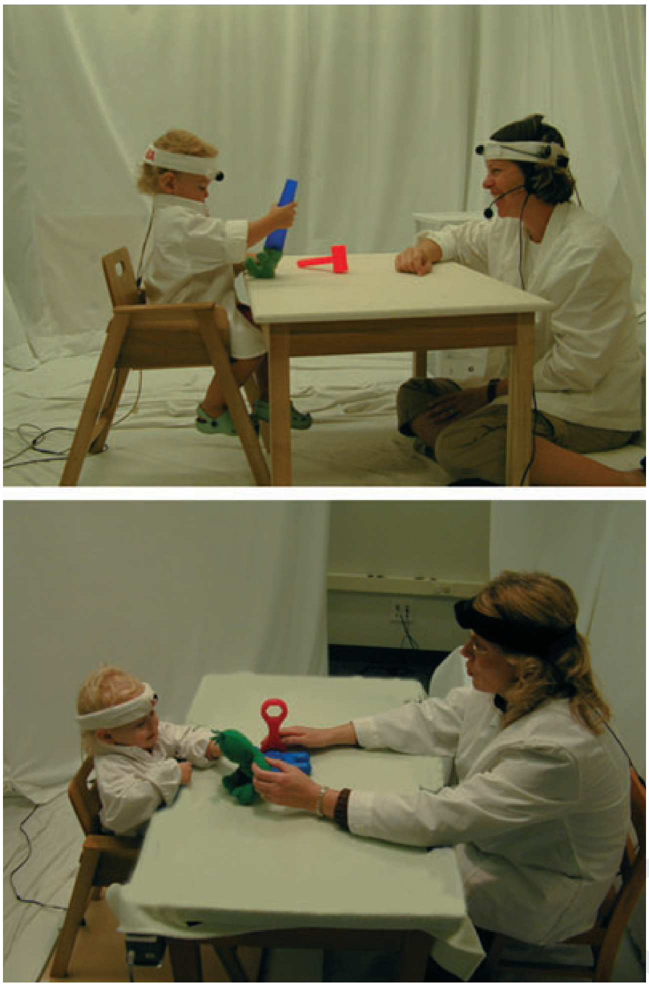

Figure 1 *The table-top set-up: White background, head cameras on parents and toddlers, and parents in one of two possible sitting arrangements.*

camera such that the button being pushed by the child was near to the center of the head camera image. The parent's head camera was then put on and similarly adjusted. Parents were told that the goal of the study was simply to observe how their child played with toys and that they should try to interact as naturally as possible. They were specifically told to give their child three toys at a time and to change toys when they were signaled to do so by the experimenter. The experimenters then left the room and the play session began. There were six three-object trials, each about 1 minute. The entire study, including set-up, lasted 15 minutes.

Data processing

The recording rate for each camera is 10 frames per second, yielding approximately 190,000 image frames from each dyad. The locations and sizes of objects and body parts (skin-colored blobs) were extracted automatically using information from all three time-locked cameras to resolve ambiguities. The first step was separation of the background and object pixels in the raw images. Since the experimental room was white except for hands, faces and toys, the procedure treats close-to-white pixels as background. Non-background pixels were then broken into several blobs using a segmentation algorithm that creates groups from adjacent pixels that have color values within a small threshold of each other and then creates larger groups from these initial groups by using a much tighter threshold. This second step of the algorithm attempts to determine which portions of the image belong to the same object even if that object is broken up visually into multiple segments (e.g. when a hand decomposes a single object into several blobs). These blobs were then input to a pre-trained object recognition model that was also helped by the simple shapes and single colors of the objects. The model yields a probabilistic map of the likelihood that each segmented blob in an image belonged to the candidate object. The object detection algorithm assigned an object label for each blob by putting probabilistic maps of all the possible objects together, and by considering the spatial coherence of an object. Comparison of object labels by this automatic procedure to frame-by-frame hand coding (for about 1000 frames) yields over 95% agreement (records of the specific objects on the table at each moment, recorded from the bird's-eye camera, indicate that object recognition from the head camera views was slightly more accurate under automatic coding than under human coding). Hand coding was also used on selected frames from the bird's-eye view (about 80,000 frames) to determine who was holding a particular object; two coders independently coded the same 25% of the frames (checking head camera images to resolve any ambiguities) with 100% agreement.

