Cognition, Brain, Behavior. An Interdisciplinary Journal

ISSN: 1224-8398
Volume XV, No. 4 (December), 535-552

A NEW PERSPECTIVE ON EMBODIED SOCIAL ATTENTION

Hanako YOSHIDA, Joseph M. BURLING *

Department of Psychology, University of Huston, Texas, USA

ABSTRACT

Over the years observational studies have made great progress in characterizing children's visual experiences and their sensitivity to social cues and their role in language development. Recent technological advancements have allowed researchers to study these issues from the child's perspective, leading to a new understanding of the dynamic involvement of bodily events. A number of recent studies have suggested that bodily actions play an important role in perception and that social partners' bodily actions may become synchronized. In the present perspective paper, we will provide a new perspective on how children's own views are generated individually and play a dynamic role in learning. By doing so, we first discuss the role of early social input in language learning as it has been treated in the literature and then introduce recent studies in which typically developing hearing children, deaf children of deaf families, and children with autism were observed in a social context using the new child-centered technology. The hypothesis of a link between sensorimotor experiences and embodied attention — specifically how different bodies produce different kinds of attention — will be discussed. Understanding the role of bodily events (the child's and the child's social partners') in early visual experiences will provide insight into the development of learning mechanisms and the processes involved in learning disabilities.

KEYWORDS: *embodied attention, social interaction, child-centered perspective*

* Corresponding author:
E-mail: yoshida@uh.edu

Children learn about their world — about objects, actions, other social beings, and language — through social interactions. Social interactions are coordinated by a suite of bodily actions including gestures, head and eye movements, body sway, and posture (Langton, 2000; Shockley, Baker, Richardson, & Fowler, 2007; Shockley, Richardson, & Dale, 2009). A number of researchers consider the body to play a significant role in our experience of the world (Barsalou, Niedenthal, Barbey, & Ruppert, 2003; Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005). Developmental researchers have long studied the role of social input by careful observations of what is happening in the learning environment (e.g., Bakeman & Adamson, 1984; Bates, 1979; Fogel, Lyra, & Valsiner, 1997; Iverson et al., 1994; Iverson, 2010; Toda & Fogel, 1993; Toda, Fogel, & Kawai, 1990). Scientific progress has also been made through laboratory studies of children's use of social cues, particularly in the domain of language learning. These studies show that children are highly sensitive to a variety of cues and that these social cues appear instrumental to early word learning (Baldwin, 1991; Bloom, 2000; Tomasello, 1995).

All this previous work considered social interactions from the perspective of an observer who coded what each participant was doing: the signals the child sent, the adult's responses, and the child's responses to the adult behavior. Social actions were seen as forming a kind of language of messages that are sent and received. But every behavior — every pointing of a finger, every shift in eye gaze — has two parts: its direct effect on the sensory system, and the social message such a signal might send. For example, shifting one's direction of gaze changes what one sees and sends a signal about the object being attended to. A gesture by one's social partner serves as a communicative signal. Such gesture use has a direct effect on the sensory system and input. The aim of this perspective paper is to introduce a new perspective about the roles of children's own views in learning in light of the embodiment idea; specifically, we will emphasize the role of new technologies and recent developmental attempts. By doing so, we consider the direct sensory effects of bodily actions in a social context — focusing specifically on how a child's actions and the parent's actions are dynamically linked and directly influence visual input and, thus, potentially learning.

KNOWING WHERE THE EYES ARE IS NOT ENOUGH

In the long history of developmental science, the most common view used in observations is a "room view", a third-person view from a corner or from another room via a one-way mirror. This has certainly been sufficient when a controlled environment is needed in which a solitary child is given a single item to look at; however, real learning environments, where young children learn social components and language, are rarely so simple. Indeed, their everyday learning environments are very chaotic and often composed of many objects and actions. Also, learning environments are dynamic. For example, children and their social partners constantly change the elements of the environment (i.e., objects and actions) and their own locations within it. This makes the environment even more complex, and yet all these elements are relevant for learning, especially in social contexts. Constantly changing coordinated eye gaze, language use, and body posture function to facilitate communication and mutual understanding (Pickering & Garrod, 2004; Shockley et al., 2009). Thus, to be sensitive to these constant changes in visual experiences, taking a child-centered view is essential. The third-person view typically cannot distinguish which parts of a complex scene a child might be attending.

