HHS Public Access

Author manuscript

Cognition. Author manuscript; available in PMC 2017 July 01.

Published in final edited form as:

Cognition. 2016 July ; 152: 101–107. doi:10.1016/j.cognition.2016.03.005.

From faces to hands: Changing visual input in the first two years

Caitlin M. Fausey[1], **Swapnaa Jayaraman**[2], and **Linda B. Smith**[2]

[1]Department of Psychology, University of Oregon, Eugene, OR 97403

[2]Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405

Abstract

Human development takes place in a social context. Two pervasive sources of social information are faces and hands. Here, we provide the first report of the visual frequency of faces and hands in the everyday scenes available to infants. These scenes were collected by having infants wear head cameras during unconstrained everyday activities. Our corpus of 143 hours of infant-perspective scenes, collected from 34 infants aged 1 month to 2 years, was sampled for analysis at 1/5 Hz. The major finding from this corpus is that the faces and hands of social partners are not equally available throughout the first two years of life. Instead, there is an earlier period of dense face input and a later period of dense hand input. At all ages, hands in these scenes were primarily in contact with objects and the spatio-temporal co-occurrence of hands and faces was greater than expected by chance. The orderliness of the shift from faces to hands suggests a principled transition in the contents of visual experiences and is discussed in terms of the role of developmental gates on the timing and statistics of visual experiences.

Keywords

scene statistics; head camera; faces; hands; infancy; egocentric vision

The world is characterized by many regularities and human learners are sensitive to these, as evident in extensive research on vision, language, causal reasoning, and social intelligence (e.g., Anderson & Schooler, 1991; Griffiths, Steyvers, & Tenenbaum, 2007; Kahneman, 2011; Simoncelli, 2003). A core theoretical problem concerns how the learner discovers which regularities are relevant for learning and how those regularities segregate into different domains of knowledge (e.g., Aslin & Newport, 2012; Frost, Armstrong, Siegelman, & Christiansen, 2015; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). The relevant data for different domains and tasks could be determined by the regularities in the data themselves (e.g., Colunga & Smith, 2005; Rogers & McClelland, 2004; Tenenbaum et al., 2011) or from internal biases that define distinct domains (e.g., Frost et al., 2015; Spelke, 2000). Here, we present evidence for another way in which data for learning may be bundled into segregated sets, by development itself: visual experiences present different regularities

at different developmental points and in so doing development may effectively define distinct datasets of visual information.

Our example case concerns two powerful sources of information for developing infants: human faces and human hands. Faces convey information about the emotional and attentional states of social partners. Hands act on the world; they make things happen. Experimental evidence indicates that infants develop specialized knowledge about the visual properties of faces, enabling the rapid recognition of faces and the meaningful interpretation of facial gestures (see Johnson, 2011). Infants also develop specialized knowledge about seen hand movements, knowledge that supports causal inferences about instrumental actions on objects (e.g., Cannon & Woodward, 2012; Woodward, 2009) and that links gestures and points to reference and word learning (e.g., Carpenter, Nagell, & Tomasello, 1998; Namy & Waxman, 1998; Rader & Zukow-Goldring, 2012). Overall, the evidence suggests a protracted course of development of both kinds of knowledge (see De Heering, Rossion, & Maurer, 2012; Goldin-Meadow & Alibali, 2013) and mature cortical visual representations for faces and hands that are distinct (e.g., Bracci, Ietswaart, Peelen, & Cavina-Pratesi, 2010; Peelen & Downing, 2007).

Although human beings, with their faces and hands, are plentiful in the larger dataset that is human experience, we hypothesize that early visual samples of people are dense with faces (regularities relevant to face processing) and that later samples are dense with hands (regularities relevant for instrumental acts on objects). This hypothesis is suggested by recent discoveries using a new technology, head cameras worn by infants. Although conducted for a variety of purposes by different investigators, all of these studies aimed to capture the visual world of infants and in aggregate they have provided a set of new insights pertinent to the present hypothesis: First, the scenes directly in front of infants are highly selective with respect to the visual information in the larger environment (e.g., Smith, Yu, Yoshida, & Fausey, 2015; Yu & Smith, 2012). Second, properties of these scenes differ systematically from adult-perspective scenes (e.g., Smith, Yu, & Pereira, 2011), from third-person perspective scenes (e.g., Aslin, 2009; Yoshida & Smith, 2008; Yurovsky, Smith, & Yu, 2013), and are not easily predicted by adult intuitions (e.g., Franchak, Kretch, Soska, & Adolph, 2011; Yurovsky, Smith, & Yu, 2013). Third, and most critically, the properties of these scenes are different for children of different ages and developmental abilities (e.g. Frank, Simmons, Yurovsky, & Pusiol, 2013; Kretch, Franchak, & Adolph, 2014; Pereira, James, Jones, & Smith, 2010; Raudies & Gilmore, 2014). Infant-perspective scenes change systematically with development because they depend on the perceiver's body morphology, typical postures and motor skills, abilities, interests, motivations, and caretaking needs. These all change dramatically over the first two years of life, and thus collectively serve as developmental gates to different kinds of visual datasets. In brief, the overarching hypothesis is that development bundles visual experiences into separate datasets for infant learners (see also Adolph & Robinson, 2015; Bertenthal & Campos, 1990; Campos et al., 2000).

One result that has now been reported from studies using head cameras to record everyday at-home experiences is that faces were very frequent in infant-perspective scenes for infants younger than 4 months of age (e.g., Jayaraman, Fausey, & Smith, 2015; Sugden, Mohamed-Ali, & Moulson, 2014). In contrast, laboratory studies of toddler-perspective views found

that the faces of social partners were rarely in the toddlers' views but the hands of the partners were frequently in view (e.g., Deák, Krasno, Triesch, Lewis, & Sepeta, 2014; Franchak et al., 2011; Yu & Smith, 2013). Because the contexts of these studies with younger and older infants were different, this developmental pattern – from visual experiences dense with faces to those that were dense with hands – could be the product of the home versus laboratory contexts of the social interactions. Alternatively, the developmental pattern could be broadly characteristic of age-related changes in infant experiences and could indicate a more pervasive temporal segregation of visual datasets about social agents. Here, we provide evidence by using head cameras to collect a large corpus of infant-perspective scenes during unconstrained at-home activities for infants as young as 1 month and as old as 24 months.