Results

On each experimental trial there are three objects on the table and thus three objects that could be in the child's view. These three objects are all approximately the same actual size and thus when measured from the overhead bird's-eye camera, each object takes up roughly the same amount of area in the images from that camera. Further, if the child were to sit back and take a broad view of the table, not moving his or her head, all three objects would be in view in the 70^{o} head camera image and would all have approximately the same image size. However, if the child moves his body and/or moves the objects so that one object is closer to the head and eyes than other objects, then that selected object will be larger than the other objects and, being closer to the sensors, it could even obstruct the view of the other objects. If the child's head movements or manual actions on the objects focus successively on one then another object, then the head camera images should show dynamic variation in the objects in view and in the relative sizes of those objects in the head camera view. Accordingly, the objects in the head camera image and their image sizes provide a

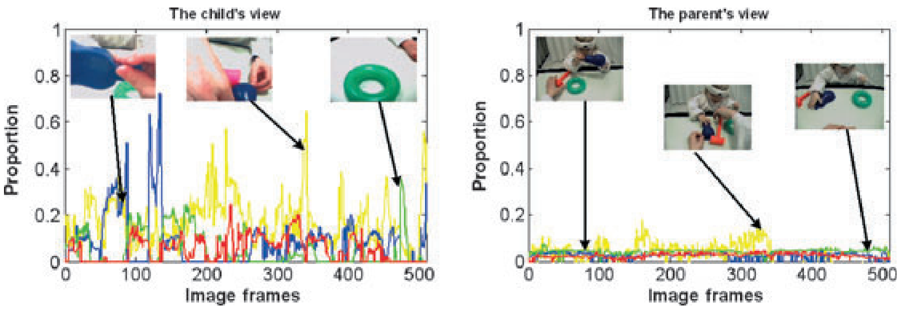

Figure 2 *Time series of the changing dominance of the objects in the head camera images from child and parent dyad. The figures show proportion of the head camera field (size of object in terms of pixels relative to the size of the whole head camera image) taken up by each of the three toy objects and by hands in the images from the child camera and the parent camera. The changing frame-by-frame sizes of the three toy objects (red, green and blue) are indicated by the corresponding colored lines. The yellow line indicates the proportion of the field that is images of exposed body parts (combined mother and child hands and faces). The text provides aggregate statistics across all participating dyads.*

measure of how the child's own activity selects visual information. These, then, are the principal dependent measures in the following analyses.

Figure 2 shows the frame-by-frame changes in the head camera image sizes of the three objects for one dyad for one trial. For this figure, size for each object is calculated in terms of the proportion of pixels in the image that belong to each of the three toys. Also shown is the proportion of the image taken up by body parts (faces and hands of both participants in aggregate as these are not discriminated in the automatic coding). The pattern shown in the figure is characteristic of all dyads on all trials and is the main result. The toddler view is one in which, at any one moment, one toy is much larger than the other toys in the image and the largest object in the image changes often. In contrast, the parent view is broad, stably containing all three objects, with each taking up a fairly constant and small portion of the head camera field.

Statistics of the sizes of objects calculated over all dyads show the same pattern. Over all head camera frames, the toys took up three times as much area in the child's head camera image as in the parent's head camera image (M = .15 versus .05, $t(18) = 8.78$, $p < .001$) which means that the toys were closer to the toddlers' heads and eyes than to the parents' heads and eyes. Moreover, the two sitting arrangements for parents did not differ ($t(8) = 0.711$; $p > .491$), on this (or any other) measure. Finally, the average proportion of the head camera image occupied by body parts (faces and hands) was small (.05) and was the same for both parents and children.

Dominating objects

A visual world in which one object is often closer to the sensors than others is a form of selection, potentially reducing competition among scene objects for attention and processing. Accordingly, our first measures of selectivity asked whether there was a 'dominating object' in the child and parent views, with the dominating object defined in terms of its relative size, that is, as being the largest – and thus closest to the sensors – compared to the other in-view objects. More specifically, each frame was defined as having a single dominating object if the size of one object was at least twice the combined size of all other objects (or object fragments). Only .08 of the frames from the parent view but .30 of the frames from the child view had a single dominating object; thus, substantially more toddler views were dominated by a single object than were parent views, both when parents sat on chairs ($t(8) = 5.48$, $p < .001$) and when parents sat on the floor ($t(8) = 4.86$, $p < .005$). These differences were calculated in terms of proportions of all frames; however, sometimes children were 'off-task' – not playing with the toys and not looking at the table top but rather looking the ceiling camera, the door, the floor, or the parent's face. Specifically, for .21 of the frames there was no object in the child's view, compared with .07 of the adult view frames. If we exclude all the no-object frames from consideration, then the difference between the child and parent head camera images in terms of a dominating object is even larger (.38 of children's head camera frames are characterized by one dominating object whereas only .09 of parent head camera frames are). In sum, the adult view includes and is equal distance from all of the objects on the table top; but in marked contrast, the child's view often contains one dominating object that is closer to the head and eye and thus often blocks the view of the other objects.