Another issue affecting the choice of perspective is that adult coders are influenced by where they think the child should be looking. If a child is looking at the wall behind his or her social partner when the partner says "oh, no" it is likely that coders will score the look as if the child had looked at the partner, not at the wall. This is due to our strong tendency to use intuition to describe what should be happening. This type of coding issue and intuition bias can be particularly critical in social domains where many social contacts and complex emotional exchanges might be present to reinforce such overinterpretation. It is a particular problem when measuring social development. Infants look back and forth between their social partner's face (and gestures) and an object. Indeed, looking between a social partner's gaze and potential referents has been thought to supply information about the partner's gaze direction and affect (Butterworth, 1995; Tomasello, 1995). Thus, coding requires precision. Observations containing social contexts benefit most from the correction of these intuition biases. Acquiring an unbiased and more precise coding of what the child is looking at can be essential for this type of measure.

Recent technological developments have made it possible to capture input from the viewer's perspective using eye-tracking equipment. Yet until very recently, with the advent of head-mounted eye trackers (which have not been used much in infant–parent social settings), eye tracking has typically been used with stationary babies watching a TV screen, and thus researchers were unable to capture the role the whole body plays in social behavior or the relation between the child's own actions and what the child sees (but see Franchak, Kretch, Soska, & Adolph, in press, for a child-centered perspective in a self-navigation context).

Two recent eye-tracking studies that recorded social scenes from infants' perspectives successfully documented natural tendencies in visual exploration when the videotaped scenes were shown to different groups of young infants and adults (Aslin, 2009; Frank, Vul, & Johnson, 2009). The studies reported substantial rates of looking at faces as well as age-related differences in fixation. Both studies also noted interesting attention shifts from face to hands when the hands were performing actions, indicating that social attention is sensitive to the content of actions. These studies successfully captured what children look at when scenes are presented in two dimensions. Observational precision (i.e., how precisely one can detect where an infant is looking at any given moment) will continue to improve as we persist in making technological advancements.

However, although these studies made use of eye-tracking technology, they did not address the idea of embodied attention because they studied infants looking at videos produced by equipment on other infants' bodies. Embodied attention means that what is attended to in the learning environment is partly influenced by body structure. Body structure in this case refers to the child's own physical constraints and actions, such as eye gaze, head movements, posture shifts, grabbing or holding, and the same actions performed by another body — a social partner. What this means is that our perceptions (moment-to-moment visual experiences) may not be sufficiently understood in isolation from our own body and movements. Children "select" their visual input both by their own actions and by how their actions influence the behaviors of others. This core idea of there being a meaningful link between bodily experiences and perceptual experiences has received much attention (Barsalou, Pecher, Zeelenberg, Simmons, & Hamann, 2005; Bertenthal, Campos, & Barrett; 1984; Martin & Chao, 2001), and recent behavioral evidence has further suggested the importance of sensorimotor coordination (e.g., seeing, touching, manipulating objects) in children and the dynamic linkage to the learning environment for our understanding of visual selection and social input (Adolph & Berger, 2006; Needham, Barrett, & Peterman, 2002; Pereira, Smith, & Yu, 2008; Thelen & Smith,1994; see Sheya & Smith, 2010, for a review).

New methods have been developed to explore this dynamic loop between social attention from the child's point of view and the child's own moment-to-moment interactions. One study explored early visual behavior in a naturalistic free-play context where 14-month-old infants' eye gaze was recorded using a head-mounted eye-tracking device (Franchak et al., in press). This study demonstrated that infants' selection of their own visual field depended heavily on their own actions, and they seldom looked at faces (in this case, their mother's face). Further, the study revealed a new relation between infants looking at their own hands and their own actions while engaged in manual actions, and a direct impact of the social partner's movement on the child's view.

HEAD CAMERAS AND THE HEAD-CENTERED VIEW

Efforts have been made to document children's own natural visual exploration while they participate in moment-to-moment social interactions. These studies have used relatively less complex technology yet are still suitable for addressing embodied social attention. Typically, a small head camera (about the size of a quarter) is placed on the child as the child and parent play with a toy. The camera captures visual events in which the children themselves are participating. The children are creating their own visual input through social exchange, and in contrast to the eye-tracking methods described above, this is captured from the child's perspective (often a child-mother play session). This perspective is taken from the forehead (as opposed to tracking eye gaze) and thus is a relatively less complicated form of data collection that does not interfere with natural viewing and their interactions.

One major concern about using the head-mounted camera has been that it is positioned at some distance from the eyes, and thus it may be difficult to identify the focal point of the child's attention with precision. Still, several researchers have found a strong head–eye coupling in infants (e.g., Daniel & Lee, 1990; von Hofsten & Roseander, 1996; Savelsbergh, von Hofstein, & Jonsson, 1997; Smith, Thelen, Titzer, & McLin, 1999), and more recently Yoshida and Smith (2008) documented how the head view can secure sufficient measurement of where children attend to in common social contexts. This study assessed 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 (e.g., horizontal movement, reaching for objects). The study verified the sufficient overlap between the head camera images and the child's visual field, suggesting the effectiveness of the head camera in the interactive task context, which simulates a natural learning environment. Moreover, in social settings in which people are continually moving their head and eyes, eyes alone are unlikely to be a good cue.