Our use of head cameras builds on the prior developmental research using this method (see Smith et al., 2015, for review) as well as growing multi-disciplinary efforts directed toward understanding egocentric vision (e.g., Fathi, Ren, & Rehg, 2011; Pirsiavash & Ramanan, 2012). Considerable progress in understanding adult vision has been made by studying "natural scenes" (e.g., Geisler, 2008; Simoncelli, 2003). However, these scenes are photographs taken by adults and differ systematically in content and visual properties from the scenes sampled by perceivers as they move about in the world (e.g., Pinto, Cox, & DiCarlo, 2008; see also Foulsham, Walker, & Kingstone, 2011). As noted by Braddick and Atkinson (2011), body-worn cameras are especially important for building a developmentally-indexed corpus of scenes that captures how the visual data change as infants' bodies, postures, interests, and activities change with development. Here, we provide evidence for the general importance of a developmentally-indexed description of egocentric scenes by showing that the content of those scenes changes systematically with age for two important classes of social information.

Method

Participants

The participating infants (n = 34, 17 male) varied in age from 1 to 24 months (see Fausey, Jayaraman, & Smith, 2015, for additional participant information). Prior work suggests that a shift from scenes dense with the faces of social partners to those dense with their hands could occur with increasing engagement in instrumental acts (e.g., in the period around 5 to 11 months; Rochat, 1992; Soska & Adolph, 2014; Woodward, 1998) or perhaps around one year when infants show increased interest in and imitation of others' instrumental acts (e.g., Fagard & Lockman, 2010; Karasik, Tamis-LeMonda, & Adolph, 2011). Because there is no strong prior basis for making fine-grained predictions about the ages across which a transition from many faces to many hands might occur, we sampled infants continuously within the expected broad age range of this transition – from 1 to 16 months. Because some of the laboratory studies indicating a toddler focus on hands have included older infants (near their second birthday, e.g., Smith et al., 2011; Yu & Smith, 2013), we also included more advanced 24-month-olds to measure the distribution of hands and faces in experiences at the end of infancy. The sample of infants was recruited from a database of families maintained for research purposes that is broadly representative of Monroe County, Indiana:

84% European American, 5% African American, 5% Asian American, 2% Latino, 4% Other) and consisted of predominantly working- and middle-class families.

Capturing the scenes

Recording the availability of faces and hands in infants' everyday environments requires a method that does not distort the statistics of those daily environments. Accordingly, we used a commercial wearable camera that was easy for parents to use (Looxcie). The diagonal field of view (FOV) was 75 degrees, vertical FOV was 41 degrees, and horizontal FOV was 69 degrees, with a 2″ to infinity depth of focus. The camera recorded at 30 Hz. The battery life of each camera was approximately two continuous hours; parents were given multiple cameras to use and could alternate and charge the cameras to full battery capacity as they needed. Video was stored on the camera until parents had completed their recording and then was transferred to laboratory computers for storage and processing.

The camera was secured to a hat that was custom fit to the infant so that when the hat was securely placed on the infant the lens was above the nose and did not move. Because the central interest of this project was the faces and hands of others (not the infant's own hands), the camera was situated and adjusted to capture the broad view in front of the infant; as a result, the camera could miss the infant's own in-view hands if those hands were below the infant's chin and close (within 2 inches) to the infant's body (see Smith et al., 2015, for a discussion of these issues). Parents were not told that we were interested in faces or hands but were told that we were interested in their infant's everyday activities and to try to record six hours of video when their child was awake. Hours of recording did not always reach the six hour goal and varied across participants ($M = 4.22$, $SD = 1.76$), but did not vary with age ($t(32) = -.12$, *n.s.*). The total number of scenes collected across all infants was 15,507,450; the analyzed scenes were sampled from this larger dataset as described below. Activities and contexts were primarily captured at home (over 80% of all scenes) but also included some out-of-home settings such as stores and group activities. A time-sampling study of the larger population from which these families were selected (Jayaraman, Fausey, & Smith, submitted) indicated similar proportions of (awake) time in the home that changed little over this age range.

Coding for the presence of faces and hands

To estimate the rate of faces and hands in the collected scenes, scenes were sampled at 1/5 Hz (Figure 1; see also Fausey et al., 2015, for example videos and corresponding 1/5 Hz scenes) leading to a total of 103,383 coded scenes. Sampling at 1/5 Hz should not be biased in any way to faces or hands and appears sufficiently dense to capture major regularities: First, a coarser sampling of scenes at 1/10 Hz yielded the same reliable patterns reported below. Second, a sampling of a different set of scenes (72,000 frames) at 1/5 Hz using new starting points was partially recoded and yielded no reliable differences in the reported patterns (see also Jayaraman et al., 2015).

Each sampled scene was coded by four naïve coders who saw the scenes in a randomly ordered presentation and were asked, in separate passes, one question answerable with "yes" or "no": whether there was a human face present or whether there was a hand present.

Coders were instructed to indicate “yes” if there was whole face or hand or if there was a clearly identifiable part of a face or part of a hand. A scene was defined as “reliably coded” if at least three coders gave the same answer – that is, three “yes” responses or three “no” responses (Faces = 96.5%, Hands = 94.75%); thus, a scene was categorized as containing a face or hand if at least three of four coders had affirmed the presence of the queried entity. Note that three-of-four is a criterion; all the data that contribute to main findings received the same judgment from at least three naïve and independent coders. Scenes that contained a hand were subsequently coded by four independent and naïve coders using the same three out of four agreement criterion, again with either at least three “yes” or at least three “no” judgments defining reliable coding. The four hand measures, coded in separate passes, were: the hand in the scene was the infant's own hand (99.75% reliably coded), the hand in the scene was touching something (89.08% reliably coded), the hand in the scene was holding onto something (86.36% reliably coded), and the hand in the scene was holding a small, carry-able object (95.48% reliably coded).

Results

Each infant's data consists of a set of scenes ($M = 3041$, $SD = 1265$). Thus, there are on average about 3,000 data points per subject and all data are reported in terms of the individual participant. The principal analyses use linear regression to examine whether the frequency of faces and hands in these scenes change as function of age. As indexed by the presence of a face or hand, a person appeared in roughly one-quarter of the captured scenes (.27) and this did not vary with age ($t(32) = .04$, *n.s*). That is, people were just as likely to be in view (with a face and/or hand) for the youngest and oldest infants. The results that follow, therefore, are not due to the differential presence of other people in younger and older infants’ scenes.