Figure 3 provides converging evidence for these conclusions. Here we define the dominating object as simply the largest of the three objects (that is, as having a head camera size that is greater than .33 of the total size of all three objects combined). Figure 3 shows a histogram of the proportion of all frames with objects in view in which the dominating object dominates the other objects by varying degrees (beginning at .33 when three objects are all in view and roughly the same size). Several aspects of these results are noteworthy. First, a dominating object constitutes 100% of the size of the in-view objects when it

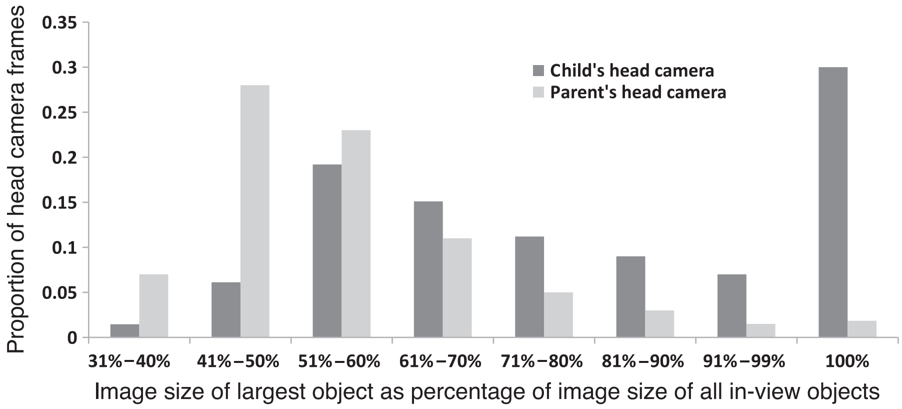

Figure 3 *Histogram showing the magnitude of domination of the largest object in the head camera image: The proportion of head camera frames (with at least one object in the frame) in which the largest object dominated the other objects in the image by varying degrees (from 33% of the size of all in-view objects to 100% of the size of all in view objects).*

is the only object in view. This occurs quite frequently for the children (on 30% of all frames with at least one object in view), but infrequently for adults. Because all three objects are on the table and potentially in view, this significant reduction of information in the head camera view can only occur by one object being much closer to the sensors than the others (see Figure 4). Also noteworthy is the fact that the image size of the dominating object for children is virtually always greater than 50% of the total size of objects in the head camera view. This could be due to there being three objects in view with one much larger than the other two or two objects in view with one larger (to varying degrees) than the other. Either way constitutes a selection, and possibly therefore, more focused attention. Overall, the distribution of head camera sizes of the dominating object in Figure 3 provide converging support for the conclusion that children's views are highly selective, approximating a one-object-at-a-time form of attention.

The above analyses of the dominating object sizes all derive from measures of the *relative* size of objects in view. Given a 70^{o} field, objects that take up 3% of the image are roughly comparable to the size of the fovea and any size greater than 10% of the head camera field is a substantial object in the visual field. To help the reader understand the significance of these absolute image size measures, Figure 4 shows head camera images with different object size properties. Figure 5 shows the histogram of total object sizes in child and parent head camera images. As can be seen, over 60% of object sizes in the parent head camera images are less than 10% of the head camera field; in contrast, over 60% of the total object sizes in the child head camera field are *greater* than 10%. This fact combined with the measures of the dominating object indicate two very different views of the same table top events for parents and their children. For the parents, not only are all three objects often in view and of roughly equal size, they are small in the visual field. For the toddlers, fewer objects are in view at any one time and one object often dominates by being close to the head and therefore large in the visual field. Indeed, on 28% of all child head camera frames with at least one object in view, the largest object *by itself* takes up more than 10% of the head camera image.

All together, these results on dominating objects make clear that the pattern so evident in Figure 2 characterizes the head camera images of the participating parents and children more generally. That is, the child's view of the table top events in joint play are highly selective and centered on one object that is close to the sensors; the parent's view is broad and distant and encompasses all objects.