Although studies have suggested early sensitivity to eye gaze and its relevance to language learning (e.g., Brooks & Meltzoff, 2008), little work has been done on how temporal dynamics of eye-gaze following can emerge in complex social contexts where heads, eyes, and bodies are constantly engaging in actions. Indeed, most experiments on the following of eye gaze, in infants and adults, manipulated eye-gaze direction in a fixed, straight-on face (see Langton, Watt, & Bruce, 2000, for a review). Thus these studies have not accounted for the natural social contexts in which heads and eyes move both together and independently (Einhäuser et al., 2007).

Additional support for head–eye coupling comes from studies showing that gaze and head orientation independently influence decisions concerning social attention. Adults, children, and infants demonstrated difficulty ignoring the direction of the head when judging eye-gaze direction (Doherty & Anderson, 1999, 2001; Doherty, Anderson, & Howieson, 2009; Langton, 2000, Loomis, Kelly, Pusch, Bailenson, & Beall, 2008; Moore, 2006; Moore & Corkum, 1998). Accordingly, the head-camera method seems to be sufficient for studying children's perspective while engaging in social actions, which occurs in most natural learning environments. Recent results support the importance of taking a head-camera view when evaluating children's own creation of visual input.

A group of researchers (Pereira & Smith, 2009; Pereira et al., 2008; Pereira, Yu, Smith, & Shen, 2009; Smith, Yu, & Pereira, 2011; Yoshida & Smith, 2008) explored children's proximate visual behavior in the context of naturalistic play with the mother playing and teaching a set of words to her child. The typical procedure used with studies of this kind attaches a small head camera to the child's forehead with an elastic band while the child engages in toy play with a parent. The goal is to capture visual events — in which the child creates his or her own moment-to-moment social exchange — from the child's perspective. Each study has contributed to the understanding of what components are most likely to be available to the child and how they might relate to social events, and how that might depend on the individual's learning experiences.

These studies focused on exploring the visual field of toddlers (ranging from 12 to 24 months old) by using a variant of the head-camera method, documenting several novel and enlightening results concerning the dynamics of toddlers' first-person view. One study (Yoshida & Smith, 2008) noted that the room view and the child view (head-camera view) are very different (see Figure 1 for synchronized view frames from Yoshida & Smith, 2008). Indeed, it is relatively clear from the head-camera view what is available to the child's attention at any given moment, but not when observed from parents' views (also via head cameras). Children's' unique behavior of bringing objects closer to their eyes can reduce the information available in their view. This is a potentially significant constraint in early learning.

Figure 1.
Left: third-person view; right: first-person view.

Furthermore, the closer view of an object is also indicative of the relation between a child's view and the child's body structure — a preference for closer views is related to a child's shorter arms and the location of arms relative to the head. Yoshida and Smith's (2008) and Smith et al.'s (2011) results also suggest a strong tendency to look more at objects that are in hands — their own and their parents' — and to shift attention to new objects as parents' hands move in that direction. These results using a different methodology appear to contrast with earlier findings of a preference for faces (e.g., Aslin, 2009; Frank et al., 2009), highlighting the potential role of parents' action in organizing early attention. These studies indicate how early attention is dynamically linked to a child's own body and actions and is embedded in a social partner's action.

DIFFERENT BODIES, DIFFERENT SOCIAL ATTENTION?

Studies that captured the visual content generated by an individual child have emphasized the importance of considering embodied attention to fully understand children's visual selection and social input. The seemingly tight loop connecting embodied attention, an individual's own actions in a social environment, and social partners' actions suggests that such embodied attention might differ when it is generated by children whose sensorimotor experiences and social interaction are different and uniquely integrated.

Our current attempts to understand embodied attention — specifically feedback from sensorimotor, language, and social experiences — have focused on social contexts that mimic word-learning environments that diverge from those of the typically developing, hearing child (Yoshida, 2010; Yoshida & Kushalnagar, 2009). We have concentrated on three groups in an effort to explore the dynamic structure of embodied attention: typically developing hearing children, typically developing deaf children of deaf families, and hearing children with Autism spectrum disorder (ASD). The sensorimotor, language, and social experiences differ among these children both quantitatively and qualitatively.