The hypothesis is that the likelihood of the two body parts in these scenes changed systematically with age. As predicted and as shown in Figure 2, faces were more frequent in the scenes captured from the youngest infants and declined with age (linear trend: $F(1, 32) = 10.73$, $p < .005$, Figure 2a). By contrast, the frequency of hands increased with age (linear trend: $F(1, 32) = 26.11$, $p < .001$, Figure 2b). The relative frequency of faces and hands within the scenes captured from individual infants also showed an orderly transition from “relatively more faces” to “relatively more hands” (delta score: proportion faces minus proportion hands; linear trend: $F(1, 32) = 55.05$, $p < .001$, Figure 2c). Figure 2c shows that the age-related decline in faces and the age-related increase in hands leads to an early period in which faces are dominant, a later period in which hands are dominant, and a middle period in which faces and hands are both more similarly prevalent.

The orderliness of this transition is notable given that these scenes were sampled from several hours of everyday activities of different infants with no constraints on those activities. Thus, the findings may indicate a systematic transition in the contents of visual experiences, a transition in the datasets for statistical learning.

The hands captured in these infant-perspective scenes were overwhelmingly the hands of other people (.92 of all scenes with hands) and did not vary by age, $t(30) = .15$, *n.s.*,

excepting one outlier, a two-year-old, whose frequency of own hands exceeded 4 SD above the group mean. Hands were touching (.76 of scenes with hands) or holding (.48 of scenes with hands) something and this key property of hands acting-on-objects also did not vary by age: touching $r(31)$, = .15, *n.s.*; holding $r(31)$ = .16, *n.s.*; note that data from one infant who was three weeks old, an age at which faces dominate, did not contribute to these and subsequent analyses because no hands appeared in her scenes. Because hands were much more frequent in infant-perspective scenes at older than at younger ages, and because these hands were typically in contact with objects, the changing contents of visual scenes may be understood as a developmental shift from data about faces to data about manual actions on objects.

This developmental segregation of visual scenes with faces versus those with hands does not necessarily imply that they are completely segregated in experience (see Libertus & Needham, 2011; Slaughter & Heron-Delaney, 2011). Although there were very few hands in the head-camera scenes of the youngest infants, the hands they did see may be spatiotemporally proximal to faces. To test this possibility, we measured whether the presence of a hand (infrequent for young infants) signaled the presence of a face in that same scene or in a temporally nearby scene. More specifically, each infant's sampled (at 1/5 Hz) head camera scenes were assembled into their real time order (Figure 3a). For each scene in this stream that contained a hand, the nearest scene that contained a face was identified. The proportions of hands that occurred with a face simultaneously, within five seconds of a face or within ten seconds are shown in Figure 3b. For very young infants, hands occurred at the same time as, or shortly before or after, a face. Despite the relative infrequency of hands in the scenes from very young infants, this spatiotemporal co-occurrence provides an early basis for integrating face and hand information about a single person.

To evaluate whether this structure is due to the base rates of faces and hands or whether the stream of experience provides more spatiotemporal structure than random co-occurrence, we compared each infant's actual stream to a shuffled stream.

Specifically, each scene could contain a face, a hand, both or neither; thus, each infant's data was decomposed into a face stream and a hand stream. Each infant's face stream was shuffled 100 times and paired with the real hand stream. This preserves each infant's frequency of faces and hands but randomizes the proximity of faces to hands in the stream. The proportion of hands with a face simultaneous, within five seconds, and within ten seconds was calculated on each shuffle. The structure of each infant's real stream was compared to the median of their shuffled streams. A difference score greater than zero indicates non-random spatiotemporal structure. The results indicate structure greater than that expected by the base frequencies across the sample of infants (simultaneous: $t(32)$ = 15.19, $p < .001$, $d = 2.64$; within five seconds: $t(32) = 14.04$, $p < .001$, $d = 2.44$; within ten seconds $t(32) = 6.60$, $p < .001$, $d = 1.15$). Further, this systematicity appears to be particularly dramatic for the youngest infants, with the degree to which the available structure differs from random declining with age (simultaneous: $F(1,31) = 32.54$, $p < .001$; within five seconds: $F(1,31) = 6.12$, $p = .02$; within ten seconds: $F(1,31) = 1.37$, *n.s.*). Faces dominated the visual scenes of the youngest infants and less frequent hands systematically co-occurred with faces for these youngest infants; for older infants, hands did not as

frequently co-occur with faces and thus constitute a class of experiences more segregated in real time from faces.

Discussion

The contents of infant-perspective scenes change over the first two years of life, an unsurprising fact given the remarkable changes in abilities and interests over this period. What is perhaps surprising, though hinted at by previous head-camera studies, is that the visual information about the body parts of social agents in the lives of infants also changes. The present findings document that earlier visual experiences about people are dense with faces and that later experiences are dense with hands. With age, the rate of decreasing faces and the rate of increasing hands in the input both appear to be incremental; the joint effect of these two changes over the first two years of life leads to an early period in which faces dominate and to a later one in which hands dominate. In brief, visual experiences of people are developmentally bundled into datasets. This bundling may be a key component in explanations of how visual processes become specialized to different sources of social information.

An extensive literature indicates that human face processing is characterized by special properties, including its developmental course (see McKone, Kanwisher, & Duchaine, 2007; Nelson, 2003). Newborn infants are biased to look at very simple “face-like” arrays consisting of two dark blobs (eyes) within a face-shaped contour (e.g., Goren, Sarty, & Wu, 1975; Johnson, Dziurawiec, Ellis, & Morton, 1991). This neonatal bias has been interpreted in terms of an “experience-expectant innate template” (e.g., McKone et al., 2007) that directs infant attention to faces and ensures the engagement of the visual system with face stimuli (e.g., Morton & Johnson, 1991). These face experiences lead ultimately to the development of visual processes highly specialized for extracting the relevant information from faces for rapid identification, categorization, and social judgment (see McKone, Crookes, Jeffery, & Dilks, 2012; Pascalis & Kelly, 2009; Scherf & Scott, 2012, for reviews). An early visual environment that is sufficiently dense in faces *relative to other body parts* may be essential. By keeping the visual signal about meaningful social events relatively clean with faces, the constrained input may tune (or maintain; Aslin, 1981) experience-expectant neural processes in the direction of face specific regularities.