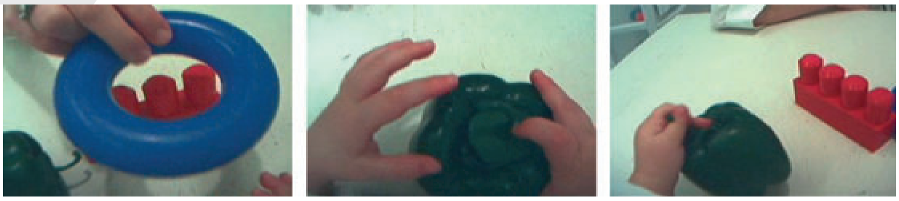

Figure 4 *Three images from the child head camera in which the image size of the largest object was 37% of the total image, 23% of the total image, and 14% of the total image. Figure 5 Histogram of absolute image sizes of all object images (combined): Proportion of all frames in which the total size of all objects (and object parts) in view took up from 0 to 100% of the head camera image for child and parent head camera images.*

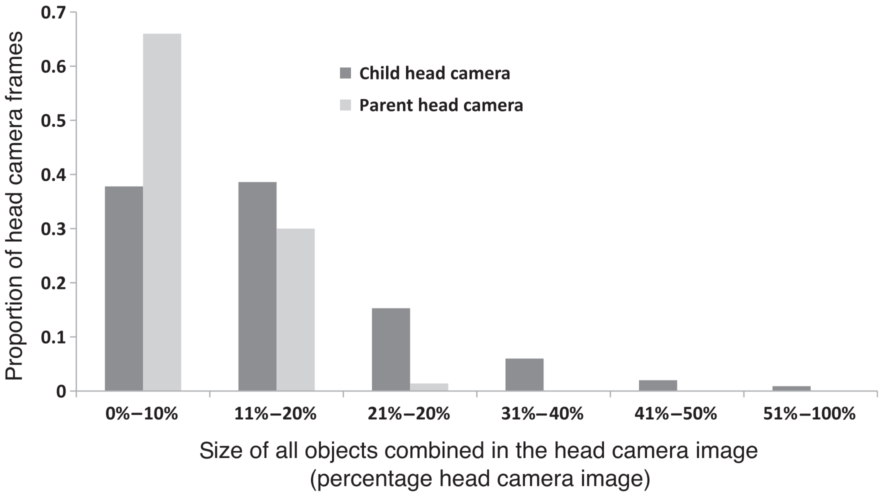

Figure 5 *Histogram of absolute image sizes of all object images (combined): Proportion of all frames in which the total size of all objects (and object parts) in view took up from 0 to 100% of the head camera image for child and parent head camera images.*

Changes in views

Because the head camera view is tightly tied – moment by moment – to the child's own actions, the dominating object will also change as the child moves and moves objects. Accordingly, we calculated the number of times that the dominating object changed, that is shifted from one toy dominating to a different toy dominating. For this measure, the dominating object was defined as an object at least twice the size of the other in-view objects. By this measure, there are on average 19.2 switches (*SD* = 2.5) in the dominating object per minute in the child view but only 7.8 switches per minute in the adult view (*SD* = 2.31), $t(18) = 4.39$, $p < .002$). For those frames containing at least one object, on average 20.2% of object pixels were changing frame by frame in the child's view but only 5% were changing in the adult's view. This difference is expected both by larger body movements and the closeness of the objects to the head camera (and sensors) for the toddlers than the parents. In brief, the toddler view is characterized by a dynamically varying dominant object, with the location of that object in view and the particular object that is dominating changing frequently.

Hands and the social partner

The changes in the image sizes of objects in the child's head camera view may be caused by the child's head rotation, the child's holding and moving of an object, or the parent's holding and moving of an object. The results thus far implicate the second two kinds of actions – hand actions – as the most likely major source of visual selection in the present study. This is because head movements in general (though not always) will increase or decrease the size in the head camera image of all the objects on the table. But hand movements literally can select one object to bring close to head and eyes. As shown in Figure 4, both parent action and child action may contribute to the character of the toddlers' dynamic views; however, the results also suggest that the child's own hand actions may be the most critical. Specifically, over all frames parents were manually interacting with an object on .64 of the frames and children were on .59 of the frames, which indicates that the parents as well as the children were engaged with these objects (*ns*, $t(18) = 1.08$, $p > .31$). However, if we consider only those frames in which there was an *object in the in the child's head camera view*, then children were holding at least one object in 72% of all frames whereas parents were holding an object in 49% of the frames ($t(18) = 7.39$, $p < .001$). Finally, if we consider only those frames with a dominant object in the child's view (defined as twice the size of the sum of all other objects in the view), the dominating object was in the child's hands 54% of the time and in the parent's hand 23% of the time ($t(18) = 6.07$, $p < .001$). On the remaining 24% of the frames, the dominating object was sitting on the table close to the child. Together, these results suggest that the child's own hand actions play an important role in selecting the information available to the visual system.