In one study (Yoshida & Kushalnagar, 2009), typically developing deaf children aged 4 to 5 years were compared with typically developing hearing children of the same age range in an extension of the parent–child interaction sessions used in earlier work by Yoshida and Smith (2008). Parents were instructed to interact naturally with their child with the goal of teaching a series of words by demonstrating the meanings with toys and actions. Head-mounted cameras were attached to the young participants in order to provide a child-centered perspective of the learning environment. The head-mounted camera (Figure 2) is a Watec (WAT-230A) miniature color camera weighing approximately 30g (36×30×15mm). The camera was sewn into a headband that could be placed on the forehead such that the mounted camera is close to the child's eyes. The angle of the camera is adjustable. A second camera — the third-person camera — was set up to record a broad view of the task context that included the tabletop, the child, and the parent. The two cameras were synchronized by an Edriol (model V-4) four-channel video mixer. The head camera moves with the head (not with eye movements). However, a prior calibration study directly comparing independent measures of head and eye direction in typically developing toddlers found a 90% correspondence in direction. After fitting the child with the headband and giving instructions, the experimenters left the testing section and hid behind a curtain to monitor the head-camera view to continue ensuring the camera angle. The entire play session lasted about 12 minutes: parents taught 16 words, each word was taught for 40 seconds.

Figure 2.
Head camera

A total of 6 coders coded the video from the head camera (2 deaf coders for American Sign Language transcribing) and 4 coders for nonlinguistic contents frame-by-frame using MacShapa, an application for observational data analysis. They calculated the amount of looking time for coding categories including the individual toys, the parent's face, the parent's hands, and the child's hands to obtain the proportion of looking per session.

This approach — namely, exploring the social interactions among attention, body, and the environment — is used here to consider how different populations of children (e.g., deaf, hearing) partition their visual scenes and to observe the dynamic processes that shape their word-learning experiences. Parent–child interactions are inevitably influenced by the capabilities of both the child and the parent. For instance, the use of either sign language or verbal speech as a means of communication has a direct impact on what is observed from a child's point of view. Which simultaneously appearing objects are attended to seems to differ between hearing and deaf children, and the proportion of time spent attending to certain objects over others varies, as well. This could be related to differences in bodily communication styles. Communication through speech among hearing individuals frees up the hands of both parent and child so that objects within their immediate environment can be adapted for use in teaching new words. Dyadic interactions involving parents who use sign language to communicate with their children involve more prolonged periods of attention to crucial communicative components, such as the parent's hands or face, and the default location of hands might be higher and closer to the face (Yoshida, 2010; see Figure 3 for synchronized view frames from Yoshida & Kushalnagar, 2009).

Figure 3.
Left: third person view; right: first person view.

The structure of deaf children's embodied communicative acts also constrains their views in that their language (hands) dominates their view, and their sensory experiences (mostly through the visual modality) assign more attentional weight to visual components. Because the deaf children in this study were all being raised by parents who are also deaf, their initial encounter with language perception was likely to be through visual experiences, and their language production is also through the visual system. Their visually organized language use relies on meanings and ideas, and the meanings children learn early are often considered to be concrete and visually tangible at an early stage in general (Gentner, 1982). Accordingly, for deaf children who have access to sign language in the beginning of life, language perception, production, and the mapping of language to meaning are all linked through visual experiences, suggesting early dominance of visual processing. The question is how this visual dominance influences attention to social components, such as a partner and her action. Figure 4 illustrates the overall trend in viewing the parent's hands or face for deaf children and typically developing hearing children (Yoshida & Kushalnagar, 2009). The graph compares the proportion of time spent looking at the parent's hand or face, averaged across 16 teaching trials. Deaf children on average captured these items more frequently than hearing children, especially at the beginning of a word-learning session.

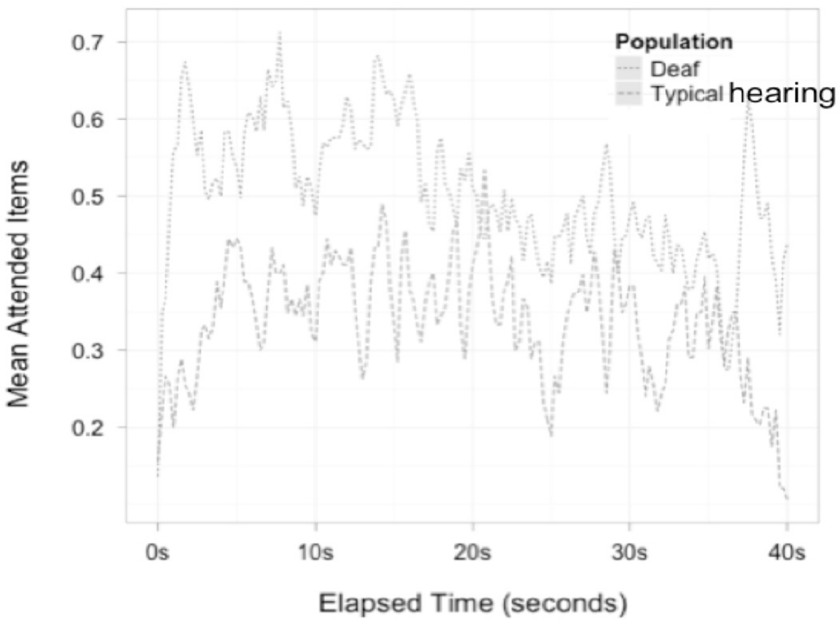

Figure 4.
Proportion of time spent looking at the parent's hands or face averaged across trials.