We know much less about the development of hand processing. However, findings from several somewhat disjointed literatures suggest that hands are also characterized by specialized visual processes, albeit ones that may be specifically about manipulable objects (e.g., Borghi et al., 2007; Vainio, Symes, Ellis, Tucker, & Ottoboni, 2008). For example, a large and varied literature studying adults shows that hands direct attention to objects (e.g., Abrams, Davoli, Du, Knapp, & Paull, 2008; Reed, Grubb, & Steele, 2006; Tseng, Bridgeman, & Juan, 2012) and that hand actions and shapes directly inform perceivers about object properties (e.g., Klatzky, Pellegrino, McCloskey, & Doherty, 1989). Evidence from infants shows that they are sensitive to the causal and semantic structure of manual actions (see Sommerville, Upshaw, & Loucks, 2012, for review) and how points, gestures, and manual actions guide visual attention to objects (e.g., Butterworth, 2003; Goldin-Meadow & Butcher, 2003; Tomasello, Carpenter, & Liszkowski, 2007; Volterra, Caselli, Capirci, &

Pizzuto, 2005; Yu & Smith, 2013). These phenomena are principally studied, and show their most systematic patterns of growth, late in infancy, from just before the first birthday to well into the second year. The present results suggest that the developmental timing of this growth in knowledge about the information conveyed by hands may be in part determined by the increased prevalence of hands of social partners acting on objects in the visual input.

We propose that the segregation of visual information about faces and hands supports the development of face and hand visual processing that becomes optimized to the specific social information provided by each, a hypothesis in need of more direct test in future research. But if faces and hands are separate datasets in developmental time, how do infants learn to coordinate the social cues provided by each? The extant evidence shows that very young infants follow another's gaze in highly restricted viewing contexts (e.g., Farroni, Johnson, Brockbank, & Simion, 2000; Farroni, Massaccesi, Pividori, & Johnson, 2004; Vecera & Johnson, 1995), but also shows that the spatial resolution of gaze following is often not sufficient for navigating real-time social interactions in more spatially complex social settings (e.g., Doherty, Anderson, & Howieson, 2009; Loomis, Kelly, Pusch, Bailenson, & Beall, 2008; Vida & Maurer, 2012; Yu & Smith, 2013). Critically, the spatial complexity of social interactions explodes as infants become more physically active and transition from social interactions dominated by face-to-face play to social interactions that are dominated by shared engagement with objects (see Striano & Reed, 2006). In a study using simultaneous head-mounted eye trackers worn by toddlers and parents, Yu and Smith (2013) found that one-year-old infants coordinated their own gaze with that of the parent, not by following parent eye-gaze, but by fixating on and following parent hand movements to objects (to which parent eye gaze was also dynamically coordinated). Computational modelers have further proposed that hand-following – with its superior spatial precision – may tune and refine gaze following (e.g., Triesch, Teuscher, Deák, & Carlson, 2006; Ullman, Harari, & Dorfman, 2012; Yu & Smith, 2013), which in principle could enable gaze skills to increasingly meet the challenge of complex interactions with objects. Other evidence suggests that gaze following may emerge later in childhood, potentially after opportunities to learn from hands acting on objects (e.g., Deák et al., 2014). These issues highlight the critical need to continue the task begun here, determining how the regularities in infant visual experiences of faces and others' hands change with age, and the importance of a new line of research only possible given the study of developmentally-indexed egocentric scenes: how the changing regularities in those scenes align with infants' developing abilities to use face and hand information.

What underlies the age-related changes in infant-perspective scenes? One possibility is that the timetable is driven by changes in infant interests and motivations. Studies in which infants view experimenter-selected scenes indicate a greater visual interest in faces in early infancy (e.g., Ahtola et al., 2014; Amso, Haas, & Markant, 2014; DiGiorgio, Turati, Altoè, & Simion, 2012; Frank, Amso, & Johnson, 2014; Frank, Vul, & Johnson, 2009; Gluckman & Johnson, 2013; Libertus & Needham, 2014) and greater looking to hands and instrumental actions on objects with increasing age (Aslin, 2009; Frank, Vul, & Saxe, 2012; see also Reddy, Markova, & Wallot, 2013; Slaughter & Heron-Delaney, 2011). Changing interests, in turn, may be driven by infants' changing abilities, including, and perhaps especially, changing motor skills. The transitions to reaching (e.g., Libertus & Needham, 2014), sitting

stably (e.g., Soska & Adolph, 2014), and crawling and walking (e.g., Karasik et al., 2011) are all associated with changes in social interactions. The present study was focused on social information and thus used a head camera adjusted to a geometry that captures others' hands. Infants' visual experiences with their own hands could also be a contributing factor in the developmental changes (e.g., Woodward, 2009). A clearly needed next step is the joint study of motor skills, input statistics, and developing perceptual expertise about faces, others' hands, and own hands in order to understand the detailed pathways of cause and consequence over developmental time (e.g., Byrge, Sporns, & Smith, 2014).

Theories of how evolution works through developmental process have noted how evolutionarily important outcomes are often restricted by the density and ordering of different classes of sensory experiences (e.g., Gottlieb, 1991; Lord, 2013; Turkewitz & Kenny, 1982). This idea is often conceptualized in terms of "developmental niches" that provide different environments with different regularities (e.g., Gottlieb, 1991; West & King, 1987) at different points in time. These niches – like a developmental period dense in face inputs or dense in hand inputs – may be jointly determined and constrained by evolutionary and developmental processes in multiple ways. That evolution, across species and across domains, has chosen to developmentally bundle kinds of input data suggests that systematically segregated and ordered datasets may play a key role in helping organisms extract the relevant information for the many tasks that have to be solved.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research was supported in part by the Faculty Research Support Program at Indiana University, and in part, by NIH grants R21HD068475 and NSF BCS-1523982 to Smith. Fausey was supported by NIH T32 HD007475. We thank Char Wozniak and Ariel La for help with data collection and video pre-processing.