These data, however, do not distinguish whether the parent in some way instigated the selected object – by pushing it forward, by handing it to the child, or by pointing to or naming it and thereby perhaps starting the cascade that leads the child to bring it closer to the head and eyes. These results, of course, also do not mean that *only* hand actions are important (as compared to head and whole-body movements or to shifts in eye gaze) but they do show that self-generated hand actions play a critical role in toddler visual attention, a role that has not been well studied in the past.

Discussion

Everyday learning contexts are highly cluttered, with many objects and many potential targets for attention and for learning. Theorists of natural and artificial intelligence have often noted the daunting demands of attention in the 'wild' (e.g. Breazeal & Scassellati, 1999). The present findings suggest that in a complex context such as toy play with a partner, the toddler's first-person view of the events is highly selective, indeed, often centered on one object at time. This object is closer to the sensors than others and thus bigger in the visual field, and this is often (though not always) due the child's holding the object close to the body. Although the present analyses just demonstrate this fact, it may prove to be crucial for understanding toddler learning and attention in everyday cluttered contexts.

Before considering the implications of the present observations, we consider possible causes for the dynamic structure of the toddler's first-person views in this task. The properties of the toddler's view most likely derive from body size, movement patterns, and interest in the objects. Toddlers have short arms (on average 23 cm for the child subjects versus 50 cm for the parents, shoulder to wrist). This leads naturally to a constrained visuo-motor workspace that is located nearer the body for toddlers than for adults (see Newell, 1991) and thus to a visual geometry in which objects that are held are close and likely to at least partially block the view of other objects. Further, motor behavior by toddlers is highly synergistic, often involving the whole body (Thelen & Smith, 1994) and therefore may often result in large changes in the relation of the sensors to the objects in the scene. Finally, just about everything is somewhat novel and interesting to an 18-month-old and thus worthy of close manual and visual exploration. Thus the causes behind the observed toddler visual dynamics may be mundane. However, this does not mitigate their theoretical importance. Factors such arm length, synergistic large movements, and curiosity are stable organizing principles for toddler experience. And the visual dynamics they help create are the very data on which learning – and real-time attention and social engagement – must depend.

Implications

A small and near visuo-motor workspace sets up a context in which manual engagement naturally leads to one object dominating the view by being close to the sensors and thus blocking the view of other objects. This is a cheap but effective solution to visual selection that has been used successfully in robotics research (e.g. Ballard, 1991; Metta & Fitzpatrick, 2003; Fitzpatrick & Metta, 2003; Lungarella, Metta, Pfeifer & Sandini, 2003). From the perspective of that research, the one-dominating-object-at-a time dynamics is likely to be a good thing, one that would aid learning about objects. In particular, the robotics research indicates (see especially, Fitzpatrick & Metta, 2003; Metta & Fitzpatrick, 2003) how holding and moving objects naturally segment the object of focus from other objects in the scene, minimize competition from potential competitors by making the selected object larger in the visual field, and create multimodal loops of perception and action that stabilize attention and may also play a role in binding object properties together. These ideas lead to testable predictions for future work. Following this line of reasoning, for example, the optimal moment for naming objects for toddler learning might be when the toddler is not just looking at the intended referent but is also holding it.

The one-dominating-object-at-a-time dynamics of the toddler views also raise new challenges to understanding joint attention. As is apparent in Figure 2, the dynamics of the parent and child head camera images are fundamentally different. The dynamics for the children imply selection based on the closeness to the sensors of the object that is attended to. Attended objects are close and big in the visual image. Competitors for attention are small or out of view. The dynamics for the parents, as measured by the head camera, are stable such that all objects are equally distant and continually in view. Although we did not measure eye gaze in this study, the adult participants are likely to have visually selected objects by shifting eye gaze to bring the selected object to the fovea, thereby recentering the visual field around that selected object. For adults, then, attended objects are small in the visual image but are centered. In this attentional system, competitors *and* potential next targets are always in view.