Allocating attentional resources to hands and faces lessens the possibility of other objects — such as the toys in the word-learning session — being the focus of the child's attention for a prolonged period of time. Attention to the intricacies of hand movements and facial expressions is less likely in hearing children. Hearing children are also less likely to continually switch from hands and faces to other objects, leading to a more local perspective with attention focused on individual objects for prolonged periods of time. Deaf children, on the other hand, are more likely to simultaneously keep track of the dynamics inherent in their visual scene, leading to a much more global perspective (Yoshida & Kushalnagar, 2009).

Despite apparent differences in their language-learning environments, typically developing deaf children often manage to achieve typical language development, and their adaptability may lie in how they carved up their visual environment. This is evident in the results suggesting that the necessary referents for word learning exist for both deaf and hearing populations. However, it is not our intention to proclaim that the developmental trajectories of deaf and hearing individuals are in any way identical. Other factors, such as the timing of the onset of hearing loss, the implantation of electronic cochlear devices, the hearing capabilities of parents with deaf children, or of hearing children with deaf parents, have their own complexities that can give rise to different word-learning environments. Nevertheless, the unique structure of embodied attention generated by typically developing deaf children influenced by their unique bodily experiences (via an alternative language-learning environment including their ASL use) seems to generate relevant visual experiences. This suggests that selectively attending to relevant aspects within their visual environment leads to outcomes similar to those of children without hearing loss, but with different means of reaching those outcomes.

THE CASE OF AUTISM DEVELOPMENT

The dynamic coordination of hand movements, eye gaze, and posture that can be present in individuals is a crucial component of communication (Chartrand & van Baaren, 2009; Shockley et al., 2009). Such coordination seems to differ depending on individuals' bodily interactions with the environment and their sensorimotor experiences. Yoshida (2010) addressed the role of bodily structured attention by exploring embodied attention in children with autism, which represents one case of atypical development. This is a particularly interesting developmental question. Many cognitive developmental disorders involve attention, and there is considerable evidence of comorbidity of these cognitive disorders with early sensorimotor patterns (Hartman, Houwen, Scherder, & Visscher, 2010; Larroqueet al., 2010). For instance, a number of studies have suggested that the function of shared attention (e.g., child and caretaker jointly attending to the same item) develops differently in children with Autism spectrum disorder. Shared attention has often been discussed in terms of its effectiveness in guiding attention in language learning. The development of shared attention thus can be a key element in language development.

In the dynamic social embodiment framework, shared attention can be viewed as an essential developmental process in which children *initiate* a demonstration of their attention (e.g., clear, sustained attention to a toy), which parents can then follow — parents can adjust their topic to their children's moment-to-moment attentional shifts (Adamson & Bakeman, 1991). However, sustaining attention to objects has been considered particularly challenging for young children with autism (Pierce, Muller, Ambrose, Allen, & Courchesne, 2001). From the perspective of social embodiment, the body and what the body does both influence sensory inputs and the social setting. Understanding these links in children with atypical development will lead to recognition of potential developmental consequences.

Yoshida (2010) used a head camera set-up to address how a deficiency in sustained attention influences embodied attention in 4- to 5-year-old children with ASD. Preliminary findings suggest distinct differences between ASD and typically developing children in their focus of attention. The child's eye view revealed that objects used to teach novel and familiar words, when paired with hand gestures and manipulations, were present in the field of view of children with ASD for very little time. Irrelevant objects within their immediate environment occupied a substantial proportion of their visual field. Extended periods of time spent observing nonreferential items such as the floor, ceiling, or a table meant that there was less opportunity during the word-learning session for relevant objects to be introduced into the child's visual field.

In addition to these distinct patterns of looking, observations taken from the head camera revealed a tendency to extend the gaze to the parent's face only, a significant departure compared to typically developing children. This learned behavior may be related to common practice and a form of communicative therapy provided to those with autism to support their development of social interactions. Such learned tendencies to observe the parent's face may lead to overall looking patterns that differ from those of typically developing children. This propensity for not attending to some types of relevant information during the word-learning process could provide insight into the language development of children with ASD.