References

Abrams RA, Davoli CC, Du F, Knapp WK, Paull D. Altered vision near the hands. Cognition. 2008; 107:1035–1047. [PubMed: 17977524]

Adolph, KE.; Robinson, SR. Motor development.. In: Damon, W.; Lerner, R., editors. Handbook of child psychology. 7th ed.. Wiley; New York: 2015. p. 113-156.

Ahtola E, Stjerna S, Yrttiaho S, Nelson CA, Leppänen JM, Vanhatalo S. Dynamic eye tracking based metrics for infant gaze patterns in the face-distractor competition paradigm. PloS One. 2014; 9(5):e97299. [PubMed: 24845102]

Amso D, Haas S, Markant J. An eye tracking investigation of developmental change in bottom-up attention orienting to faces in cluttered natural scenes. PloS One. 2014; 9(1):e85701. [PubMed: 24465653]

Anderson JR, Schooler LJ. Reflections of the environment in memory. Psychological Science. 1991; 2(6):396–408.

Aslin, RN. Experiential influences and sensitive periods in perceptual development: A unified model.. In: Aslin, RN.; Alberts, JR.; Petersen, MR., editors. Development of perception. Vol. 2. Academic Press; New York: 1981. p. 45-93.

Aslin RN. How infants view natural scenes gathered from a head-mounted camera. Optometry and vision science: official publication of the American Academy of Optometry. 2009; 86(6):561–565. [PubMed: 19417702]

Aslin RN, Newport EL. Statistical learning from acquiring specific items to forming general rules. Current Directions in Psychological Science. 2012; 21(3):170–176. [PubMed: 24000273]

Bertenthal, BI.; Campos, JJ. A systems approach to the organizing effects of self-produced locomotion during infancy.. In: Rovee-Collier, C.; Lipsitt, LP., editors. Advances in infancy research. Vol. 6. Ablex; Norwood, NJ: 1990. p. 1-60.

Borghi AM, Bonfiglioli C, Lugli L, Ricciardelli P, Rubichi S, Nicoletti R. Are visual stimuli sufficient to evoke motor information?: Studies with hand primes. Neuroscience Letters. 2007; 411(1):17–21. [PubMed: 17052839]

Bracci S, Ietswaart M, Peelen MV, Cavina-Pratesi C. Dissociable neural responses to hands and non-hand body parts in human left extrastriate visual cortex. Journal of Neurophysiology. 2010; 103(6): 3389–3397. [PubMed: 20393066]

Braddick O, Atkinson J. Development of human visual function. Vision research. 2011; 51(13):1588–1609. [PubMed: 21356229]

Butterworth, G. Pointing is the royal road to language for babies.. In: Kita, S., editor. Pointing: Where language, culture, and cognition meet. Lawrence Erlbaum; Mahwah, NJ: 2003. p. 9-33.

Byrge L, Sporns O, Smith LB. Developmental process emerges from extended brain–body–behavior networks. Trends in Cognitive Sciences. 2014; 18(8):395–403. [PubMed: 24862251]

Campos JJ, Anderson DI, Barbu-Roth MA, Hubbard EM, Hertenstein MJ, Witherington D. Travel broadens the mind. Infancy. 2000; 1(2):149–219.

Cannon EN, Woodward AL. Infants generate goal-based action predictions. Developmental Science. 2012; 15(2):292–298. [PubMed: 22356184]

Carpenter M, Nagell K, Tomasello M. Social cognition, joint attention, and communicative competence from 9 to 15 months of age. Monographs of the Society for Research in Child Development. 1998; 63(4) Serial No. 255.

Colunga E, Smith LB. From the lexicon to expectations about kinds: A role for associative learning. Psychological Review. 2005; 112(2):347–382. [PubMed: 15783290]

De Heering A, Rossion B, Maurer D. Developmental changes in face recognition during childhood: Evidence from upright and inverted faces. Cognitive Development. 2012; 27(1):17–27.

Deák GO, Krasno AM, Triesch J, Lewis J, Sepeta L. Watch the hands: infants can learn to follow gaze by seeing adults manipulate objects. Developmental Science. 2014; 17(2):270–281. [PubMed: 24387193]

DiGiorgio E, Turati C, Altoè G, Simion F. Face detection in complex visual displays: an eye-tracking study with 3-and 6-month-old infants and adults. Journal of Experimental Child Psychology. 2012; 113(1):66–77. [PubMed: 22695758]

Doherty MJ, Anderson JR, Howieson L. The rapid development of explicit gaze judgment ability at 3 years. Journal of Experimental Child Psychology. 2009; 104:296–312. [PubMed: 19640550]

Fagard J, Lockman JJ. Change in imitation for object manipulation between 10 and 12 months of age. Developmental Psychobiology. 2010; 52(1):90–99. [PubMed: 19937747]

Farroni T, Johnson MH, Brockbank M, Simion F. Infants' use of gaze direction to cue attention: The importance of perceived motion. Visual Cognition. 2000; 7:705–718.

Farroni T, Massaccesi S, Pividori D, Johnson MH. Gaze following in newborns. Infancy. 2004; 5:39–60.

Fathi, A.; Ren, X.; Rehg, JM. Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference On. IEEE; Jun. 2011 Learning to recognize objects in egocentric activities.; p. 3281-3288.

Fausey CM, Jayaraman S, Smith LB. From faces to hands: Changing visual input in the first two years. Databrary. 2015

Foulsham T, Walker E, Kingstone A. The where, what and when of gaze allocation in the lab and the natural environment. Vision Research. 2011; 51(17):1920–1931. [PubMed: 21784095]

Franchak JM, Kretch KS, Soska KC, Adolph KE. Head-mounted eye-tracking: A new method to describe infant looking. Child Development. 2011; 82(6):1738–1750. [PubMed: 22023310]

Frank MC, Amso D, Johnson SP. Visual search and attention to faces during early infancy. Journal of Experimental Child Psychology. 2014; 118:13–26. [PubMed: 24211654]

Frank MC, Simmons K, Yurovsky D, Pusiol G. Developmental and postural changes in children's visual access to faces. Proceedings of the 35th annual meeting of the Cognitive Science Society. 2013

Frank MC, Vul E, Johnson SP. Development of infants' attention to faces during the first year. Cognition. 2009; 110(2):160–170. [PubMed: 19114280]

Frank MC, Vul E, Saxe R. Measuring the Development of Social Attention Using Free-Viewing. Infancy. 2012; 17(4):355–375.