Here then is the explanatory challenge: Considerable research shows that parents and toddlers do successfully coordinate attention and, moreover, that parent actions are a strong force guiding and scaffolding toddler attention and learning (e.g. Liebal, Behne, Carpenter & Tomasello, 2009; Tomasello, 2008; Pereira, Smith & Yu, 2008). But the present results suggest (at least in complex active tasks) that parent and toddler attentional systems are based on fundamentally different principles of selection. There are also many demonstrations in simple laboratory tasks showing that toddlers use the direction of eye gaze of the mature partner to focus attention on an object (for review, see Poulin-Dubois, Demke & Olineck, 2007). But it is unclear how useful eye gaze tracking by the infant can be in the noisier and more dynamic settings of everyday and cluttered tasks (e.g. Kaplan & Hafner, 2006; Brand & Shallcross, 2008; Pereira *et al*., 2008). How, then, are the apparently very different attentional systems of the toddler and parent coordinated in joint and active contexts such as toy-play? Emerging research suggests that the answer may lie in whole-body movements – including rhythms of posture, head, and hand movements movements (Shockley, Santana & Fowler, 2003; Pereira *et al*., 2008).

Limitations

One contribution of the present approach is the use of the head camera which provides information about toddlers' experience that is profoundly different from a third-person camera, which is the standard approach used in child development research. The difference between a head camera and a third-person camera is that the first-person camera captures the momentary dynamics of available visual information *as it depends* on the child's own actions. The limitation, however, is that not all actions influence the head camera view; in particular, the head camera moves with head movements, not eye movements. Thus, the head camera is *not* a substitute for direct measures of eye gaze direction (see Yoshida & Smith, 2008; Aslin, 2008, 2009) but instead provides information about the dynamics of *available* visual information with larger body movements. Ideally, one would jointly measure both the dynamics of the larger visual field (as given by a head camera) and also focal attention as indicated by eye gaze direction within that field. Prior calibration studies (Yoshida & Smith, 2008) with the head camera tell us that head and eye movements are highly coordinated, but even this coordination leaves open the possibility of transitory and very brief glances to the mother without head movements that may, nonetheless, play an important role in coordinating the social interaction. Because the automatic coding does not clearly distinguish faces from hands, the present analyses do not provide sufficient detail on the role of glances to the face by either participant.

A second limitation concerns the definition of the dominant object in the head camera image. An object that is very large in the visual field – that the child has brought close to their own face – has considerable face-validity as the object being attended to. However, given that there has been no prior work in this area, it is unclear just how big an object needs to be in a head camera field to count as dominating attention. A next step needed to validate this approach is to link the dominating object, as measured here, to some other behavioral outcome related to attention, for example, to learning about the object or its name or to the ease of distraction by some other salient object in the periphery. Related to this limitation is the size of the head camera field itself; at 70° it is considerably smaller that the full visual field (about 180° for toddlers) and thus does not provide a full measure of the potential peripheral influences on the toddler's attention (influences that may include faces and hands in the periphery). Capturing a broader view (via a wider lens head camera) and measuring the size of the effective attentional field will be critical to understanding attention shifting as it likely to be events in the periphery that instigate head and hand movements.

A final limitation concerns the task itself. Parents were asked to actively engage their children with the toys and were instructed to play with three toys at a time, keeping them on the tabletop. Although this is *one* common context in the life of toddlers, it is also one in which the child is sitting and thus larger body movements are constrained. Further, the room is designed (all white) so that the most interesting events are on the table top, perhaps leading to more on-task attention than would be observed in a freely moving toddler in the more cluttered environments of everyday life which must provide much more competition for attention. Future work is needed that examines the dynamics of toddler attention in broader contexts. Some contexts that might be particularly revealing include contexts with a greater number of competitors for attention, contexts in which the child is playing alone, and contexts in which the parent is explicitly guiding attention as in object name teaching tasks. Nonetheless, even in the present constrained tabletop task with just three objects and a parent engaged in joint play, the dynamics of the child's visual experience are dramatically different from those of adults. It will be important to understanding toddler learning to know how different contexts – and different actions by social partners – emphasize or minimize these differences.

Conclusion

There is much that we do not know about the dynamic patterns that comprise sensory-motor experience, including the across-task generality of the patterns observed here, the timing and nature of developmental change in these patterns, and whether there are task contexts in which adults might behave – and generate visual experiences – like those of toddlers. However, the present results make clear the insights that may emerge from addressing these questions head on, that is by trying to capture the contents of children's first-person visual experiences as a function of their body movements and actions. The dynamic properties of toddler active vision are most certainly relevant to the mechanisms of real-time attention in cluttered fields and to real-time learning. The present results suggest that the small visual-motor work-space of the toddler may actually create an advantage to learning by creating a dynamic structure in which manual engagement naturally leads to one object dominating the view by being close to the sensors and thus blocking the view of other objects. This is a cheap but effective solution to visual selection that may bootstrap processes of object segregation, integration of multiple object views, and the stabilization of attention.