DISCUSSION

Children's early learning environments contain a number of elements that are dynamically linked in real time as they and their social partners move and act on these elements. Researchers attempting to carefully document the focus of children's attention in complex social settings have suggested that the patterns of children's looking at faces and their social context may be mutually influential, especially when manual actions are involved. However, in recent approaches in social cognition, it has been suggested that it is not possible to fully understand social interactions without addressing how all the bodily interactions (gestures, head and eye movements, body sway, and posture) are coordinated (Langton et al., 2000; Langton, 2000; Shockley, Richardson, & Dale, 2009; Shockley et al., 2007). This promotes the idea that a child's perspective is part of a social dynamic loop.

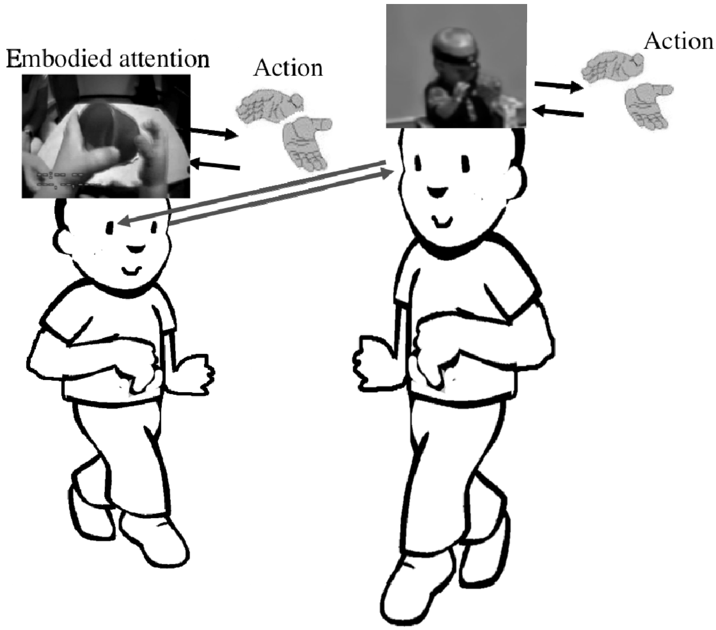

Figure 5.
Perception–action–body loop

We have discussed ideas and empirical work suggesting how different bodily actions (both the child's and the parent's) are dynamically linked and directly influence the moment-to-moment generation of visual input and thus potentially matter for learning. This is far more dynamic and complex than it may seem at first glance, given that different bodies generate different embodied attention based on different body structures, experience different sensory inputs, and perceive different social signals that guide attention. The schema is illustrated in Figure 5. In the figure, a child is attending to an aspect of a scene — in this case, a cup. What is being attended to determines the in-the-moment input (a cup). The hands and body structure of the child play a role in guiding the child's attention, but this tight loop between visual experience and action is dynamically coupled with a mature social partner. This means that the embodied attention is generated through at least two perception–motor loops (here, the child's and the parent's) and that each action changes the sensory input of the child and the parent. What the parent does affects the child's moment-to-moment visual experiences. Thus, it is not sufficient to study social interactions solely from the perspective of an observer, nor to consider a child's perspective that does not take into consideration his or her own body and actions.

The role of bodily cues in social attention (e.g., hand movement, hands acting on objects, eye gaze) appears to be relatively clear, yet recent research findings suggest that such a role may well be subtle and possibly implicit in nature. For instance, in whole-body cues, and coordination itself such as in global body rhythms generate synchronous effect (Shockley, Santana, & Fowler, 2003; Smith et al., 2011; Spivey, Richardson, & Dale, 2009). One line of studies documented a relation between communication effectiveness and body coordination between social partners through body sway and posture (Langton, 2000; Shockley, Richardson & Dale, 2009; Shockley et al., 2007). Another study suggested a dynamic link between interaction quality — including smoothness of turn-taking behaviors and bodily coupling (e.g., of speakers) — and effective word learning in the context of parent and toddler social events (Shockley et al., 2007). Accordingly, particular body movements that are obvious and meaningful may not independently contribute to the creation of social attention. Rather, synchrony of movements and more global body structures as a whole are more likely to participate in the loop. There might be dynamic relational local bodily elements and larger body cues creating attention and possibly influencing learning (Bangerter, 2004; Kita, 2003; Richardson, Dale, & Kirkham, 2007). This sensitivity to synchronous elements (invariants) is consistent with the early emergence of multimodal correspondences in learning (Gogate, Bahrick, & Watson, 2000) and thus may well be the key to our understanding of human learning and communication.

Traditional views of early social input have been refined by technological advancement, and certainly the field has benefited from our ability to observe from their own point of view what children scan in complex visual scenes. The new technology has also helped eliminate coding biases. Use of the technology has led to new findings about early sensitivity to action and the context-dependent nature of attention, and it also has confirmed previous observations about early attentional preferences for human faces and eye gaze for particular task contexts.