Frost R, Armstrong BC, Siegelman N, Christiansen MH. Domain generality versus modality specificity: the paradox of statistical learning. Trends in Cognitive Sciences. 2015; 19(3):117–125. [PubMed: 25631249]

Geisler WS. Visual perception and the statistical properties of natural scenes. Annual Review of Psychology. 2008; 59:167–192.

Gluckman M, Johnson SP. Attentional capture by social stimuli in young infants. Frontiers in Psychology. 2013; 4 doi: 10.3389/fpsyg.2013.00527.

Goldin-Meadow S, Alibali MW. Gesture's role in speaking, learning, and creating language. Annual Review of Psychology. 2013; 64:257–283.

Goldin-Meadow, S.; Butcher, C. Pointing toward two-word speech in young children.. In: Kita, S., editor. Pointing: Where language, culture, and cognition meet. Lawrence Erlbaum; Mahwah, NJ: 2003. p. 85-107.

Goren CC, Sarty M, Wu PY. Visual following and pattern discrimination of face-like stimuli by newborn infants. Pediatrics. 1975; 56(4):544–549. [PubMed: 1165958]

Gottlieb G. Experiential canalization of behavioral development: Theory. Developmental Psychology. 1991; 27(1):4–13.

Griffiths TL, Steyvers M, Tenenbaum JB. Topics in semantic representation. Psychological Review. 2007; 114(2):211–244. [PubMed: 17500626]

Jayaraman S, Fausey CM, Smith LB. The faces in infant-perspective scenes change over the first year of life. PLoS One. 2015 doi: 10.1371/journal.pone.0123780.

Jayaraman S, Fausey CM, Smith LB. Why are faces denser in the visual experiences of younger than older infants?. submitted.

Johnson, MH. Face perception: A developmental perspective.. In: Calder, AJ.; Rhodes, G.; Johnson, MH.; Haxby, JV., editors. The Oxford handbook of face perception. Oxford University Press; Oxford, UK: 2011. p. 3-14.

Johnson MH, Dziurawiec S, Ellis H, Morton J. Newborns' preferential tracking of face-like stimuli and its subsequent decline. Cognition. 1991; 40(1):1–19. [PubMed: 1786670]

Kahneman, D. Thinking, fast and slow. Farrar, Straus and Giroux; New York: 2011.

Karasik LB, Tamis-LeMonda CS, Adolph KE. Transition from crawling to walking and infants' actions with objects and people. Child Development. 2011; 82(4):1199–1209. [PubMed: 21545581]

Klatzky RL, Pellegrino JW, McCloskey BP, Doherty S. Can you squeeze a tomato? The role of motor representations in semantic sensibility judgments. Journal of Memory and Language. 1989; 28(1): 56–77.

Kretch KS, Franchak JM, Adolph KE. Crawling and walking infants see the world differently. Child Development. 2014; 85(4):1503–1518. [PubMed: 24341362]

Libertus K, Needham A. Reaching experience increases face preference in 3-month-old infants. Developmental Science. 2011; 14(6):1355–1364. [PubMed: 22010895]

Libertus K, Needham A. Face preference in infancy and its relation to motor activity. International Journal of Behavioral Development. 2014; 38(6):529–538.

Loomis JM, Kelly JW, Pusch M, Bailenson JN, Beall AC. Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. Perception. 2008; 37:1443–1457. [PubMed: 18986070]

Lord K. A comparison of the sensory development of wolves (Canis lupus lupus) and dogs (Canis lupus familiaris). Ethology. 2013; 119(2):110–120.

McKone E, Crookes K, Jeffery L, Dilks DD. A critical review of the development of face recognition: Experience is less important than previously believed. Cognitive Neuropsychology. 2012; 29(1-2): 174–212. [PubMed: 22360676]

McKone E, Kanwisher N, Duchaine BC. Can generic expertise explain special processing for faces? Trends in Cognitive Sciences. 2007; 11(1):8–15. [PubMed: 17129746]

Morton J, Johnson MH. CONSPEC and CONLERN: a two-process theory of infant face recognition. Psychological Review. 1991; 98(2):164–181. [PubMed: 2047512]

Namy LL, Waxman SR. Words and gestures: Infants' interpretations of different forms of symbolic reference. Child Development. 1998; 69(2):295–308. [PubMed: 9586206]

Nelson, CA. The development of face recognition reflects an experience-expectant and activity-dependent process.. In: Pascalis, O.; Slater, A., editors. The development of face processing in infancy and early childhood: Current perspectives. Nova Science Publishers; New York, NY: 2003. p. 79-97.

Pascalis O, Kelly DJ. The origins of face processing in humans: Phylogeny and ontogeny. Perspectives on Psychological Science. 2009; 4(2):200–209. [PubMed: 26158945]

Peelen MV, Downing PE. The neural basis of visual body perception. Nature Reviews Neuroscience. 2007; 8(8):636–648. [PubMed: 17643089]

Pereira AF, James KH, Jones SS, Smith LB. Early biases and developmental changes in self-generated object views. Journal of Vision. 2010; 10(11):1–13.

Pinto N, Cox DD, DiCarlo JJ. Why is real-world visual object recognition hard?. PLoS Computational Biology. 2008; 4(1):e27. [PubMed: 18225950]

Pirsiavash, H.; Ramanan, D. Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE; 2012. June). Detecting activities of daily living in first-person camera views.; p. 2847-2854.

Rader NDV, Zukow-Goldring P. Caregivers' gestures direct infant attention during early word learning: the importance of dynamic synchrony. Language Sciences. 2012; 34(5):559–568.

Raudies F, Gilmore R. An analysis of optic flow observed by infants during natural activities. Journal of Vision. 2014; 14(10):226.

Reddy V, Markova G, Wallot S. Anticipatory adjustments to being picked up infancy. PLoS One. 2013; 8(6):e65289. [PubMed: 23840324]

Reed C, Grubb JD, Steele C. Hands up: Attentional prioritization of space near the hand. Journal of Experimental Psychology: Human Perception and Performance. 2006; 32:166–177. [PubMed: 16478334]

Rochat P. Self-sitting and reaching in 5-8 month old infants: The impact of posture and its development on early eye-hand coordination. Journal of Motor Behavior. 1992; 24(2):210–220. [PubMed: 14977620]

Rogers, TT.; McClelland, JL. Semantic cognition: A parallel distributed processing approach. MIT Press; Cambridge, MA: 2004.