Acknowledgements

This research was supported by NIH R21 EY017843 to the first author, NSF BCS0544995 to the second author and by a PhD scholarship from the Gulbenkian Foundation to the third author. We thank Charlotte Wozniak for help in the data collection, Hanako Yoshida and Bill Freeman for invention of the head camera, and Andrew Filopowicz for assistance with the figures.

References

Aslin, R.N. (2008). Headed in the right direction: a commentary on Yoshida and Smith. *Infancy*, **13** (3), 275–278.

Aslin, R.N. (2009). How infants view natural scenes gathered from a head mounted camera. *Optometry and Vision Science*, **86**, 561–565.

Ballard, D (1991). Animate vision. *Artificial Intelligence*, **48**, 57–86.

Bertenthal, B.I., Campos, J.J., & Kermoian, R. (1994). An epigenetic perspective on the development of self-produced locomotion and its consequences. *Current Directions in Psychological Science*, **3** (5), 140–145.

Bojczyk, K. E., & Corbetta, D. (2004). Object retrieval in the 1st year of life: Learning effects of task exposure and box transparency. *Developmental Psychology*, **40** (1), 54–66.

Brand, R.J., & Shallcross, W.L. (2008). Infants prefer motionese to adult-directed action. *Developmental Science*, **11** (6), 853–861.

Breazeal, C., & Scassellati, B. (1999). *A context-dependent attention system for a social robot*. Proceedings of the Sixteenth International Joint Conference on Artifical Intelligence (IJCAI 99). Stockholm, Sweden, 1146–1151.

Carey, S. (2009). *The origin of concepts*. New York: Oxford University Press.

Fitzpatrick, P., & Metta, G. (2003). Grounding vision through experimental manipulation. *Philosophical Transactions of the Royal Society of London A*, **361**, 2165–2185.

James, K.H. (in press). Sensori-motor experience leads to changes in visual processing in the developing brain. *Developmental Science*.

Kaplan, F., & Hafner, V. (2006). The challenges of joint attention. *Interaction Studies*, **7**, 135–169.

Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Developmental Science*, **12** (2), 264–271.

Lungarella, M., Metta, G., Pfeifer, R., & Sandini, G. (2003). Developmental robotics: a survey. *Connection Science*, **4** (15), 151–190.

Metta, G., & Fitzpatrick, P. (2003). Better vision through manipulation. *Adaptive Behavior*, **11**, 109–128.

Needham, A., Barrett, T., & Peterman, K. (2002). A pick me up for infants' exploratory skills: early simulated experiences reaching for objects using 'sticky' mittens enhances young infants' object exploration skills. *Infant Behavior and Development*, **25** (3), 279–295.

Newell, K.M. (1991). Motor skill acquisition. *Annual Review of Psychology*, **42**, 213–237.

Pereira, A., Smith, L., & Yu, C. (2008). Social coordination in toddler's word learning: interacting systems of perception and action. *Connection Science*, **20**, 73–89.

Poulin-Dubois, D., Demke, T.L., & Olineck, K.M. (2007). The inquisitive eye: infants' implicit understanding that looking leads to knowing. In R. Flom, K. Lee & D. Muir (Eds.), *Gaze-following: Its development and significance* (pp. 263–281). Mahwah, NJ: Lawrence Erlbaum Associates.

Shockley, K., Santana, M., & Fowler, C.A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, **29** (2), 326–332.

Smith, L.B., & Gasser, M. (2005). The development of embodied cognition: six lessons from babies. *Artificial Life*, **11** (1–2), 13–29.

Soska, K.C., Adolph, K.E., & Johnson, S.P. (in press). Systems in development: motor skill acquisition facilitates three-dimensional object completion. *Developmental Psychology*.

Thelen, E., & Smith, L.B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Tomasello, M. (2007). Cooperation and communication in the 2nd year of life. *Child Development Perspectives*, **1** (1), 8–12.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Xu, F., & Tenenbaum, J.B. (2007). Word learning as Bayesian inference. *Psychological Review*, **114** (2), 245–272.