This perspective paper focused on the embodied nature of social attention, i.e., how it emerges through interaction between a child and the child's social partner, each dealing with his or her own physical constraints, sensorimotor input, and cognitive capacity. Within this framework, we can expect to see different types of social attention for different groups of individuals whose everyday experiences — through sensorimotor input and cognitive development — are uniquely constrained. Systematic observations of children's proximate visual experiences in social contexts have provided insight into the dynamic relations (Yoshida, 2010; Yoshida & Kushalnagar, 2009; Yoshida & Smith, 2008). Studies of deaf children of deaf parents (Yoshida & Kushalnagar, 2009) revealed sustained attention to face and hands — which has a communicative function — and more attentional shifts to capture a global perspective compared to typically developing hearing children. Observations of hearing children with ASD (Yoshida, 2010) revealed that they pay less attention to objects that are the topic of conversation (e.g., held in hands and acted on) and more attention to nonreferential items. These studies suggest that attentional biases are generated through the dynamic social loops of all participants: perception, action, and body structure. Differences in the dynamics can generate differences in attentional biases and thus in language development.

Taking the child's perspective of the perception–action–body loop is a crucial step in understanding social structure and early visual experiences and their influence on cognition, from language to problem solving. This may lead to further insight into the learning mechanisms and processes involved in learning disabilities.

ACKNOWLEDGMENTS

The preparation of the manuscript and research introduced in the manuscript were supported by the University of Houston, a National Institutes of Health grant (R01 HD058620), and the Foundation for Child Development.

REFERENCES

Adamson, L. B., & Bakeman, R. (1991). The development of shared attention during infancy. *Annals of Child Development*, *8*, 1–41.

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

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

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, *55*, 1278–1289.

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*, 874–890.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*, 415–419.

Barsalou, L. W., Niedenthal, P. M., Barbey, A. K., & Ruppert, J. A. (2003). Social embodiment. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (Vol. 43, pp. 43–92). San Diego, CA: Academic Press.

Barsalou, L. W., Pecher, D., Zeelenberg, R., Simmons, W. K., & Hamann, S. B. (2005). Multimodal simulation in conceptual processing. In W. Ahn, R. L. Goldstone, B. C. Love, A. Markman, & P. Wolff (Eds.), *Categorization inside and outside the lab: Festschrift in honor of Douglas L. Medin* (pp. 249–270). Washington, DC: American Psychological Association.

Bates, E. (1979). *The emergence of symbols: Cognition and communication in infancy*. New York, NY: Academic Press.

Bertenthal, B., Campos, J., & Barrett, K. (1984). Self-produced motion: An organizer of emotional, cognitive, and social development in infancy. In R. Emde & R. Harmon (Eds.), *Continuities and discontinuities* (pp. 175–210). New York, NY: Plenum Press.

Bloom, P. (2002). *How children learn the meanings of words*. Cambridge, MA: MIT Press.

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*, 207–220.

Butterworth, G. (1995). Origins of mind in perception and action. In C. Moore & P. Dunham (Eds.) *Joint attention: Its origins and role in development* (pp. 29–40). Hillsdale, NJ: Lawrence Erlbaum.

Chartrand, T. L., & van Baaren, R. (2009). Human mimicry. *Advances in Experimental Social Psychology, 41*, 219–274.

Corkum, V., & Moore, C. (1998). Origins of joint visual attention in infants. *Developmental Psychology*, *34*, 28–38.

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

Doherty, M. J., & Anderson, J. R. (1999). A new look at gaze: Preschool children's understanding of eye-direction. *Cognitive Development*, *14*, 549–571.

Doherty, M. J., & Anderson, J. R. (2001). People don't keep their heads still when looking to one side, and other people can tell. *Perception*, *30*, 765–767.

Doherty, M. J., Anderson, J. R., & Howieson, L. (2009). The rapid development of explicit gaze judgment ability at 3 years. *Journal of Experimental Child Psychology*, *104*, 296–312.

Einhäuser, W., Schumann, F., Bardins, S., Bartl, K., Böning, G., Schneider, E., König, P. (2007). Human eye-head co-ordination in natural exploration. *Network: Computation in Neural Systems*, *18*, 267–297.

Fogel, A., Lyra, M., & Valsiner, J. (1997). *Dynamics and indeterminism in developmental and social processes*. Mahwah, NJ, Lawrence Erlbaum Associates.

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (in press). Head-mounted eye-tracking: A new method to describe the visual ecology of infants. *Child Development*.

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*, 160–170.

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, *71*, 878–894.

Hartman, E., Houwen, S., Scherder, E., & Visscher, C. (2010). On the relationship between motor performance and executive functioning in children with intellectual disabilities. *Journal of Intellectual Disability Research*, *54*, 468–477.