Scherf KS, Scott LS. Connecting developmental trajectories: Biases in face processing from infancy to adulthood. Developmental Psychobiology. 2012; 54(6):643–663. [PubMed: 22711622]

Simoncelli EP. Vision and the statistics of the visual environment. Current Opinion in Neurobiology. 2003; 13(2):144–149. [PubMed: 12744966]

Slaughter V, Heron-Delaney M. When do infants expect hands to be connected to a person?. Journal of Experimental Child Psychology. 2011; 108(1):220–227. [PubMed: 20864118]

Smith LB, Yu C, Pereira AF. Not your mother's view: The dynamics of toddler visual experience. Developmental Science. 2011; 14(1):9–17. [PubMed: 21159083]

Smith LB, Yu C, Yoshida H, Fausey CM. Contributions of head-mounted cameras to studying the visual environments of infants and young children. Journal of Cognition and Development. 2015; 16(3):407–419. [PubMed: 26257584]

Sommerville, JA.; Upshaw, MB.; Loucks, J. The nature of goal-directed action representations in infancy.. In: Xu, F.; Kushnir, T., editors. Advances in child development. Vol. 43. Academic Press; Waltham, MA: 2012. p. 351-384.

Soska KC, Adolph KE. Postural position constrains multimodal object exploration in infants. Infancy. 2014; 19(2):138–161. [PubMed: 24639621]

Spelke ES. Core knowledge. American Psychologist. 2000; 55(11):1233–1243. [PubMed: 11280937]

Striano T, Reid VM. Social cognition in the first year. Trends in Cognitive Sciences. 2006; 10(10): 471–476. [PubMed: 16942896]

Sugden NA, Mohamed-Ali MI, Moulson MC. I spy with my little eye: Typical, daily exposure to faces documented from a first-person infant perspective. Developmental Psychobiology. 2014; 56(2): 249–261. [PubMed: 24285109]

Tenenbaum JB, Kemp C, Griffiths TL, Goodman ND. How to grow a mind: Statistics, structure, and abstraction. Science. 2011; 331(6022):1279–1285. [PubMed: 21393536]

Tomasello M, Carpenter M, Liszkowski U. A new look at infant pointing. Child Development. 2007; 78(3):705–722. [PubMed: 17516997]

Triesch J, Teuscher C, Deák GO, Carlson E. Gaze following: why (not) learn it?. Developmental Science. 2006; 9(2):125–147. [PubMed: 16472311]

Tseng P, Bridgeman B, Juan CH. Take the matter into your own hands: a brief review of the effect of nearby-hands on visual processing. Vision Research. 2012; 72:74–77. [PubMed: 23010259]

Turkewitz G, Kenny PA. Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement. Developmental Psychobiology. 1982; 15(4): 357–368. [PubMed: 7106395]

Ullman S, Harari D, Dorfman N. From simple innate biases to complex visual concepts. Proceedings of the National Academy of Sciences. 2012; 109(44):18215–18220.

Vainio L, Symes E, Ellis R, Tucker M, Ottoboni G. On the relations between action planning, object identification, and motor representations of observed actions and objects. Cognition. 2008; 108(2): 444–465. [PubMed: 18452910]

Vecera SP, Johnson MH. Gaze detection and the cortical processing of faces: Evidence from infants and adults. Visual Cognition. 1995; 2:59–87.

Vida M, Maurer D. Fine-grained sensitivity to vertical differences in triadic gaze is slow to develop. Journal of Vision. 2012; 12:634.

Volterra, V.; Caselli, MC.; Capirci, O.; Pizzuto, E. Gesture and the emergence and development of language.. In: Tomasello, M.; Slobin, D., editors. Beyond nature-nurture: Essays in honor of Elizabeth Bates. Lawrence Erlbaum Associates; Mahwah, NJ: 2005. p. 3-40.

West MJ, King AP. Settling nature and nurture into an ontogenetic niche. Developmental Psychobiology. 1987; 20(5):549–562. [PubMed: 3678619]

Woodward AL. Infants selectively encode the goal object of an actor's reach. Cognition. 1998; 69(1): 1–34. [PubMed: 9871370]

Woodward AL. Infants' grasp of others' intentions. Current Directions in Psychological Science. 2009; 18(1):53–57. [PubMed: 23645974]

Yoshida H, Smith LB. What's in view for toddlers? Using a head camera to study visual experience. Infancy. 2008; 13(3):229–248. [PubMed: 20585411]

Yu C, Smith LB. Embodied attention and word learning by toddlers. Cognition. 2012; 125(2):244–262. [PubMed: 22878116]

Yu C, Smith LB. Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. PLoS One. 2013; 8(11):e79659. [PubMed: 24236151]

Yurovsky D, Smith LB, Yu C. Statistical word learning at scale: the baby's view is better. Developmental Science. 2013; 16(6):959–966. [PubMed: 24118720]

Research Highlights

Visual frequency of faces and hands in 15 million infant-perspective everyday scenes.

Faces and hands of social partners not equally available throughout first two years.

Earlier period of dense face input and later period of dense hand input.

Hands of social partners in contact with objects.

Systematic spatiotemporal co-occurrence of hands and faces.

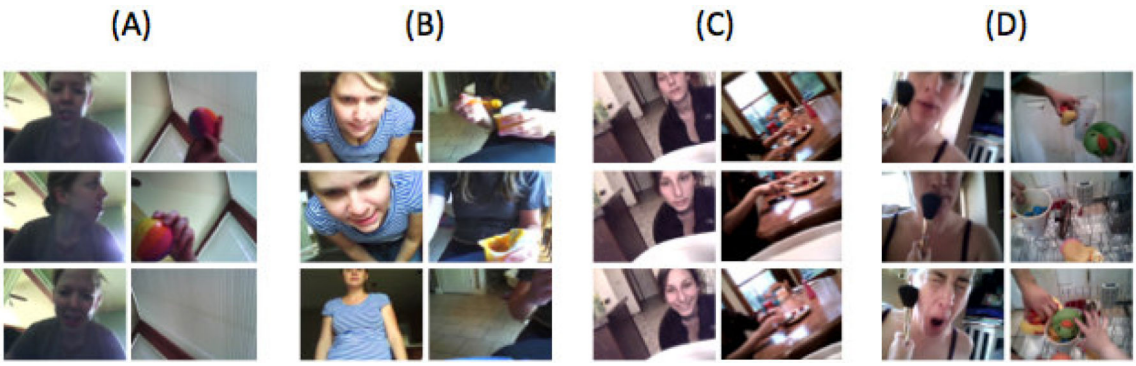

Figure 1.
Example streams of 15 seconds of continuous recording (left: faces; right: hands) sampled at 1/5 Hz from (A) 6-week-old, (B) 31-week-old, (C) 53-week-old, (D) 102-week-old.