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

Received: 25 March 2009
Accepted: 2 October 2009

Author Query Form

Journal: DESC

Article: 947

Dear Author,
During the copy-editing of your paper, the following queries arose. Please respond to these by marking up your proofs with the necessary changes/additions. Please write your answers on the query sheet if there is insufficient space on the page proofs. Please write clearly and follow the conventions shown on the attached corrections sheet. If returning the proof by fax do not write too close to the paper's edge. Please remember that illegible mark-ups may delay publication.

Many thanks for your assistance.

Query reference	**Query**	**Remarks**
Q1	**AUTHOR: Figure 2 has been saved at a low resolution of 256 dpi. Please resupply at 600 dpi. Check required artwork specifications at http://www.blackwellpublishing.com/authors/digill.asp**	

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	• • • under matter to remain	✓
Insert in text the matter indicated in the margin	⋏	New matter followed by ⋏ or ⋏ ⓧ2
Delete	/ through single character, rule or underline or ⊢—⊣ through all characters to be deleted	δ or δ ⓧ2
Substitute character or substitute part of one or more word(s)	/ through letter or ⊢—⊣ through characters	new character / or new characters /
Change to italics	— under matter to be changed	⎵
Change to capitals	≡ under matter to be changed	≡
Change to small capitals	= under matter to be changed	=
Change to bold type	∼ under matter to be changed	∼
Change to bold italic	≂ under matter to be changed	≂
Change to lower case	Encircle matter to be changed	‡
Change italic to upright type	(As above)	⊥
Change bold to non-bold type	(As above)	∼∤∼
Insert 'superior' character	/ through character or ⋏ where required	Y or ⋏ under character e.g. $\overset{2}{Y}$ or $\overset{2}{\curlywedge}$
Insert 'inferior' character	(As above)	⋏ over character e.g. $\underset{2}{\curlywedge}$
Insert full stop	(As above)	⊙
Insert comma	(As above)	,
Insert single quotation marks	(As above)	$\overset{‘}{Y}$ or $\overset{‘}{\curlywedge}$ and/or $\overset{’}{Y}$ or $\overset{’}{\curlywedge}$
Insert double quotation marks	(As above)	$\overset{“}{Y}$ or $\overset{“}{\curlywedge}$ and/or $\overset{”}{Y}$ or $\overset{”}{\curlywedge}$
Insert hyphen	(As above)	⊢-⊣
Start new paragraph	_⌐	_⌐
No new paragraph	⤺	⤺
Transpose	⊔⊓	⊔⊓
Close up	linking ⁀‿ characters	⁀‿
Insert or substitute space between characters or words	/ through character or ⋏ where required	Y
Reduce space between characters or words	| between characters or words affected	⊤

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	• • • under matter to remain	✓⃝
Insert in text the matter indicated in the margin	⋏	New matter followed by ⋏ or ⋏ⓧ2
Delete	/ through single character, rule or underline or ⊢—⊣ through all characters to be deleted	ꝺ/ or ꝺ/ⓧ2
Substitute character or substitute part of one or more word(s)	/ through letter or ⊢—⊣ through characters	new character / or new characters /
Change to italics	— under matter to be changed	⌙
Change to capitals	≡ under matter to be changed	≡
Change to small capitals	= under matter to be changed	=
Change to bold type	∼ under matter to be changed	∼
Change to bold italic	≂ under matter to be changed	≂
Change to lower case	Encircle matter to be changed	≢
Change italic to upright type	(As above)	┼
Change bold to non-bold type	(As above)	∼┼∼
Insert ‘superior’ character	/ through character or ⋏ where required	⋎ or ⋏ under character e.g. $\overset{2}{\vee}$ or $\overset{2}{\wedge}$
Insert ‘inferior’ character	(As above)	⋏ over character e.g. $\underset{2}{\wedge}$
Insert full stop	(As above)	⊙
Insert comma	(As above)	,
Insert single quotation marks	(As above)	$\overset{‘}{\vee}$ or $\overset{‘}{\wedge}$ and/or $\overset{’}{\vee}$ or $\overset{’}{\wedge}$
Insert double quotation marks	(As above)	$\overset{“}{\vee}$ or $\overset{“}{\wedge}$ and/or $\overset{”}{\vee}$ or $\overset{”}{\wedge}$
Insert hyphen	(As above)	⊢-⊣
Start new paragraph	_┌	_┌
No new paragraph	↩	↩
Transpose	⊔⊓	⊔⊓
Close up	linking ◠◡ characters	◠◡
Insert or substitute space between characters or words	/ through character or ⋏ where required	Y
Reduce space between characters or words	| between characters or words affected	⊤