Hofsten, C. V., & Rosander, K. (1997). Development of smooth pursuit tracking in young infants. *Vision Research*, *37*, 1799–1810.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of Child Language*, *37*, 229–261.

Iverson, J. M., Capirci, O., & Caselli, M. (1994). From communication to language in two modalities. *Cognitive Development*, *9*, 23–43.
Kita, S. (2003). Pointing: A Foundational Building Block of Human Communication. In Kita S. (Ed.), *Pointing: Where language, culture, and cognition meet* pp. 9–33. Mahwah, NJ: Erlbaum.
Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences*, *4*, 50–59.
Langton S. R. H, (2000) The mutual influence of gaze and head orientation in the analysis of social attention direction. *The Quarterly Journal of Experimental Psychology, 53* 825-845.
Larroque, B., Ancel, PY., Marret, S., Marchand, L., Andre, M., Arnaud, C., . (2008). Neurodevelopmental disabilities and special care of 5-year-old children born before 33 weeks of gestation (the EPIPAGE study): A longitudinal cohort study. *The Lancet*, *371*, 813–820.
Loomis, J. M., Kelly, J. W., Pusch, M., Bailenson, J. N., & Beall, A. C. (2008). Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. *Perception*, *37*, 1443–1457.
Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology, 11*, 194-201.
Moore, C., & Corkum, V. (1998). Infant gaze following based on eye direction. *British Journal of Developmental Psychology, 16*, 495–503.
Moore, C. (2006). *The Development of Commonsense Psychology*. Mahwah, NJ: Lawrence Erlbaum Associates.
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*, 279–295.
Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality & Social Psychology Review*, *9*, 184–211.
Pereira, A. F., & Smith, L. B. (2009). Developmental changes in visual object recognition between 18 and 24 months of age. *Developmental Science*, *12*, 67–80.
Pereira, A. F., Smith, L. B., & Yu, C. (2008). Social coordination in toddler's word learning: Interacting systems of perception and action. *Connection Science*, *20*, 73–89.
Pereira, A.F., Yu., C., Smith, L.B., Shen, H. (2009) A First-Person Perspective on a Parent-Child Social Interaction During Object Play. In N. Taatgen, H. van Rijn, J. Nerbonne, & L. Schomaker (Eds.) *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.
Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*, 169–190.
Pierce, K., Muller, R. A., Ambrose, J., Allen, G., & Courchesne, E. (2001). Face processing occurs outside the fusiform 'face area' in autism: Evidence from functional MRI. *Brain*, *124*, 2059–2073.
Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological Science*, *18*, 407–413.

Savelsbergh, G., Hofstein, C. von, & Jonsson, B. (1997). The coupling of head, reach and grasp movement in nine months old infant prehension. *Scandinavian Journal of Psychology*, *38*, 325–333.

Sheya, A. & Smith, L. B. (2011). Development through Sensory-Motor Coordinations. In J. Stewart, O. Gappenne and E.A. Di Paolo (Eds), *Enaction: Towards a New Paradigm for Cognitive Science*. (p. 123). Cambridge, MA: MIT Press.

Shockley, K., Baker, A. A., Richardson, M. J., & Fowler, C. A. (2007). Articulatory constraints on interpersonal postural coordination. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 201–208.

Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and coordinative structures. *Topics in Cognitive Science*, *1*, 305–319.

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

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*, 235–260.

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

Spivey, M. J., Richardson, D., & Dale, R. (2009). The movement of eye and hand as a window into language and cognition. In E. Morsella, J. A. Bargh, and P. M. Gollwitzer (Eds.) *Oxford Handbook of Human Action*, 225–249. New York, NY: Oxford University Press.

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

Toda, S., & Fogel, A. (1993). Infant response to the still-face situation at 3 and 6 months. *Developmental Psychology*, *29*, 532–538.

Toda, S., Fogel, A., & Kawai, M. (1990). Maternal speech to three-month-old infants in the United States and Japan. *Journal of Child Language*, *17*, 279–294.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint Attention: Its Origin and Role in Development*, pp. 103-130. Hillsdale, N. J.: Lawrence Erlbaum.

Yoshida, H. (2010, March). *Early attentional shifting in typically developing children and children with autism*. Talk presented at XVIIth biennial international conference on infant studies*,* Baltimore, MD.

Yoshida, H., & Kushalnagar, P. (2009, April). Attentional flexibility: *Head-cameras reveal different strategies for hearing children, deaf children, and children with ASD.* Talk presented at preconference for the biennial meeting of the Society for Research on Child Development, Denver, CO.

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

Yu, C., Smith, L., Shen, H., Pereira, A., & Smith, T. (2009). Active information selection: Visual attention through the hands. *IEEE Transactions on Autonomous Mental Development*, *1*, 141–151.