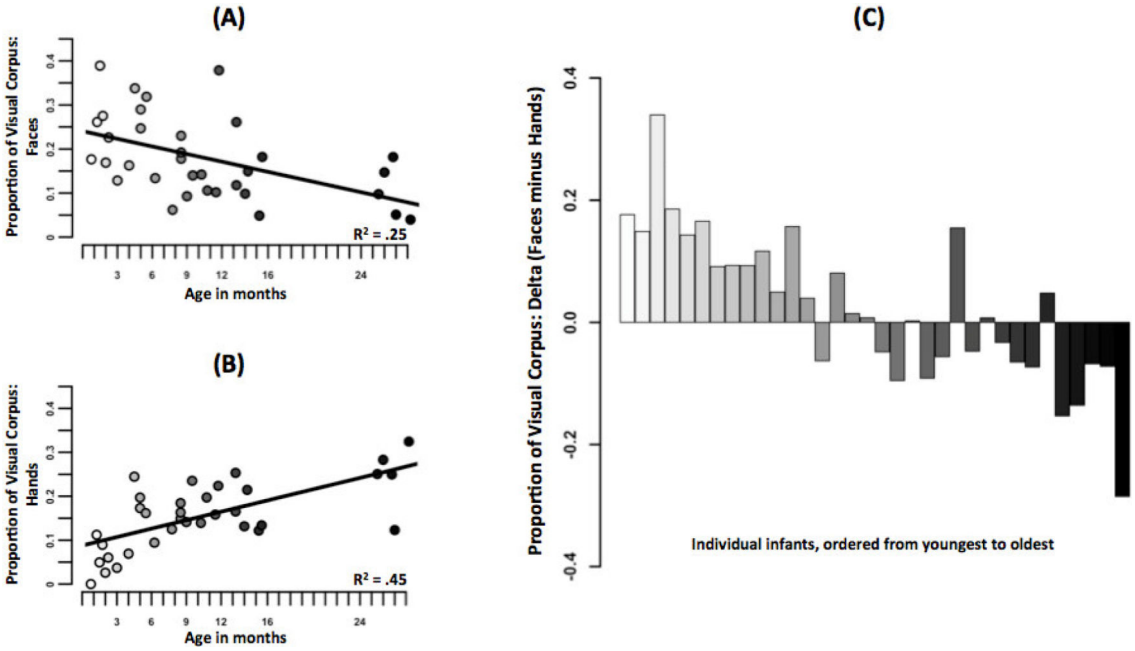

Figure 2.

The changing contents of developmentally-indexed scenes. (A) Decreasing availability of faces, (B) Increasing availability of hands, (C) Relative frequency of faces and hands for each infant in this visual corpus.

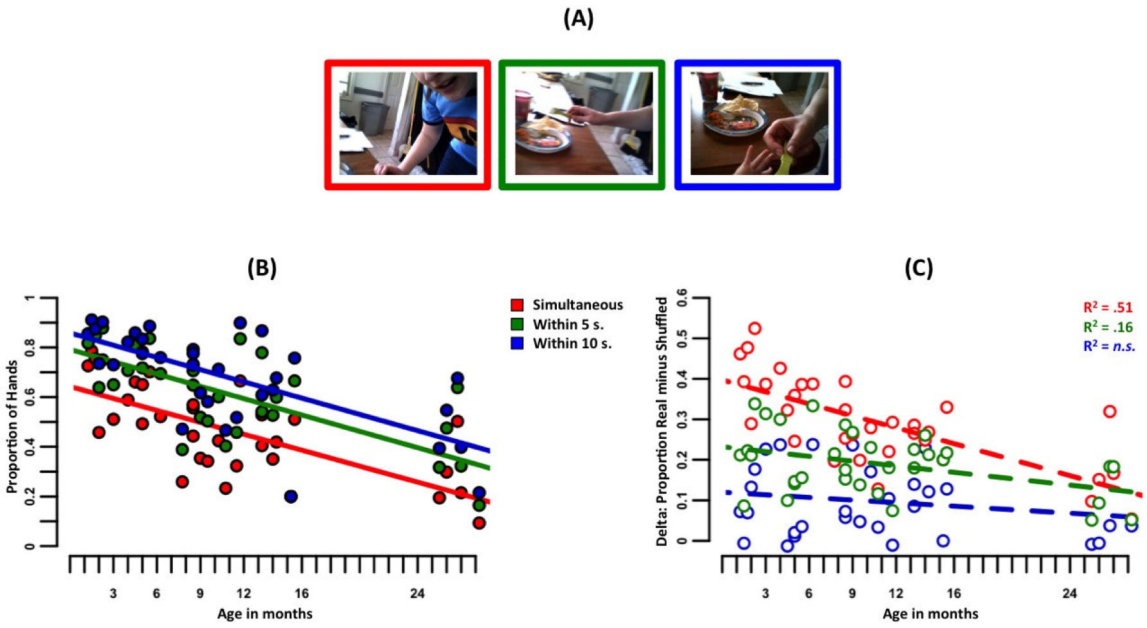

Figure 3.

Temporal proximity of faces to hands. (A) Example 15 second continuous episode. A face appears simultaneously with the hand in the red scene, five seconds from the hand in the green scene, and ten seconds from the hand in the blue scene. (B) Structure in time. The proportion of hands with a face available within three time windows, for each infant in this visual corpus. (C) Structure in time is non-random, especially for the youngest infants. Each point represents the difference between temporal structure available in real and shuffled sequence data (see text for details).